

# Interventions to Educate Family Physicians to Change Test Ordering: Systematic Review of Randomized Controlled Trials

Roger Edmund Thomas, MD, PhD, CCFP, MRCGP<sup>1</sup>, Marcus Vaska, MLIS<sup>2</sup>,  
 Christopher Naugler, BScH, MSc, CCFP, FCFP, FRCPC<sup>1,3</sup>,  
 and Tanvir Turin Chowdhury, PhD, MBBS<sup>1</sup>

## Abstract

The purpose is to systematically review randomised controlled trials (RCTs) to change family physicians' laboratory test-ordering. We searched 15 electronic databases (no language/date limitations). We identified 29 RCTs (4,111 physicians, 175,563 patients). Six studies specifically focused on reducing unnecessary tests, 23 on increasing screening tests. Using Cochrane methodology 48.5% of studies were low risk-of-bias for randomisation, 7% concealment of randomisation, 17% blinding of participants/personnel, 21% blinding outcome assessors, 27.5% attrition, 93% selective reporting. Only six studies were low risk for both randomisation and attrition. Twelve studies performed a power computation, three an intention-to-treat analysis and 13 statistically controlled clustering. Unweighted averages were computed to compare intervention/control groups for tests assessed by >5 studies. The results were that fourteen studies assessed lipids (average 10% more tests than control), 14 diabetes (average 8% > control), 5 cervical smears, 2 INR, one each thyroid, fecal occult-blood, cotinine, throat-swabs, testing after prescribing, and urine-cultures. Six studies aimed to decrease test groups (average decrease 18%), and two to increase test groups. Intervention strategies: one study used education (no change); two feedback (one 5% increase, one 27% desired decrease); eight education + feedback (average increase in desired direction >control 4.9%), ten system change (average increase 14.9%), one system change + feedback (increases 5-44%), three education + system change (average increase 6%), three education + system change + feedback (average 7.7% increase), one delayed testing. The conclusions are that only six RCTs were assessed at low risk of bias from both randomisation and attrition. Nevertheless, despite methodological shortcomings studies that found large changes (e.g. >20%) probably obtained real change.

## Keywords

family doctors, randomized controlled trials, lab tests, systematic review, meta-analysis

Received October 14, 2015; Revised January 14, 2016; Accepted January 23, 2016.

## Introduction

There is concern in several countries about the increasing numbers of laboratory tests ordered by community family physicians and the wide variation in test ordering by family physicians. The increase in testing can be illustrated for several countries. In 2003, the Australian government's initiative to improve the quality of care of chronic illnesses by family physicians and general practitioners (GPs; defined as general primary care physicians without specialty training in family

<sup>1</sup> Department of Family Medicine, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada

<sup>2</sup> Knowledge Resource Service, Holy Cross Centre, Calgary, Alberta, Canada

<sup>3</sup> Departments of Pathology & Laboratory Medicine, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada

### Corresponding Author:

Roger Edmund Thomas, Faculty of Medicine, University of Calgary, 3330 Hospital Drive NW, Calgary, Alberta, Canada T2M 1M1.

Email: [rthomas@ucalgary.ca](mailto:rthomas@ucalgary.ca)



Creative Commons CC-BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 3.0 License (<http://www.creativecommons.org/licenses/by-nc/3.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

medicine) had a marked effect on specific areas of laboratory test ordering. Although the number of family physicians/GPs increased by 10.6% between 2003 and 2007/2008, clinical activity increased by 16.7% and test ordering increased even more. Between 2004 and 2008, 20 patient problems that accounted for <20% of all problems managed by family physicians/GPs were responsible for 73% of growth in pathology testing, preventive health interventions accounted for 32% of this pathology test growth, and management of 3 chronic diseases (diabetes, hypertension, and lipid disorders) accounted for a further 27% of pathology test growth.<sup>1,2</sup>

In the United Kingdom, the quality and outcomes framework offered financial rewards to GPs for more intensive monitoring of patients, and its introduction was associated in 2002 to 2005 with a 20% increase in laboratory tests and from 2005 to 2009 a 24.2% increase in tests, mainly due to testing more patients than more tests/patient. The largest increases were in fecal occult blood (121%), C-reactive protein (86%), hematinics (75%), immunoglobulins (73.4%), and serum iron testing (72.2%).<sup>3</sup> A review of the United Kingdom National Health Service estimated that 25% of all pathology tests ordered were unnecessary.<sup>4</sup>

In Calgary, Alberta, which has a large integrated laboratory system, the number of laboratory tests increased 6% to 8% annually between 2004 and 2014, whereas the annual population growth was 2.2%.<sup>5</sup> During 2005 to 2011, 125 million tests were processed, with a 24% increase/capita in chemistry tests, 10% increase/capita in microbiology, 7% increase/capita in anatomical pathology, and a 15% decrease/capita in cytopathology.<sup>6</sup> There is also a striking variability in test ordering by family physicians (Table 1). Two examples are Mindemark et al<sup>7</sup> who found test ordering by GPs across 8 counties in Sweden on average varied by a factor of 2.5, and for some tests by a factor of 8, and O'Kane et al<sup>8</sup> across 58 practices in Northern Ireland found that electrolyte tests ordered varied between 158 and 1056/1000 patients.

With such a rapid increase in laboratory testing volumes, identifying effective strategies to slow the rate down without affecting the quality of patient treatment is important to restrain health costs. Therefore, we wished to perform a systematic review of test ordering behavior by family physicians/GPs. We identified 3 systematic reviews: one of 70 randomized controlled trials (RCTs) of audit and feedback, which found 4 RCTs on test ordering behavior involving family physicians,<sup>12,13</sup> a systematic review of on-screen point-of-care computer reminders, which identified 3 studies of test ordering in primary care,<sup>14</sup> and a systematic review of laboratory test ordering with 109 RCTs and nonrandomized studies, which also identified only 4 RCTs of test ordering practices.<sup>15</sup> Thus, the purpose of this systematic review and meta-analysis is to identify all published RCTs that educated family physicians about test utilization and assess whether studies succeeded, which planned to (a) increase desired testing, (b) decrease undesired testing, and (c) decrease variability among physicians.

## Methods

### Search Strategy

We searched the following databases using predetermined search strategies discussed between the librarian and the principal and coinvestigators (Figure 1): MEDLINE (1946-February 2015), EMBASE (1980-February 2015), EBM Reviews (1980-February 2015; Cochrane Database of Systematic Reviews, ACP Journal Club, Database of Abstracts of Reviews of Effects, Cochrane Central Register of Controlled Trials, Cochrane Methodology Register, Health Technology Assessment, NHS Economic Evaluation Database), PubMed (1966-February 2015), PubMed Central (1900-February 2015), Scopus (1960-February 2015), Web of Science (1900-February 2015), and CINAHL (1982-February 2015). No limits on publication date were applied; the search included studies in all languages and from all countries. All included studies were entered in the PubMed Single Citation Matcher on October 1, 2015, and all references to these studies followed up to identify any additional relevant studies.

### Searching Other Resources

Reference lists of the included studies were searched to identify additional potentially relevant studies. Studies in systematic reviews of health maintenance and screening interventions; physician education, on-screen, telephone, and paper reminders; audit and feedback; computerized clinical decision support systems; and pathology test use were searched for relevant RCTs. We identified 23 reviews of related areas and searched their reference lists. Experts in the field (ie, laboratory directors and managers) were consulted to identify additional unpublished studies or studies in press.

### Inclusion Criteria

Inclusion criteria were all RCTs with an intervention to change family physicians' test ordering behavior.

### Exclusion Criteria

Exclusion criteria were studies that on review of the abstract met the inclusion criteria, but on reading the full text were not RCTs or in which the outcomes of family physicians were not separable from those of other physicians. We wished to identify a "pure intervention cohort" of family physicians so that later systematic reviewers could compare outcomes for other professional groups such as diabetologists or nurse practitioners.

### Study Assessment and Data Entry

All titles and abstracts were independently assessed by 2 authors for inclusion, and data were independently entered.

**Table 1.** Examples of Variability in Testing Between Physicians and Between Jurisdictions.

Author, Date, Country	Practice Setting	Metric of Comparison*	Results
Britt et al, 2008, <sup>1</sup> Bayram et al, 2009, <sup>2</sup> Australia	Australian Department of Health and Ageing Study	Comparison of GP test ordering to guidelines	Aligned with guidelines: 75.5% for lipid disorders, 71.7% for weakness/tiredness, 72% for type 2 diabetes, 65% for hypertension, 50.9% for overweight/obesity, and 24.3% for health checks Conclusion: The guidelines that advise family physicians about optimum test ordering often are not designed by or for GPs, length is a barrier, information about optimum test ordering behavior and frequency is either not present or difficult to locate, and advice about optimum test ordering and frequency is limited for the patients with multiple chronic diseases who constitute a large part of family physicians' work)
Busby et al, 2013, <sup>3</sup> United Kingdom	United Kingdom General Practice Database, 13 regions (660 000 tests recorded in 230 000 person-years of follow-up, 2005-2009)	Tests/10 000 person-years	Largest increases in tests 2005-2009: Fecal occult blood 121% (attributed to introduction of National Bowel Cancer Screening Program); CRP 86% (attributed to new clinical guidelines for rheumatoid arthritis); hematinics 75% Between-regions standard deviations: Plasma viscosity 3.14; cardiac enzymes 2.01; blood trace elements and vitamins 1.25; creatinine phosphokinase 0.93. For plasma viscosity, there were no tests in 2 regions (East of England and Southeast Coast) but 770/10 000 person-years in the Southwest, and for some regions rates of testing for plasma viscosity, cardiac enzymes, blood trace elements, and vitamins were 3 times the national average Conclusion: "Much regional variability remained unexplained"
Mindemark et al, 2010, <sup>7</sup> Sweden	223 primary health-care centers, 8 counties (2 177 973 patients)	Number of tests/1000 inhabitants/year	Test numbers varied by average factor of 2.5 between counties and ranged by factor of 1.6 to 8.8 depending on test
O'Kane et al, 2011, <sup>8</sup> Northern Ireland	58 practices (284 609 patients)	Median number of tests/1000 patients and range	Electrolytes (median 451; range 158-1065); liver profile (386; 146-1084); lipid profile (282; 131-813); thyroid profile (202; 98-583); immunoglobulins (2.5; 0.5-13); PSA (22; 7-143) Per diabetic patient: HbA <sub>1c</sub> (1.8; 0.9-3.4); albumin/creatinine ratio 1.3 (0.5-4.7)
Salinas et al, 2011, <sup>9</sup> Spain	Valencia, 8 health districts (2 011 475 patients)	Tests/1000 inhabitants comparing lowest and highest districts	For pairs of related tests, the ratio of ordering one or both tests varied between districts: Aspartate amino transferase/alanine amino transferase 0.246 to 1.000; urea/creatinine 0.198 to 0.918; Free T4/TSH 0.255 to 1.000
Smellie et al, 2002, <sup>10</sup> United Kingdom	22 general practices in 1 district (population 165 000)	Spearman correlation coefficients for 28 tests for practices in bottom and top fifth of test ordering activity	For upper and lower fifths of practices: No differences in test ordering by average patient age in practice, women age 15-44, Townsend deprivation score for practice, or number of GPs/practice

(continued)

**Table 1.** (continued)

Author, Date, Country	Practice Setting	Metric of Comparison*	Results
Smellie et al, 2000, <sup>11</sup> United Kingdom	22 general practices in 1 district (population 165 000)	Highest and lowest decile of test ordering practices	Conclusion: "The large differences observed in general practice pathology requesting probably result mostly from individual variation in clinical practice" Test ordering varied by median 700% between lowest and highest decile of practices

Abbreviations: CRP, C-reactive protein; GP, general practitioner; HbA<sub>1c</sub>, glycated hemoglobin; PSA, prostate-specific antigen; T4/TSH, thyroxine/thyroid-stimulating hormone.

\* The metric of comparison differed widely between studies and could not be brought to a common metric.

("lab\* test\*" OR "clinical lab\* test\*") AND (frequency OR rate OR regularity OR volume\*) AND (cost\* OR "cost analysis" OR "cost effective\*" OR "cost benefit analysis" OR expens\* OR expenditure\* OR "health expenditure\*") AND ("family physician\*" OR "family practice" OR "family practitioner\*" OR "general practice" OR "general practitioner\*")

("lab\* test\*" OR "clinical lab\* test\*") AND ("test ordering" OR requisition OR "test utilization" OR "test volume\*") AND ("family physician\*" OR "family practice" OR "family practitioner\*" OR "general practice" OR "general practitioner\*")

("lab\* test\*" OR "clinical lab\* test\*") AND (cost\* OR "cost analysis" OR "cost effective\*" OR "cost benefit analysis" OR expens\* OR expenditure\* OR "health expenditure\*") AND ("test ordering" OR requisition OR "test utilization" OR "test volume\*") AND ("family physician\*" OR "family practice" OR "family practitioner\*" OR "general practice" OR "general practitioner\*")

("lab\* test\*" OR "clinical lab\* test\*") AND reduc\* AND (cost\* OR "cost analysis" OR "cost effective\*" OR "cost benefit analysis" OR expens\* OR expenditure\* OR "health expenditure\*") AND ("test ordering" OR requisition OR "test utilization") AND ("family physician\*" OR "family practice" OR "family practitioner\*" OR "general practice" OR "general practitioner\*")

**Figure 1.** Literature search strategy.

### Classification of Interventions

Kobewka et al<sup>15</sup> in 2014 performed a systematic review of the effect of education, audit, and feedback on physicians' laboratory test ordering but only identified 4 RCTs about family physicians' test ordering, and nearly all of the RCTs they found were of hospital-based test ordering. To enable comparison to the study by Kobewka et al,<sup>15</sup> we adopted their classification of interventions: educational (teaching appropriate test ordering guidelines), audit and feedback (physicians were presented with their test utilization results compared to a previous period or to peers), system-based interventions (order form modifications, computer clinical decision support systems), and incentives.

### Data Extraction and Risk of Bias Assessment

Data were independently extracted by 2 reviewers and discrepancies solved by discussion or referral to a third reviewer. Risk of bias was assessed using the methodology of the *Cochrane Handbook*.<sup>16,17</sup>

### Data Analysis

Because there was marked heterogeneity in populations, practice settings, comparators, numbers and types of tests assessed,

and outcome measures, a meta-analysis was performed only within groups of similar tests (eg, cholesterol). Studies reported either percentage change or total change in test numbers or both, and we modified the approach by Kobewka et al<sup>15</sup> and for a simple meta-analysis appropriate to the data computed (tests ordered at follow-up) minus (tests ordered at baseline) for each of the intervention group minus the comparator group.

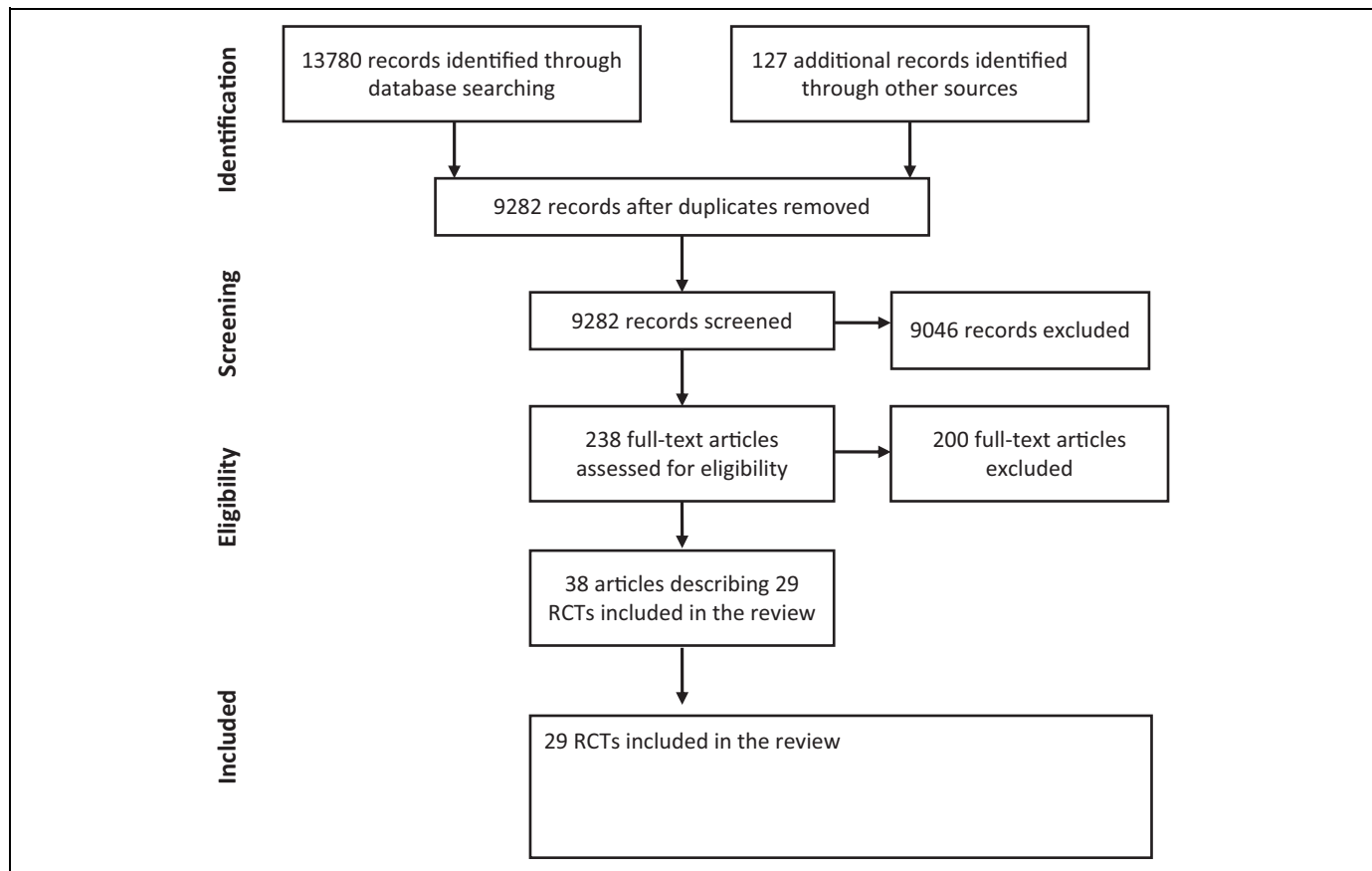
## Results

### Search

The searches excluding duplicates identified 9282 titles and abstracts, of which 238 were read in full text and 29 RCTs were included in this review (Figure 2).

### Description of Studies

The intervention in 10 studies was to reduce unnecessary testing<sup>18-30,32,33</sup> and in the other 19 studies was to increase the numbers of tests to improve screening. There were 7 studies from each of the United States and the Netherlands, 5 from the United Kingdom, 3 from Canada, 2 each from Australia, Norway, and Belgium, and 1 from New Zealand. The studies that reported data included collectively 4111 physicians and



**Figure 2.** Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow sheet of assessment of studies. Interventions to educate family physicians to change test ordering: systematic review.

175 563 patients (2 studies did not report the number of physicians,<sup>34,35</sup> and 8 studies did not report the number of patients,<sup>20,22,23,28-30,32-34,35</sup> and those numbers were also not available in related papers). There were 20 studies<sup>36-55</sup> in which outcomes for family physicians were not separable from those of other professional groups, and these were excluded from this review.

### Risk of Bias

The Cochrane Collaboration is the most authoritative method of analyzing risk of bias in RCTs. The *Cochrane Handbook*<sup>17</sup> asks authors of systematic reviews to independently search the text of each RCT and copy verbatim how the author describes the methods used in order to provide a transparent and reproducible method of recording risk of bias. Assessment of the risks of bias in an RCT is the key information in deciding whether the results of the trial are trustworthy and can be acted upon. The *Handbook*<sup>17</sup> assesses the risk of bias as low, unclear, or high for 6 key aspects of study execution (randomization, concealment of randomization from the researchers, blinding of participants and personnel, blinding of outcome assessors, attrition, and selective reporting of results). If authors provide no information for the above risk of bias categories that could

place their study at either low or high risk of bias, the risk of that item is assessed as “unclear” (*Handbook*).<sup>17</sup> For example, for randomization, the unclear designation most frequently occurs when the authors only say that physicians or patients were “randomly assigned” without stating a strong randomization method as defined by the *Handbook*.<sup>17</sup> The unclear category thus includes studies with data that are unclear because the authors did not perform the maneuver to reduce the risk of bias or did not report it or both.

### Results of the Risk of Bias Assessments

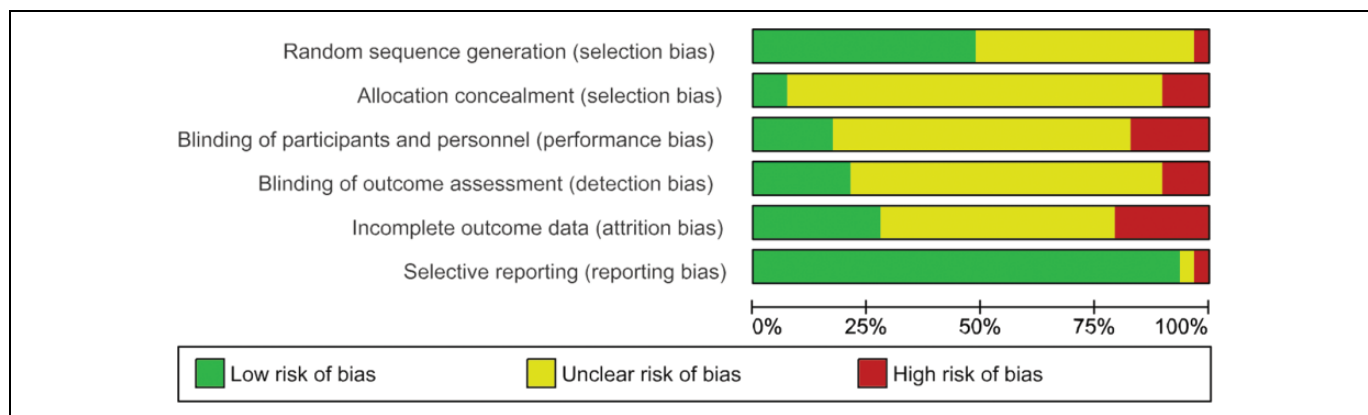
In Table 2, we present an overview of the risks of bias for the 6 items of research design, a sensitivity analysis identifying 6 RCTs at lowest risk of bias, and whether studies performed a power computation, an intention-to-treat analysis, and corrected for clustering in cluster randomized trial (C-RCTs; Figure 3).

- (1) Overview of the risk of bias assessments: Only 48.5% of studies were at low risk of bias from randomization (they used a strong method of randomization such as by computer), 7% from concealment of allocation from the researchers, 17% from blinding of participants and personnel, 21% from blinding of outcome

**Table 2.** Results of the Risk of Bias Assessments of 29 Included RCTs.

Risk of Bias Item	Low Risk, %	Unclear, %	High Risk, %	Problems and Overall Assessment
Randomization	<b>48.5</b>	<b>48.5</b>	3	Only 48.5% of the RCTs used a strong method of randomization such as computer randomization
Concealment	7	<b>83</b>	10	In only 7% of the RCTs was it not possible for the researchers to know to which group participants were assigned
Blinding of participants and personnel	17	<b>66</b>	17	Rates of blinding of personnel are only 17%, and this is an important defect when one is using techniques such as education and feedback
Blinding of outcome assessors	21	<b>69</b>	10	Only 21% of the outcome assessors were blind as to which study arm the participants were in
Attrition	27.5	<b>52</b>	<b>20.5</b>	Only 8 RCTs were assessed as at low risk for attrition. <sup>18,19,26,27,56-61</sup> NB: Three of these studies <sup>27,58,59</sup> had high attrition rates, but the authors conducted intention-to-treat analyses (which assume that those who did not complete the trial failed to benefit from the intervention). Intention-to-treat analyses provide a conservative estimate of the effect of the intervention, and the studies were thus assessed at low risk from attrition. Six RCTs were assessed at high risk of bias for attrition. The frequency of attrition across studies and the lack of clarity about attrition in these studies are major causes of weakness of this entire research enterprise
Selective reporting of results	<b>93</b>	3.5	3.5	A 93% rate of full reporting is excellent, meaning no “cherry picking” by presenting only the better results
Summary of Risk of Bias				
Summary for all 29 RCTs			Weaknesses in execution across 5 of 6 items of study execution mandate caution in interpreting results of studies at risk of bias	
Six studies with a strong method of randomization and minimal attrition (which are the 2 key aspects of research design)			We can have confidence in the results of these 6 RCTs: Baker et al (no change) <sup>26</sup> Buntinx et al (no change possible as 99% Paps satisfactory) <sup>56,57</sup> Holbrook et al (18% improvement) <sup>58</sup> Kenealy et al (8.2% to 16.3% change) <sup>59</sup> McClellan et al (0.1% and 3.8% change) <sup>60</sup> van Wyk et al (1.4 fewer tests/form $P = .003$ ) <sup>61</sup>	
Other Aspects of Study Design: Power, Intention-to-Treat Analysis, and Correcting for Clustering in C-RCTs				
Aspect of Study Design	Problems		Overall Assessment	
Power computation	12 studies made a power computation <sup>21,24,25,28-30,32,56-60,62-67</sup> but 17 did not		Of the 14 studies on lipids, 5 without a power computation showed no or minimal effects Of the 14 studies on diabetes tests, 6 without a power computation showed no or minimal change, and of the other studies, 4 showed no effect Of the 17 studies without a power computation, if they had inadequate sample size, they could likely report no effect, whereas an appropriate sample size might be associated with a significant effect	
Intention-to-treat analysis	Only 3 studies (Holbrook et al, <sup>58</sup> Kenealy et al, <sup>59</sup> and Bunting and Van Walraven <sup>27</sup> ) reported that that they conducted an intention-to-treat analysis		An intention-to treat analysis is a conservative approach to assessing results and treats dropouts as failures. Not conducting intention-to-treat analyses if the study has even modest attrition (eg <10%) may exaggerate results	
Correction for delivering interventions to clusters of physicians rather than individual physicians	27 studies were C-RCTs, and families were randomized in one and physicians in groups in the other 26. Only 13 studies used statistical techniques such as generalized estimating equations or multilevel analysis to estimate the effects of clustering on outcomes		In C-RCTs, the sample size is the number of clusters and not the number of participants. Failure to correct for clustering may overestimate the effect of the intervention	

Abbreviations: C-RCT, cluster randomized trial; RCT, randomized trial. Key data bolded.



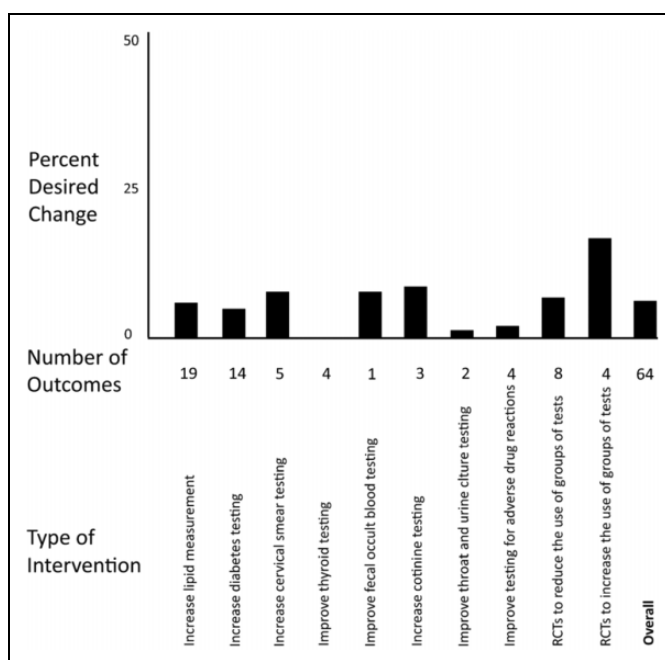
**Figure 3.** Risk of bias graph for 29 included studies.

assessors, 27.5% from attrition, but 93% did not selectively report results (and only 7% selectively reported results).

- (2) Sensitivity analysis identifying 6 RCTs at lowest risk of bias: The key aspect of study design and execution are studies with both a strong method of randomization and minimal attrition. We identified 6 studies in which we can have confidence in their results: Baker et al<sup>26</sup> (no change); Buntinx et al<sup>56,57</sup> (no change possible as 99% of Pap smears were satisfactory); Holbrook et al<sup>58</sup> (18% improvement); Kenealy et al<sup>59</sup> (8.2%-16.3% change); McClellan et al<sup>60</sup> (0.1%-3.8% change), and van Wyk et al<sup>68</sup> (1.4 fewer tests/form). The lack of clarity about whether a strong method of randomization was used, the lack of clarity about attrition, and the amount of attrition in the other 23 RCTs are major causes of weakness of this entire research enterprise. No study performed a differential attrition analysis (proving that those dropping out of the intervention and control groups were similar and thus unlikely to affect the results).
- (3) Identification of studies that performed a power computation, intention-to-treat analysis, and corrected for clustering in C-RCTs: Only 12 studies<sup>21,24,25,28-30,32,56-60,62-67</sup> made a power computation for needed sample size, 3 studies<sup>27,58,59</sup> made an intention-to-treat analysis, and only 13 studies used statistical techniques such as generalized estimating equations or multilevel analysis to estimate the effects of clustering on outcomes<sup>21,24,25,28-30,33,58,60,62,64,65-67,69,70,71,72,73</sup> (Table 2). The failure to correct the analyses in the other studies means that the conclusions need to be treated with considerable caution.

### Analysis of the Results

We analyzed studies according to 2 criteria of interest: (1) by the tests for which the researchers wished to optimize test ordering (Figure 4, Table 3) and (2) by the 4 intervention



**Figure 4.** Unweighted average desired changes in behavior for various tests and groups of tests. For interventions designed to increased test orders, an increase is considered a positive change. For interventions designed to decrease test orders, a decrease is considered a desired change. See Table 2 for an explanation of individual studies and associated statistical significance of individual studies.

strategies used (audit and feedback, system change [computerized reminders, computerized decision support systems, other reminders to physicians or patients], and practice system changes; Figure 5, Table 4).

- (1) Results analyzed according to tests of interest: We present a graphic overview (Figure 4), followed by further details about the studies in Table 3 (studies are listed beginning with the most frequent tests and then for each test in increasing order of magnitude of the intervention effect). Studies with interventions to increase testing for single illness included 14 for

**Table 3.** Interventions to Change Family Physicians' Test Ordering for Specific Tests.

Study, Author, Date, Country	No. of Physicians	No. of Patients	Intervention	Control	Test	Results and Significant Difference From Control
Interventions to increase measurement of lipids Eccles et al, 2000, <sup>69</sup> Eccles et al, 2002, <sup>62</sup> United Kingdom	270	2335	31 practices randomized to computerized angina guidelines	31 practices randomized to computerized asthma guidelines served as controls	"Cholesterol or other lipid"	Control 3.6% more than intervention (no significance statement)
van der Weijden et al, 1999, <sup>70</sup> the Netherlands O'Connor et al, 2009, <sup>71</sup> United States	32 123	3950 3703	Guidelines, 3-hour education session 1. Patient feedback (mailed information). 2. Physician feedback (prioritized lists of patients with recommended clinic actions). 3. Both patient and physician	No intervention No feedback	Cholesterol LDL	Control 1.4% more than intervention (NS) Compared to control: patient intervention 0.8% more ( $P < .05$ ); physician intervention 3% less ( $P < .01$ ); both interventions 1.1% less ( $P < .01$ )
Borgiel et al, 1999, <sup>63</sup> Canada	56	4401	Practice assessment report (PAR), Continuing medical education plan (CMEP) visit from mentor	Practice Assessment Report	Physician discussed cholesterol testing	Control 1% more (no significance statement)
Frank et al, 2004, <sup>61</sup> Australia	10	10 957	Reminders during consultation	"Usual care"	"Lipid tests"	Control 0.3% more (no significance statement) 0% difference
Hobbs et al, 1996, <sup>34</sup> United Kingdom	No N stated	No N stated	Computer asks for cardiovascular risk factors, family and medical history, and cholesterol level and then a coronary risk score is computed; advice on lipid medication, dosage, and dietary guidelines is offered	No intervention		
Sequist et al, 2005, <sup>73</sup> United States	194	2199 with coronary artery disease	Electronic medical record algorithm checked if patient had received care according to 4 evidence-based guideline reminders for coronary artery disease, reminders, medication, and problem lists displayed	"Usual care" (reminders suppressed)	Cholesterol	0% difference
Kiefe et al, 2001, <sup>72</sup> United States	70	2978	Chart review, specific feedback, achievable benchmarks	Chart review, specific feedback	Cholesterol; TG	Cholesterol 5% more ( $P = .13$ ); TGs 2% more ( $P = .22$ )
Frame et al, 1994, <sup>74</sup> United States	6	1008	Computerized reminders	Manual flowchart tracking system with reminders triggered by physician	Cholesterol	Intervention 8% more ( $P < .001$ )
Bonevski et al, 1999, <sup>64</sup> Australia	19	2784	Computerized continuing medical education program to increase screening, guidelines, feedback	Computerized continuing medical education program to increase screening, guidelines	Cholesterol	Intervention 12% more ( $P < .001$ )

(continued)



**Table 3.** (continued)

Study, Author, Date, Country	No. of Physicians	No. of Patients	Intervention	Control	Test	Results and Significant Difference From Control
Hetlevik et al, 2000, <sup>65</sup> Norway (cf also Hetlevik et al, 1998, <sup>66</sup> Hetlevik et al, 1999 <sup>67</sup> for details of study)	53	2014	GPs participated in seminar on risk intervention in diabetes and hypertension, computer-based decision support system	"Usual care"	Cholesterol	Intervention 15.4% more (no significance statement)
Holbrook et al, 2009, <sup>58</sup> Canada	46	511	Web-based COMPETE II diabetes tracker monitoring 13 variables and setting targets from Canadian and American Diabetes Association guidelines, with automated telephone system for patients, patients had access to Web tracker and received color-coded tracker page 4 × year to take to physician appointments, and monthly reminders for laboratory and physician visits	"Usual care"	LDL	Intervention 18% more (no significance statement)
van Wyk et al 2008, <sup>68</sup> the Netherlands	77	87 886	(1) Practices received alerts patients needed screening for dyslipidemia, (2) On-demand decision support	No intervention	Cholesterol; HDL; TG	Compared to control: alert group 39.5% more (RR: 1.76; 95% CI 1.41-2.20); on-demand decision support group 9.5% more (RR: 1.28, 95% CI 0.98-1.68; no significance statement)
Moher et al, 2001, <sup>75</sup> United Kingdom	100	1906	(1) Audit with feedback to primary health-care team; (2) Assistance setting up disease register and systematic recall of patients to GP; (3) Assistance setting up disease register and systematic recall to nurse-led clinic	No intervention	Cholesterol	Compared to baseline: audit group 25% more, GP recall 35% more, nurse recall 44% more (differences between groups $P = .001$ )
Interventions to increase diabetic screening and testing						
O'Connor et al, 2009, <sup>71</sup> United States	123	3703	(1) Patient feedback (mailed information), (2) Physician feedback (prioritized lists of patients with recommended clinic actions), (3) Both patient and physician	No feedback	HbA <sub>1c</sub>	(1) Patient intervention 0.6% more in control; (2) physician intervention 3.4% more in control; (3) both interventions 3.6% more in controls (All NS)
Frank et al, 2004, <sup>61</sup> Australia	10	10 957	Reminders during consultation	"Usual care"	"Diabetes screening"	0.3% more in control (no significance statement)
Kiefe et al, 2001, <sup>72</sup> United States	70	2978	Chart review, specific feedback, achievable benchmarks	Chart review, specific feedback	"Glucose"	5% more in intervention (no significance statement)
Eccles et al, 2000, <sup>69</sup> Eccles et al, 2002, <sup>62</sup> United Kingdom	270	2335	31 practices randomized to computerized angina guidelines	31 practices randomized to computerized asthma guidelines served as controls	"Blood glucose or HbA <sub>1c</sub> recorded"	Intervention 2% more (NS)

(continued)

Table 3. (continued)

Study, Author, Date, Country	No. of Physicians	No. of Patients	Intervention	Control	Test	Results and Significant Difference From Control
McClellan et al, 2003, <sup>60</sup> United States	473	22 971	Quality Improvement Organization feedback, patient education, clinical practice guidelines, practice aids to implement guidelines	No feedback	HbA <sub>1c</sub> ; quantitative urine protein	HbA <sub>1c</sub> : intervention 3.8% more ( $P = .02$ ); urine protein: intervention 0.1% more (NS)
Hetlevik et al, 2000, <sup>65</sup> Norway (cf also Hetlevik et al, 1998, <sup>66</sup> Hetlevik et al, 1999 <sup>67</sup> for details of study)	53	2014	GPs participated in seminar on risk intervention in diabetes and hypertension, computer-based decision support system	"Usual care"	HbA <sub>1c</sub>	Intervention 3.4% more (no significance statement)
Kenealy et al, 2005, <sup>59</sup> New Zealand	107	5628	(1) Patient intervention. (2) Computer intervention. (3) Patient + computer interventions	"Usual care"	"Glucose"	Compared to usual care: patient intervention 8.4% more ( $P < .001$ ); computer intervention 16.3% more ( $P = .001$ ); patient + computer intervention 8.2% more ( $P = .08$ ) LDL: intervention 18% more; albuminuria: intervention 18% more (no statement of significance)
Holbrook et al, 2009, <sup>58</sup> Canada	46	511	Web-based COMPETE II diabetes tracker monitoring 13 variables and setting targets from Canadian and American Diabetes Association guidelines, with automated telephone system for patients, patients had access to Web tracker and received color-coded tracker page 4 × year to take to physician appointments, and monthly reminders for laboratory and physician visits	"Usual care"	LDL; albuminuria	
Sequist et al, 2005, <sup>73</sup> United States	194	2199 with coronary artery disease	Electronic medical record algorithm checked if patient had received care according to 5 evidence-based guideline reminders for diabetes, reminders, medication, and problem lists displayed	"Usual care" (reminders suppressed)	Cholesterol for diabetics; HbA <sub>1c</sub>	Cholesterol: intervention 41% more ( $P < .001$ ); HbA <sub>1c</sub> : intervention 14% more ( $P = .29$ )
Cervical smears Winkens et al, 1995, <sup>20</sup> the Netherlands	79	Not stated	Feedback (group A: electrocardiography, endoscopy, cervical smears, allergy tests). Groups A and B served as each other's control	Feedback (group B: radiographic and ultrasonographic tests)	Pap smears	Intervention 27% fewer (which was the purpose of study; no significance stated)

(continued)

**Table 3.** (continued)

Study, Author, Date, Country	No. of Physicians	No. of Patients	Intervention	Control	Test	Results and Significant Difference From Control
Buntinx et al, 1992, <sup>56</sup> Buntinx et al, 1993, <sup>57</sup> Belgium	179	Not stated	(1) Comment on technical quality of each smear and why assessed as unsatisfactory if so assessed; (2) same as 1 + quality overview of slides GP submitted previous month; (3) Same as 1 + specific advice about deficiencies in technique + advice	No intervention (n not stated)	Satisfactory Pap smears	During 1-year study period, (1) <1% of Pap smears scored unsatisfactory; (2) many physicians submitted well-fixed smears at the baseline, leaving no room for improvement; (3) no statistically significant effects of intervention (to demonstrate improvement would require 739 physicians per intervention group); (4) Use of spatula + cytobrush increased from 33% to 66%
Frank et al, 2004, <sup>61</sup> Australia	10	10 957	Reminders during consultation	"Usual care"	Pap smears	Intervention 0.6% more (no significance statement)
Borgiel et al, 1999, <sup>63</sup> Canada	56	4401	Practice Assessment Report (PAR), continuing medical education plan (CMEP) visit from mentor	Practice Assessment Report	Pap smears	Intervention 5.3% more (no significance statement)
Frame et al, 1994, <sup>74</sup> United States	6	1008	Computerized reminders	Manual flowchart tracking system with reminders triggered by physician	Pap smears	Intervention 9% more ( $P = .001$ )
INR Claes et al, 2005, <sup>21</sup> Belgium	96	834	6-month retrospective analysis of INR monitoring. (1) Each group received education on oral anticoagulation, United Kingdom guidelines, anticoagulation files, and patient information booklets (groups A, B, C, and D). (2) Group B also received feedback every 2 months on their anticoagulation performance. (3) Group C determined INR with CoaguChek device in doctor's office or at patient's home. (4) Group D received Dawn AC computer-assisted advice for computing oral anticoagulation dosage. All groups received feedback every 2 months on performance of practice compared to whole set of practices	Control group (no intervention)	Time spent within 0.5 INR of target INR of 2.5 or 3.5	Education group: 13.5% more tests in range; education + feedback every 2 months: 10.5% more tests in range; education + use CoaguCheck machine in office: 7.5% more tests in range; education + Dawn C computer-assisted decision software: 5.5% more tests in range. All improved $P < .0001$ , but no significant differences between 4 physician intervention groups

(continued)

**Table 3.** (continued)

Study, Author, Date, Country	No. of Physicians	No. of Patients	Intervention	Control	Test	Results and Significant Difference From Control
van Wijk, 2001 et al. <sup>22</sup> van Wijk et al, 2002, <sup>23</sup> the Netherlands	62	No. of patients not stated	(1) BloodLink-Restricted computer-based decision support system; 22 practices, 30 GPs, and 12 742 laboratory requests). (2) BloodLink-Guideline (24 practices, 32 GPs, and 12 668 laboratory requests)	No control group	No. of INR tests ordered	BloodLink-Guideline ordered average 5.5 tests/form; BloodLink-Restricted ordered 6.9 tests/form ( $P = .003$ )
Thyroid tests Eccles et al, 2000, <sup>69</sup> Eccles et al, 2002, <sup>62</sup> United Kingdom	270	2335	31 practices randomized to computerized angina guidelines	31 practices randomized to computerized asthma guidelines served as controls	"Thyroid function"	0% difference
Fecal occult blood tests Frame et al, 1994, <sup>74</sup> United States	6	1008	Computerized reminders	Manual flowchart tracking system with reminders triggered by physician	Fecal occult blood tests	Intervention 15% more ( $P < .001$ )
Serum cotinine (to detect smoking) Moher et al, 2001, <sup>75</sup> United Kingdom	100	1906	(1) Audit with feedback to primary health-care team; (2) assistance setting up disease register and systematic recall of patients to GP; (3) assistance setting up disease register and systematic recall to nurse-led clinic	No intervention	Serum cotinine	Compared to baseline: audit group 5% more; GP recall 21% more, nurse recall 24% more (differences between groups $P = .001$ )
Throat swabs, urine cultures Flottorp et al, 2002, <sup>24</sup> Flottorp et al, 2003, <sup>25</sup> Norway	763	9887	(1) 592 practices received interventions to implement guidelines for urinary tract infection; (2) 61 practices received interventions to implement guidelines for sore throat. The 2 arms served as controls for each other	No control	"Laboratory tests" for UTI; throat swabs for sore throat	Throat swabs: no difference; UTI: intervention 5.1% fewer tests ( $P = .046$ )
Increase laboratory testing after prescribing angiotensin-converting enzyme (ACE) or ARB (angiotensin receptor blocker) Lafata et al, 2007, <sup>76</sup> United States	294	8325	Academic detailing + feedback + group problem solving to increase testing after dispensing ACE/ARBs, diuretics, or digoxin	No intervention control	Potassium, creatinine	Initial users: 3.3% intervention more tests for ACE/ARB ( $P < .01$ ); 4.9% more for diuretics ( $P < .01$ ); no difference digoxin; continuing users: 4.9% intervention more tests for ACE/ARB ( $P < .01$ ); 2.6% more for diuretics ( $P < .01$ ); no difference digoxin

(continued)

**Table 3.** (continued)

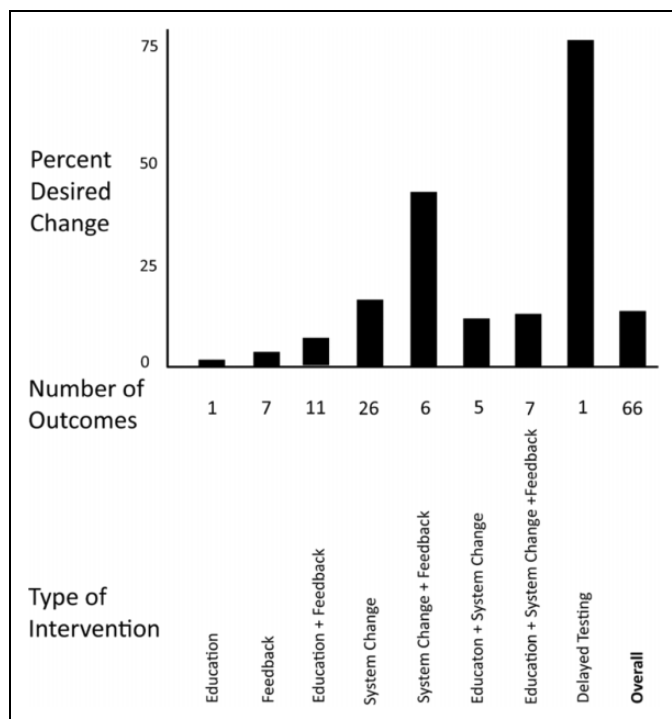
Study, Author, Date, Country	No. of Physicians	No. of Patients	Intervention	Control	Test	Results and Significant Difference From Control
RCTs to reduce use of groups of tests						
Baker et al, 2003, <sup>26</sup> United Kingdom (see below for second part of RCT to encourage lipid testing)	96	148 470	(1) Laboratory sent guidelines encouraging reduced urine tests; (2), feedback every 3 months x 1 year about test numbers; (3). Lead GP received data and convened meeting. 58 GPs in 17 practices received feedback on thyroid function, RF tests and urine cultures they ordered	No control	Changes in numbers of tests/1000 requested in either study group for any test.	No changes
Bunting and Van Walraven, 2004, <sup>27</sup> Canada	193 Drs who ordered most laboratory tests during 1 year	Intervention (3 943 000 visits); control (4 254 000 visits)	(1) Physicians visited individually up to 3 times by laboratory representatives over 2-year period, (2) educational material, (3) physician's personal laboratory	Physicians not provided with information about their own test use	No. of tests/visit	Intervention 7.9% fewer tests than control ( $p < .0001$ )
Verstappen et al, 2003, <sup>28</sup> Verstappen et al, 2004, <sup>29</sup> the Netherlands (cf Verstappen et al, 2004 <sup>30</sup> for cost data)	174	Not stated	To improve test ordering strategy: (1) feedback: education on guidelines; quality improvement sessions in small groups; (2) feedback only (arm A received intervention for cardiovascular, upper abdominal, and lower abdominal complaints; arm B received intervention for COPD/asthma, nonspecific complaints, and degenerative joint complaints; each arm served as control for other arm)	No control	Groups of tests	Arm A 12% fewer tests than arm B; arm B 5% fewer tests than arm A
Thomas et al, 2006, <sup>32</sup> United Kingdom	370	Not stated	4 senior laboratory clinicians assessed 9 laboratory tests as limited value for some patient subgroups. Practices received quarterly feedback with color graphs of their practice's test requesting rate. Half also received educational messages alongside the graphs, or brief educational reminders automatically added to test results sent to the practice, or both	No intervention	9 tests	Practices receiving enhanced feedback reduced orders by 13% for all 9 tests (with statistically significant reductions for autoantibody screen, FSH, TSH, vitamin B12); practices receiving brief educational reminders reduced orders by 11% for 8 tests (statistically significant reductions in CEA, TSH, B12)

(continued)

Table 3. (continued)

Study, Author, Date, Country	No. of Physicians	No. of Patients	Intervention	Control	Test	Results and Significant Difference From Control
Bindels et al, 2003, <sup>33</sup> the Netherlands	30	Not stated	GPs reviewed sample of 30 request forms they filled in earlier that year. An automated system displayed critical comments about nonadherence to guidelines for ordering diagnostic tests. Half the GPs received recommendations about cluster A diagnoses (7 diagnoses) and half about cluster B (9 diagnoses) and were blinded to the other diagnoses so learning effects could be assessed for the second cluster of diagnoses they assessed	No control	Tests on 30 laboratory request forms	17% reduction in number of tests in desired direction; 39% decrease in tests not in accordance with guidelines
Koch et al, 2009, <sup>18</sup> van Bokhoven et al, 2009, <sup>19</sup> the Netherlands	63	325	(1) Group 1: physicians instructed to order blood tests immediately if deemed appropriate and order either (a) limited set (Hgb, ESR, glucose, TSH) or (b) from complaint-specific list of 20 tests. (2) Group 2: same + receive systematically developed quality improvement strategy	Physicians asked to delay ordering any blood tests until 4 weeks later	No. of tests	26 GPs randomized to order blood tests immediately ordered tests on 146 of 158 patients; GPs randomized to order blood tests with 4-week delay instead ordered tests immediately on 27 of 138 patients. On the 325 patients testing established diagnoses in 11. Few patients in delay group reconsulted the GP within 4 weeks. Expanded fatigue-specific set of 13 tests resulted in more false positives than limited set of 4 tests
RCTs to increase use of groups of tests						
Baker et al, 2003, <sup>26</sup> United Kingdom	96	148 470	(1) Laboratory sent guidelines encouraging increased serum lipid tests; (2) feedback every 3 months × 1 year about test numbers; (3) lead GP received data and convened meeting; 38 GPs in 16 practices received feedback about lipid and plasma viscosity tests they ordered	No control	Changes in numbers of tests/1000 requested in either study group for any test	No changes
Smith et al, 2009, <sup>35</sup> United States	Not stated	961 patients were requested to obtain tests after new medications	(1) Medical record reminder to primary physician (with guidelines, recommended specific tests, and letter could send to patient), (2) automated voice message to patients, or (3) pharmacy team outreach (phone call from nurse then letter)	"Usual care"	AST, ALT, CBC, creatinine, potassium, sodium	Compared to usual care, medical record 26.1% more; automated voice message to patients 43.9% more; pharmacy outreach 59.6% more (no significance stated)

Abbreviations: ACE, angiotensin-converting enzyme; ALT, alanine transaminase; ARB, angiotensin receptor blocker; AST, aspartate transaminase; CBC, complete blood count; CEA, carcinoembryonic antigen; CI, confidence interval; COPD, chronic obstructive pulmonary disease; ESR, erythrocyte sedimentation rate; FSH, follicle-stimulating hormone; GP, general practitioner; HbA<sub>1c</sub>, glycated hemoglobin; HDL, high-density lipoprotein; INR, international normalized ratio; LDL, low-density lipoprotein; NS, not significant; RR, relative risk; TG, triglyceride; TSH, thyroid-stimulating hormone; UTI, urinary tract infection.



**Figure 5.** Unweighted averages of desired changes using different intervention strategies. For interventions designed to increased test orders, an increase is considered a positive change. For interventions designed to decrease test orders, a decrease is considered a desired change. See Table 3 for an explanation of individual studies and associated statistical significance of individual studies.

lipids, 14 for diabetes, 5 for cervical smears, 2 for international normalized ratio (INR), and 1 each for thyroid tests, fecal occult blood tests, serum cotinine (to detect smoking), throat swabs, testing after prescribing medications, and urine cultures. Six studies used interventions to decrease groups of tests and 2 to increase groups of tests (numbers of interventions add up to more than the total of 29 studies as some studies attempted to change more than 1 test). Unweighted averages for intervention effects are provided only for tests with >5 studies.

- (a) Lipids: Fourteen studies to increase lipid testing: (1) 5 resulted in slightly more testing in the control group, (2) 2 showed no difference between the intervention and control group, and (3) the others ranged from 5% to 44% more testing in the intervention group. Overall, the intervention group averaged 10.2% more tests ordered than the control group.
- (b) Diabetes tests: Fourteen studies to increase testing: (1) 2 resulted in slightly more testing in the control group and (2) the others ranged from 2% to 41% more testing in the intervention group. Overall, the intervention group averaged 8% more tests ordered than the control group.

- (c) Six studies to reduce use of groups of tests: (1) 1 found no decrease in the intervention group and (2) the others ranged from reductions of 5% to 17% of tests. In the unique study of patients with fatigue by Koch et al,<sup>18,19</sup> which compared immediate to delayed testing, family physicians permitted to test immediately ordered tests on 146 (92.4%) of 158 patients and those asked to delay a month ordered tests immediately on only 27 (19.5%) of 138 patients, a 72.9% reduction in the immediate testing. The entire set of tests established diagnoses in only 11 patients, and few patients in the delay group reconsulted the GP within 4 weeks. An expanded fatigue-specific set of 13 tests resulted in more false positives than a limited set of 4 tests. Overall, the average was 18% fewer tests in the intervention compared to the control group.
- (2) Results analyzed according to the type of intervention: The data are presented in a graphic overview (Figure 5), with more detail about the studies in Table 4 (studies are listed by the type of intervention and then for each intervention in increasing order of magnitude of the intervention effect). Unweighted averages for intervention effects are provided only for groups with >5 studies.
  - (a) Education: 1 RCT: There was a small increase (1.4%) in the control group.<sup>70</sup>
  - (b) Feedback: 3 RCTs: O'Connor et al<sup>71</sup> found mostly small increases in testing in the control group, Kiefe et al<sup>72</sup> found a 5% increase in testing in the intervention group, and Winkens et al<sup>20</sup> found a net desired 27% decrease in the number of Pap tests for the intervention group compared to the control group.
  - (c) Education and Feedback: 7 RCTs: (i) 2 found no changes: Baker et al<sup>26</sup> found no changes for any test (lipid, thyroid, and urine tests), Buntinx et al<sup>56,57</sup> found <1% of Paps were judged unsatisfactory so there was no room for improvement; (ii) 3 studies found changes <5%: Lafata et al<sup>76</sup> found no increase in follow-up testing after prescribing digoxin, a 3.3% increase in testing after prescribing angiotensin-converting enzyme/angiotensin receptor blockers, and a 4.9% increase after prescribing a diuretic, and Borgiel et al<sup>63</sup> found that the intervention arm that received continuing medical education and visits from a mentor over 3 years increased the number of Pap smears by 5.3% and decreased cholesterol tests by 1% compared to the less intensive physician assessment report intervention arm; (iii) 3 studies found changes >8%: Bunting and Van Walraven<sup>27</sup> in a unique study of 200 family physicians who ordered the most tests in a region found that the intervention produced change in the desired

**Table 4.** Effects on Test Ordering by Types of Intervention.

Author and Date	Tests	Percentage Difference
Education van der Weijden et al, 1999 <sup>70</sup>	Cholesterol	Control 1.4% more than intervention (NS)
Feedback O'Connor et al, 2009 <sup>71</sup>	LDL, HbA <sub>1c</sub> (compared to preintervention rates set at 100%)	LDL: compared to control, patient intervention 0.8% more ( $P < .05$ ); physician intervention 3% less ( $P < .01$ ); both interventions 1.1% less ( $P < .01$ ) HbA <sub>1c</sub> : (1) patient intervention 0.6% more in control; (2) physician intervention 3.4% more in control; (3) both interventions 3.6% more in control (all NS)
Kiefe et al, 2001 <sup>72</sup> Winkens et al, 1995 <sup>20</sup>	Glucose, triglycerides, cholesterol Paps	5% more in intervention (no significance statement) Intervention 27% fewer (which was the purpose of study; no significance stated)
Education and feedback Baker et al, 2003 <sup>26</sup> Lafata et al, 2007 <sup>76</sup>	Lipids, thyroid, urinalysis Potassium, creatinine	0 (no change in tests/1000 patients for any test) Initial users: 3.3% intervention more tests for ACE/ARB ( $P < .01$ ); 4.9% more for diuretics ( $P < .01$ ); no difference digoxin Continuing users: 4.9% intervention more tests for ACE/ARB ( $P < 0.01$ ); 2.6% more for diuretics ( $P < .01$ ); no difference digoxin
Buntinx 1992, <sup>56</sup> 1993 <sup>57</sup>	Paps	During 1 year study period, (1) <1% of Pap smears scored unsatisfactory (2) Many physicians submitted well-fixed smears at the baseline, leaving no room for improvement (3) No statistically significant effects of intervention (to demonstrate improvement would require 739 physicians per intervention group) (4) Use of spatula + cytobrush increased from 33% to 66%
Borgiel 1994 <sup>63</sup>	Cholesterol, cervical smears	Cholesterol intervention 1% less (no significance statement) Cervical smears intervention 5.3% more (no significance statement)
Bunting et al, 2004 <sup>27</sup>	All tests physicians ordered	Intervention 7.9% fewer tests than control ( $P < .0001$ ; results for intervention are a decrease and in desired direction)
Verstappen et al, 2003, <sup>28</sup> Verstappen et al, 2004, <sup>29</sup> Verstappen et al, 2004 <sup>30</sup>	Group A received education about problems involving 15 laboratory tests, group B received education about problems involving 10 laboratory tests. The physicians served as controls (without specific education about tests for the other group)	Arm A 12% fewer tests than arm B; arm B 5% fewer tests than arm A (both results in desired direction of a reduction in tests)
Thomas et al, 2006 <sup>32</sup>	9 laboratory tests assessed by laboratory as unnecessary	Practices receiving enhanced feedback reduced orders by 13% for all 9 tests (with statistically significant reductions for autoantibody screen, FSH, TSH, vitamin B12); practices receiving brief educational reminders reduced orders by 11% for 8 tests (statistically significant reductions CEA, TSH, B12). Both results in desired direction of a reduction in tests
System change Eccles et al, 2002, <sup>69</sup> Eccles et al, 2003 <sup>62</sup>	Glucose, TSH, Hgb, lipids for patients who consulted	"Cholesterol or other lipid": Control 3.6% more than intervention (no significance statement) "Blood glucose or HbA <sub>1c</sub> ": intervention 2% more (n.s.) "thyroid function": 0% difference
Frank et al, 2004 <sup>61</sup>	Percentage of preventive opportunities for Pap, "diabetes screening," "lipid screening"	Pap smears: intervention 0.6% more (no significance statement) Lipids: control 0.3% more (no significance statement) "Diabetes screening": 0.3% more in control (no significance statement)

(continued)



**Table 4.** (continued)

Author and Date	Tests	Percentage Difference
McClellan et al, 2003 <sup>60</sup>	HbA <sub>1C</sub> , quantitative urine protein	HbA <sub>1C</sub> : intervention 3.8% more ( $P = .02$ ) Urine protein: intervention 0.1% more (NS)
Frame et al, 1994 <sup>74</sup>	Fecal occult blood, Pap, cholesterol	Pap smears: intervention 9% more ( $P = .001$ ) Fecal occult blood: intervention 15% more ( $P < .001$ ) Cholesterol: intervention 8% more ( $P < .001$ )
van Wijk et al, 2001, <sup>22</sup> van Wijk et al, 2002 <sup>23</sup>	Control of INR within therapeutic limits using BloodLink-Guidelines (based on Dutch College of GPs guidelines)	BloodLink-Guideline ordered average 5.5 tests/form (14% reduction in direction of desired change) BloodLink-Restricted ordered 6.9 tests/form ( $P = .003$ )
Kenealy et al, 2005 <sup>59</sup>	Diabetes screening	Glucose compared to usual care: patient intervention 8.4% more ( $P < .001$ ); computer intervention 16.3% more ( $P = .001$ ); patient + computer intervention 8.2% more ( $P = .08$ )
Holbrook et al, 2009 <sup>58</sup>	No of tests HbA <sub>1C</sub> , LDL, albuminuria compared to guideline targets	Intervention LDL 18% more, HbA <sub>1C</sub> : 20% more, albuminuria 28% more than the control group (no significance statements)
Sequist et al, 2005 <sup>73</sup>	HbA <sub>1C</sub> , cholesterol	Diabetics: cholesterol: intervention 41% more ( $P < .001$ ); HbA <sub>1C</sub> intervention 14% more ( $P = .29$ ) Coronary artery disease: No significant change in annual cholesterol testing
Smith et al, 2009 <sup>35</sup>	Test to be performed if taking a medication: AST, ALT, CBC, creatinine, potassium, sodium	Compared to usual care: medical record 26.1% more; automated voice message to patients 43.9% more; pharmacy outreach 59.6% more (no significance stated)
van Wyk et al, 2008 <sup>68</sup>	Screening for dyslipidemia	Cholesterol, HDL, TG: compared to control: alert group 39.5% more (RR 1.76; 95% CI 1.41-2.20) On-demand decision support group 9.5% more (RR 1.28, 95% CI 0.98-1.68; no significance statement)
System change + feedback Moher et al, 2001 <sup>75</sup>	Cholesterol, cotinine (as a measure of tobacco use)	Cholesterol: compared to baseline: audit group 25% more; GP recall 35% more, nurse recall 44% more (differences between groups $P = .001$ ) Serum cotinine: compared to baseline: audit group 5% more; GP recall 21% more, nurse recall 24% more (differences between groups $P = .001$ )
Education + system change Hobbs et al, 1996 <sup>34</sup>	Cholesterol, TGs, HDL	0% (no increase in lipid tests, variation between practices remained)
Hetlevik et al, 1998, <sup>65</sup> Hetlevik et al, 1999, <sup>66</sup> Hetlevik et al, 2000 <sup>67</sup>	HbA <sub>1C</sub> , cholesterol	Cholesterol: intervention 15.4% more (no significance statement) HbA <sub>1C</sub> : intervention 3.4% more (no significance statement)
Bindels et al, 2001 <sup>33</sup>	Tests on 30 requests forms	17% reduction in number of tests in desired direction 39% decrease in tests not in accordance with guidelines
Education + system change + feedback Flottorp et al, 2002, <sup>24</sup> Flottorp et al, 2003 <sup>25</sup> Bonevski et al, 1999 <sup>64</sup> Claes et al, 2005 <sup>21</sup>	Throat swabs for streptococcal infection, "laboratory tests for urinary tract infection" Cholesterol Control of INR within therapeutic limits (all physician groups received education, anticoagulation guidelines, and patient education materials)	Throat swabs: no difference UTI tests: intervention 5.1% fewer tests ( $P = .046$ ) Cholesterol: intervention 12% more ( $P < .001$ ) Education group: 13.5% more tests in range; Education + feedback every 2 months: 10.5% more tests in range, 60%; Education + use of CoaguCheck machine in office: 7.5% more tests in range; Education + Dawn C computer-assisted decision software: 5.5% more tests in range. All improved $P < .0001$ , but no significant differences between 4 physician intervention groups

(continued)

**Table 4.** (continued)

Author and Date	Tests	Percentage Difference
Delaying testing Koch et al 2009, <sup>18</sup> van Bokhoven et al, 2009 <sup>19</sup>	91 GPs (9 withdrew, 19 included no patients, thus 63 participated), 325 patients with vague complaints	26 GPs randomized to order blood tests immediately ordered tests on 146 (92.4%) of 158 patients; GPs randomized to order blood tests with 4-week delay ordered tests immediately on 27 (19.5%) of 138 patients (a reduction in immediate testing of 72.9%). Testing established diagnoses in only 11 patients. An expanded fatigue-specific set of 13 tests resulted in more false positives than a limited set of 4 tests. Few patients in the delay group reconsulted the GP within 4 weeks

Abbreviations: ACE, angiotensin-converting enzyme; ARB, angiotensin receptor blocker; CBC, complete blood count; CEA, carcinoembryonic antigen; CI, confidence interval; FSH, follicle-stimulating hormone; GP, general practitioner; HbA<sub>1C</sub>, glycated hemoglobin; INR, international normalized ratio; LDL, low-density lipoprotein; NS, not significant; RR, relative risk; TSH, thyroid-stimulating hormone; UTI, urinary tract infection.

direction with the intervention group ordering 7.9% fewer tests/visit than the control group. Verstappen et al<sup>28-30</sup> found a desired 12% reduction in testing in a physician group asked to solve problems involving 15 laboratory tests and a 5% reduction in a group with problems involving 10 laboratory tests (cf also Verstappen).<sup>31</sup> Thomas et al<sup>32</sup> found a desired 13% reduction of tests in the enhanced feedback group for 9 tests the laboratory regarded as unnecessary and 11% in the group that received brief educational reminders. Overall, for 11 outcomes, the average increase in test ordering in the intervention group compared to the control group was 4.9% (converting the desired reductions for Bunting and Van Walraven, Verstappen et al, and Thomas et al to positive change).

- (d) System change: 10 RCTs: System change usually consisted of computer-assisted decision-making. (1) Three found minimal changes.<sup>60,61,62,69</sup> (2) Two studies found change >8%. Frame et al<sup>74</sup> found the intervention group ordered 15% more fecal occult blood tests, 9% more Pap smears, and 8% more cholesterol tests. van Wijk et al<sup>22,23</sup> found that physicians who used a computer system with guidelines ordered a desired 14% fewer INR tests than a computer system without the guidelines. (3) Two studies found change >15%: Kenealy et al<sup>59</sup> found 16.3% more eligible were screened for diabetes with a computer reminder, 8.4% with a patient reminder, and 8.2% with combined reminders compared to usual care. Holbrook et al<sup>58</sup> found that the intervention group increased testing for low-density lipoprotein by 18%, glycated hemoglobin (HbA<sub>1C</sub>) by 20%, and albuminuria by 28% more than the control group. (4) Three studies found changes 26% to 44%: Sequist et al<sup>73</sup> found a 41% increase in annual cholesterol testing for diabetics but no increase in

HbA<sub>1C</sub> and lipid testing for those with coronary artery disease. van Wyk et al<sup>68</sup> found that 39.5% more patients were screened for dyslipidemia with a computer alert, and 9.5% more with an on-demand computer-assisted decision support system the physician had to decide to use, compared to the control group (although screening increased 25.5% in the control group). Smith et al<sup>35</sup> found that for a group of follow-up tests requested to be obtained within 25 days of an intervention, 26.1% more were obtained using an electronic medical record, 43.9% more with automated voice messages to patients, and 59.6% more with a phone call from pharmacy compared to usual care. The unweighted average increase in testing for 26 outcomes in the intervention group compared to the control group was 14.9% (converting the desired reduction for van Wijk to positive change).

- (e) System + feedback: 1 RCT: Moher et al<sup>75</sup> found cholesterol screening increased 25% with audit, 35% with a facilitator identifying and recalling patients to clinic to see their GP, and 44% with recall to their nurse. Tobacco screening increased by 5%, 21%, and 24%, respectively (an average over 6 outcomes of 26%).
- (f) Education + system change: 3 RCTs: Hobbs et al<sup>34</sup> found no changes in lipids, Bindels et al<sup>33</sup> found a 17% desired decrease in 30 tests, and Hetlevik et al<sup>65-67</sup> found a 3.4% increase in HbA<sub>1C</sub> and a 15.4% increase in cholesterol tests compared to the control group. The average change for 7 outcomes was 6%.
- (g) Education + system change + feedback: 3 RCTs: Flottorp et al<sup>24,25</sup> found a 0.4% decrease in throat swabs in the intervention group and 5.1% fewer urine tests in the intervention group compared to the control groups. Bonevski et al<sup>64</sup> found a 12% increase in cholesterol testing in the intervention

group compared to the control group. Claes et al<sup>21</sup> found a 14% improvement in the percentage of time INR results were within 0.5 of the target range in the education group, 11% in the feedback group, 8% in the group that used the INR in-office test, and 8% in the group that used computer-assisted decision-making, compared to the control group. All were ( $P < .0001$ ) better than control, but there were no significant differences between the 4 physician intervention groups. For 7 outcomes, the average improvement in testing was 7.7%.

## Discussion and Conclusions

In this review of RCTs to change family physicians' laboratory test ordering, we found that although some studies achieved no change, the interventions generally produced changes in the desired direction, and some of the changes were very large (20%-40%).

- (1) How many studies are at low risk of bias and thus we can place confidence in them? The key aspect of study design and execution is studies with both a strong method of randomization and minimal attrition. We identified only 6 such studies in which we can have confidence in their results: Baker et al (no change),<sup>26</sup> Buntinx et al (no change possible as 99% of Pap smears were satisfactory),<sup>56,57</sup> Holbrook et al (18% improvement),<sup>58</sup> Kenealy et al (8.2%-16.3% change),<sup>59</sup> McClellan et al (0.1% and 3.8% change),<sup>60</sup> and van Wyk et al (1.4 fewer tests/form,  $P = .003$ ).<sup>68</sup> However, some studies without a strong method of randomization and with attrition achieved high change rates (eg, above 20%-40%), and although we should note their methodological problems, the studies clearly achieved worthwhile change.
- (2) How many studies focused specifically on increasing or decreasing testing rates? Only 6 studies were specifically designed to increase or decrease laboratory testing: Claes et al,<sup>21</sup> van Wijk et al (to reduce INR testing),<sup>22,23</sup> Bunting and Van Walraven (to decrease testing by the 193 physicians who ordered the most laboratory tests during 1 year),<sup>27</sup> Verstappen et al<sup>28,29,30</sup> and Bindels et al (to improve test ordering strategies),<sup>33</sup> and Koch et al and van Bokhoven 2009 (to reduce testing for vague complaints by delaying testing for 1 month).<sup>18,19</sup> These are the studies likely to be of most interest to laboratory directors.
- (3) Which tests were investigated? In the remaining studies, investigators were strongly focused on improving screening and monitoring chronic disease (14 RCTs testing lipids and 14 testing diabetes), with the next largest number of 6 RCTs aiming to reduce groups of heterogeneous tests and 4 to improve cervical smear testing. Surprisingly, there was only 1 RCT for each of

these areas of frequent testing: thyroid, throat swabs, urine, and fecal occult blood (Table 3). Within each of the groups with enough studies to draw conclusions, the range of improvements in testing was very wide.

- (4) Which interventions were tested? The most frequently tested intervention was system change (10 RCTs, average change 14.9%) and then education + feedback (7 RCTs, average change 4.9%). There were much smaller numbers testing other interventions, with 3 each on feedback, education + system change, and education + system change + feedback and 1 each on education and delayed testing, with the numbers in these latter groups too small to draw conclusions, so we do not know if these latter 4 combinations of interventions are effective in increasing testing.
- (5) Do we know why the interventions worked or not? Only 3 studies followed up with the physician and staff participants to assess how the RCTs had functioned and detected the sources of problems. Flottorp et al<sup>24,25</sup> conducted telephone interviews with 112 (93%) of the 120 of the practices and discussed reasons for variation between practices. They identified 3 problems: all relevant staff (such as practice assistants) participated in only 67% of the practices for the intervention (however, 89% of all GPs participated); 10% of practices spent no time discussing the guidelines and 52% spent <1 hour; only 38% had started a change process (but most said they needed more time) and 39% said they did not need to change their practice; and 13% had serious internal communication problems. The researchers themselves reported that it was difficult to run the project in 25% of the practices, 20% of the practices reported serious problems with the software installation, and 11% with the use of the software. Decision support software was available in only 2418 (48%) of 5031 sore throat and 703 (28%) of 2522 of urinary tract infection consultations. Hobbs et al<sup>34</sup> encountered many problems with the then available software. The computer program was not loadable onto a central file server in any of the practices so there was only 1 workstation per practice and physicians who wanted to participate had to go to that workstation and enter demographic and clinical data already in their practice computers. The 386 computers were very slow. Three practices were unable to record any data, and the data from another were lost in the post. The software was unable to import and export data successfully from and to the practice medical systems. Buntinx et al<sup>56,57</sup> asked family physicians if the feedback they received about their test ordering was meaningful and desirable. Those who received either a mailed comment or specific advice about their technique rated both types of feedback as 96% meaningful and desirable, whereas monthly overview reports on their tests or comparison to peers were rated lower at 74% to 78% meaningful and desirable.

- (6) Do we know why there is marked variability in test ordering between family physicians? A review identified 104 articles about factors that affect physicians' test ordering and found that test ordering was correlated with physician age, gender, specialization, geographic location, practice setting, belief systems, experience, knowledge, fear of malpractice litigation, physician regret about missed diagnoses, financial incentives, awareness of costs, and provision of written feedback.<sup>77</sup> A review of 38 studies of factors that may influence test ordering in patients with undiagnosed complaints in primary or secondary care identified 5 key factors: diagnostic, therapeutic and prognostic, patient-related, doctor-related, and policy- and organization-related factors.<sup>78</sup> None of the studies assessed in this current review explored why there is variability among physicians or intervened to specifically correct it (other than providing interventions to improve test ordering for all physicians). Smellie et al concluded that "The large differences observed in general practice pathology requesting probably result mostly from individual variation in clinical practice."<sup>10(p312)</sup> Variability between family physicians remains a key large unresolved problem. No insight was provided by the 29 studies in this review how to diminish variability between physicians.
- (7) Did studies build on previous research? Science usually progresses by improving the work of others and testing the next steps. No study explicitly built upon and improved the studies of others or recorded that they had interviewed the research team and health staff and patients who had participated in previous projects to find out the obstacles encountered and how to improve outcomes. There has been much discussion why some research projects in primary care falter, and it has been concluded that they falter if the physicians and staff are not interested, are too busy with patient care, already have a quality improvement project, or they think that a readymade research project is being imposed on them and there are no benefits for them. An alternative approach to improve participation and decrease attrition is to discover the key problems that family physicians in the practices are interested in and motivated to research and build the change projects from the ground up with their continuing involvement and advice rather than imposing a completed research design.<sup>79</sup> The skill is then to execute the project to the highest standards of research with attention to a strong method of randomization, minimizing attrition, and being present to motivate and solve problems as they arise.

## Future Research

The interventions used in these studies are appropriate and practical, but the execution of the research projects, data

analysis, and presentation of results require major improvement. Skilled trial coordinators and statisticians need to be involved in future trials from their inception. The apparently most effective interventions to increase rational testing need replicating and improving. They need to engage involved medical staff in planning the studies to be of direct interest to them in their practices. Careful attention to adherence to the protocol and manual, minimization of attrition, and ongoing engagement with participants during trials to detect obstacles to participation are essential.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## References

1. Britt H; Australian Association of Pathology Practices, Inc. *An Analysis of Pathology Test Use in Australia*. Sydney, Australia: Family Medicine Research Centre, University of Sydney; 2008.
2. Bayram C, Britt H, Miller G, Valenti L. *Evidence-Practice Gap in GP Pathology Test Ordering: A Comparison of BEACH Pathology Data and Recommended Testing*. Sydney, Australia: Family Medicine Research Centre, University of Sydney; 2009.
3. Busby J, Schroeder K, Woltersdorf W, et al. Temporal growth and geographic variation in the use of laboratory tests by NHS general practices: using routine data to identify research priorities. *Br J Gen Pract*. 2013;63:e256-e266. doi:10.3399/bjgp13X665224.
4. Report of the Second Phase of the Review of NHS Pathology Services in England. Chaired by Lord Carter of Coles. 2008. Website. <http://www.bnms.org.uk/professional-standards/professional-standards/report-of-the-second-phase-of-the-review-of-nhs-pathology-services-in-england.html>. Accessed February 17, 2016.
5. Naugler C. A perspective on laboratory utilization management in Canada. *Clin Chim Acta*. 2014;427:142-144.
6. Naugler C. Laboratory test use and primary care physician supply. *Can Fam Physician*. 2013;59:e240-e245.
7. Mindemark M, Wernroth L, Larsson A. Costly regional variations in primary health care test utilization in Sweden. *Scand J Clin Lab Invest*. 2010;70:164-170.
8. O'Kane MJ, Casey L, Lynch PL, McGowan N, Corey J. Clinical outcome indicators, disease prevalence and test request variability in primary care. *Ann Clin Biochem*. 2011;48(pt 2):155-158.
9. Salinas M, López-Garrigós M, Díaz J, et al. Regional variations in test requiring patterns of general practitioners in Spain. *Ups J Med Sci*. 2011;116:247-251.
10. Smellie WS, Galloway MJ, Chinn D, Gedling P. Is clinical practice variability the major reason for differences in pathology requesting patterns in general practice? *J Clin Pathol*. 2002;55: 312-314.

11. Smellie WSA, Galloway MJ, Chinn D. Benchmarking general practice use of pathology services: a model for monitoring change. *J Clin Pathol*. 2000;53:476-480.
12. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev*. 2012;6:CD000259.
13. Ivers NM, Grimshaw JM, Jamtvedt G, et al. Growing literature, stagnant science? Systematic review, meta-regression and cumulative analysis of audit and feedback interventions in health care. *J Gen Intern Med*. 2014;29:1534-1541.
14. Shojania KG, Jennings A, Mayhew A, Ramsay CR, Eccles MP, Grimshaw J. The effects of on-screen, point of care computer reminders on processes and outcomes of care. *Cochrane Database Syst Rev*. 2009;(3):CD001096.
15. Kobewka DM, Ronksley PE, McKay JA, Forster AJ, van Walraven C. Influence of educational, audit and feedback, system based, and incentive and penalty interventions to reduce laboratory test utilization: a systematic review. *Clin Chem Lab Med*. 2015;53:157-183.
16. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343. doi: <http://dx.doi.org/10.1136/bmj.d5928>. Published October 18, 2011. Accessed February 17, 2016.
17. Higgins JPT, Green S, eds; The Cochrane Collaboration. *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0*. 2011. Web site. [www.cochrane-handbook.org](http://www.cochrane-handbook.org). Accessed September 22, 2015. Updated March 2011.
18. Koch H, van Bokhoven MA, ter Riet G, et al. Ordering blood tests for patients with unexplained fatigue in general practice: what does it yield? results of the VAMPIRE trial. *Br J Gen Pract*. 2009;59:e93-e100.
19. van Bokhoven MA, Koch H, van der Weijden T, et al. The effect of watchful waiting compared to immediate test ordering instructions on general practitioners' blood test ordering behaviour for patients with unexplained complaints; a randomized clinical trial (ISRCTN55755886). *Implementation Sci*. 2012;7:29.
20. Winkens RA, Pop P, Bugter-Maessen AM, et al. Randomised controlled trial of routine individual feedback to improve rationality and reduce numbers of test requests. *Lancet*. 1995;345:498-502.
21. Claes N, Buntinx F, Vijgen J, et al. The Belgian improvement study on oral anticoagulation therapy: a randomized clinical trial. *Eur Heart J*. 2005;26:2159-2165.
22. van Wijk MA, van der Lei J, Mosseveld M, Bohnen AM, van Bommel JH. Assessment of decision support for blood test ordering in primary care. a randomized trial. *Ann Intern Med*. 2001;134:274-281.
23. van Wijk MA, van der Lei J, Mosseveld M, Bohnen AM, van Bommel JH. Compliance of general practitioners with a guideline-based decision support system for ordering blood tests. *Clin Chem*. 2002;48:55-60.
24. Flottorp S, Oxman AD, Havelsrud K, Treweek S, Herrin J. Cluster randomised controlled trial of tailored interventions to improve the management of urinary tract infections in women and sore throat. *BMJ*. 2002;325:367.
25. Flottorp S, Oxman AD. Identifying barriers and tailoring interventions to improve the management of urinary tract infections and sore throat: a pragmatic study using qualitative methods. *BMC Health Serv Res*. 2003;3:3.
26. 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.
27. Bunting PS, Van Walraven C. Effect of a controlled feedback intervention on laboratory test ordering by community physicians. *Clin Chem*. 2004;50:321-326.
28. Verstappen WH, van der Weijden T, Sijbrandij J, et al. Effect of a practice-based strategy on test ordering performance of primary care physicians. *JAMA*. 2003;289:2407-2411.
29. 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.
30. Verstappen WH, van Merode F, Grimshaw J, Dubois WI, Grol RP, van der Weijden T. Comparing cost effects of two quality strategies to improve test ordering in primary care: a randomized trial. *Int J Qual Health Care*. 2004;16:391-398.
31. Verstappen WH, ter Riet G, Dubois WI, Winkens R, Grol RP, van der Weijden T. Variation in test ordering behaviour of GPs: professional or context-related factors? *Fam Pract*. 2004;21:387-395.
32. Thomas RE, Croal BL, Ramsay C, Eccles M, Grimshaw J. Effect of enhanced feedback and brief educational reminder messages on laboratory test requesting in primary care: a cluster randomised trial. *Lancet*. 2006;367:1990-1996.
33. Bindels R, Hasman A, Kester AD, Talmon JL, De Clercq PA, Winkens RA. The efficacy of an automated feedback system for general practitioners. *Inform Prim Care*. 2003;11:69-74.
34. Hobbs FD, Delaney BC, Carson A, Kenkre JE. A prospective controlled trial of computerized decision support for lipid management in primary care. *Fam Pract*. 1996;13:133-137.
35. Smith DH, Feldstein AC, Perrin NA, et al. Improving laboratory monitoring of medications: an economic analysis alongside a clinical trial. *Am J Manag Care*. 2009;15:281-289.
36. Castello E, Deichmann RE, Horswell R, Friday KE. Improvements in diabetic care as measured by HbA1c after a physician education project. *Diabetes Care*. 1999;22:1612-1616.
37. Dowling PT, Alfonsi G, Brown MI, Culpepper L. An education program to reduce unnecessary laboratory tests by residents. *Acad Med*. 1989;64:410-412.
38. El-Kareh RE, Gandhi TK, Poon EG, et al. Actionable reminders did not improve performance over passive reminders for overdue tests in the primary care setting. *J Am Med Inform Assoc*. 2011;18:160-163.
39. Feldstein AC, Smith DH, Perrin N, et al. Improved therapeutic monitoring with several interventions: a randomized trial. *Arch Intern Med*. 2006;166:1848-1854.

40. Froom P, Barak M. Cessation of dipstick urinalysis reflex testing and physician ordering behavior. *Am J Clin Pathol.* 2012;137:486-489.
41. Girard A, Moreau-Gaudry A, Alpes Réseau P, Hilleret MN. Analysis of medical prescribing practices for hepatitis B serology tests. *Gastroenterol Clin Biol.* 2010;34:8-15.
42. Herrin J, Nicewander DA, Hollander PA, et al. Effectiveness of diabetes resource nurse case management and physician profiling in a fee-for-service setting: a cluster randomized trial. *Proc (Bayl Univ Med Cent).* 2006;19:95-102.
43. Lo HG, Matheny ME, Seger DL, Bates DW, Gandhi TK. Impact of non-interruptive medication laboratory monitoring alerts in ambulatory care. *J Am Med Inform Assoc.* 2009;16:66-71.
44. Lobach DF, Hammond WE. Computerized decision support based on a clinical practice guideline improves compliance with care standards. *Am J Med.* 1997;102:89-98.
45. Maclean CD, Gagnon M, Callas P, Littenberg B. The Vermont diabetes information system: a cluster randomized trial of a population based decision support system. *J Gen Intern Med.* 2009;24:1303-1310.
46. McPhee SJ, Bird JA, Fordham D, Rodnick JE, Osborn EH. Promoting cancer prevention activities by primary care physicians. Results of a randomized, controlled trial. *JAMA.* 1991;266:538-544.
47. Montori VM, Dinneen SF, Gorman CA, et al. The impact of planned care and a diabetes electronic management system on community-based diabetes care. *Diabetes Care.* 2002;25:1952-1957.
48. Ornstein SM, Garr DR, Jenkins RG, Rust PF, Arnon A. Computer-generated physician and patient reminders. Tools to improve population adherence to selected preventive services. *J Fam Pract.* 1991;32:82-90.
49. Peterson KA, Radosevich DM, O'Connor PJ, et al. Improving diabetes care in practice. Findings from the TRANSLATE trial. *Diabetes Care.* 2008;31:2238-2243.
50. Raebel MA, Lyons EE, Chester EA, et al. Improving laboratory monitoring at initiation of drug therapy in ambulatory care: a randomized trial. *Arch Intern Med.* 2005;165:2395-2401.
51. Rhyne RL, Gehlback SH. Effects of an educational feedback strategy on physician utilization of thyroid function panels. *J Fam Pract.* 1979;8:1003-1007.
52. Scholes D, Grothaus L, McLure J, et al. A randomized trial of strategies to increase chlamydia screening in young women. *Prev Med.* 2006;43:343-350.
53. Thomas JC, Moore A, Qualls PE. The effect on cost of medical care for patients treated with an automated clinical audit system. *J Med Syst.* 1983;7:307-313.
54. Sundaram V, Lazzeroni LC, Douglass LR, Sanders GD, Tempio P, Owens DK. A randomized trial of computer-based reminders and audit and feedback to improve HIV screening in a primary care setting. *Int J STD AIDS.* 2009;20:527-533.
55. Turner RC, Peden JG Jr, O'Brien K. Patient-carried card prompts vs computer-generated prompts to remind private practice physicians to perform health maintenance procedures. *Ach Intern Med.* 1994;154:1957-1960.
56. Buntinx F, Knottnerus JA, Crebolder HF, Essed GG. Reactions of doctors to various forms of feedback designed to improve the sampling quality of cervical smears. *Qual Assur Health Care.* 1992;4:161-166.
57. Buntinx F, Knottnerus JA, Crebolder HFJM, Seegers T, Essed GGM, Schouten H. Does feedback improve the quality of cervical smears? A randomized controlled trial. *Br J Gen Pract.* 1993;43:194-198.
58. Holbrook A, Thabane L, Keshavjee K, et al. Individualized electronic decision support and reminders to improve diabetes care in the community: COMPETE II randomized trial. *CMAJ.* 2009;181(1-2):37-44.
59. Kenealy T, Arroll B, Petrie KJ. Patients and computers as reminders to screen for diabetes in family practice. Randomized-controlled trial. *J Gen Intern Med.* 2005;20:916-921.
60. McClellan WM, Millman L, Presley R, Couzens J, Flanders WD. Improved diabetes care by primary care physicians: results of a group-randomized evaluation of the Medicare Health Care Quality Improvement Program (HCQIP). *J Clin Epidemiol.* 2003;56:1210-1217.
61. Frank O, Litt J, Beilby J. Opportunistic electronic reminders. Improving performance of preventive care in general practice. *Aust Fam Physician.* 2004;33(1-2):87-90.
62. Eccles M, McColl E, Steen N, et al. Effect of computerised evidence based guidelines on management of asthma and angina in adults in primary care: cluster randomised controlled trial. *BMJ.* 2002;325:941.
63. Borgiel AEM, Williams JI, Davis DA, et al. Evaluating the effectiveness of 2 educational interventions in family practice. *Can Med Assoc J.* 1999;161:965-970.
64. Bonevski B, Sanson-Fisher RW, Campbell E, Carruthers A, Reid ALA, Ireland M. Randomized controlled trial of a computer strategy to increase general practitioner preventive care. *Prev Med.* 1999;29(6 pt 1):478-486.
65. Hetlevik I, Holmen J, Kruger O, Kristensen P, Iversen H, Furuseth K. Implementing clinical guidelines in the treatment of diabetes mellitus in general practice. evaluation of effort, process, and patient outcome related to implementation of a computer-based decision support system. *Int J Technol Assess Health Care.* 2000;16:210-227.
66. Hetlevik I, Holmen J, Kruger O, Kristensen P, Iversen H. Implementing clinical guidelines in the treatment of hypertension in general practice. *Blood Press.* 1998;7(5-6):270-276.
67. Hetlevik I, Holmen J, Kruger O. Implementing clinical guidelines in the treatment of hypertension in general practice. evaluation of patient outcome related to implementation of a computer-based clinical decision support system. *Scand J Prim Health Care.* 1999;17:35-40.
68. van Wyk JT, van Wijk MAM, Sturkenboom MCJM, Mosseveld M, Moorman PW, van der Lei J. Electronic alerts versus on-demand decision support to improve dyslipidemia treatment. *Circulation.* 2008;117:371-378.
69. Eccles M, Grimshaw J, Steen N, et al. The design and analysis of a randomized controlled trial to evaluate computerized decision

- support in primary care: The COGENT study. *Fam Pract.* 2000; 17:180-186.
70. van der Weijden T, Grol RP, Knottnerus JA. Feasibility of a national cholesterol guideline in daily practice. A randomized controlled trial in 20 general practices. *Int J Qual Health Care.* 1999;11:131-137.
71. O'Connor PJ, Sperl-Hillen J, Johnson PE, Rush WA, Crain AL. Customized feedback to patients and providers failed to improve safety or quality of diabetes care: a randomized trial. *Diabetes Care.* 2009;32:1158-1163.
72. Kiefe CI, Allison JJ, Williams OD, Person SD, Weaver MT, Weissman NW. Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. *JAMA.* 2001;285:2871-2879.
73. Sequist TD, Gandhi TK, Karson AS, et al. A randomized trial of electronic clinical reminders to improve quality of care for diabetes and coronary artery disease. *J Am Med Inform Assoc.* 2005; 12:431-437.
74. Frame PS, Zimmer JG, Werth PL, Hall WJ, Eberly SW. Computer-based vs manual health maintenance tracking. A controlled trial. *Arch Fam Med.* 1994;3:581-588.
75. Moher M, Yudkin P, Wright L, et al. Cluster randomised controlled trial to compare three methods of promoting secondary prevention of coronary heart disease in primary care. *BMJ.* 2001;322:1338.
76. Lafata JE, Gunter MJ, Hsu J, et al. Academic detailing to improve laboratory testing among outpatient medication users. *Med Care.* 2007;45:966-972.
77. Sood R, Sood A, Ghosh AK. Non-evidence-based variables affecting physicians' test-ordering tendencies: a systematic review. *Neth J Med.* 2007;65:167-177.
78. Whiting P, Toerien M, de Salis I, et al. A review identifies and classifies reasons for ordering diagnostic tests. *J Clin Epidemiol.* 2007;60:981-989.
79. Westfall JM, Nearing K, Felzien M, et al. Researching together: a CTSA partnership of academicians and communities for translation. *Clin Transl Sci.* 2013;6:356-362.