



# Effective Health Care Program

## Benefits and Harms of Routine Preoperative Testing: Comparative Effectiveness

### Executive Summary

#### Introduction

Traditionally, preoperative testing has been part of the preoperative care process to inform patient selection by determining fitness for anesthesia and identifying patients at high risk for perioperative complications. The American Society of Anesthesiologists (ASA) defines routine preoperative tests as those done in the absence of any specific clinical indication or purpose; they typically include a panel of blood tests, urine tests, chest radiography, and electrocardiogram (ECG).<sup>1,2</sup> These tests are performed to find latent abnormalities, such as anemia or silent heart disease, that could impact how, when, or whether the planned surgical procedure and concomitant anesthesia are performed. Many hospitals have instituted protocols to perform a series of laboratory tests prior to any operative procedure under the assumption that this information will enhance safety for surgical patients and reduce liability for adverse events.<sup>2</sup> During the past three decades, routine preoperative testing has been challenged by several academic publications with concerns about the sizable cost of testing, overtesting, the consequences of false-positive tests (leading to unnecessary workups and treatments), and the unknown benefit to patients.<sup>3-8</sup> In addition to increasing

#### Effective Health Care Program

The Effective Health Care Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its Comparative Effectiveness Reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders, including consumers.

The full report and this summary are available at [www.effectivehealthcare.ahrq.gov/reports/final.cfm](http://www.effectivehealthcare.ahrq.gov/reports/final.cfm).

the cost of surgical care,<sup>2</sup> nonselective preoperative testing may result in false-positive or borderline results (in the absence of clinical indication), which require further investigation. Additional



Agency for Healthcare Research and Quality

Advancing Excellence in Health Care • [www.ahrq.gov](http://www.ahrq.gov)

Effective  
Health Care

investigation may cause unnecessary psychological and economic burdens, postponement of surgery, and even morbidity and mortality (e.g., complications due to unnecessary biopsies performed to follow up false-positive laboratory tests).<sup>2</sup> As all routine testing does, preoperative testing will find some abnormal test results that will lead to new diagnoses (such as previously undetected lung cancer), but it is unclear whether the benefits accrued from responses to true-positive tests outweigh the harms of false-positive preoperative tests and, if there is a net benefit, how this benefit compares with the resource utilization required for testing.

## **Considerations for Evaluation of Preoperative Testing**

### **Alternative Testing Strategies**

There is no common terminology among anesthesiologists and surgeons regarding the alternative preoperative testing strategies. For this review, we define the three main alternatives as follows: (1) routine preoperative testing, in which the tests of interest are conducted in all patients undergoing a given procedure, regardless of medical history or other patient features; (2) per-protocol preoperative testing, in which the tests of interest are conducted in a subset of patients undergoing a given procedure, such as ECG only in patients aged  $\geq 50$  years or hemoglobin only in premenopausal women; (3) ad hoc, or elective, testing, in which preoperative testing is done at the discretion of the clinician doing a preoperative evaluation, based on patient history or physical examination (H&P) findings. No tests are done routinely or based on any protocol.

### **Preoperative Tests**

There are many preoperative tests that can be ordered for a patient to determine fitness for surgery and anesthesia. Routine tests are those that may be of value to reduce the risk of procedural complications but are not directly related to the planned procedure. The specific tests under review here include hematologic, metabolic, and organ function blood tests; hemostasis tests; urinalysis; chest radiography (and related tests); ECG (and related tests); and pregnancy tests. These tests may be done alone (e.g., only a pregnancy test) or as part of a panel of tests.

### **Patient and Procedure Heterogeneity**

Patients undergoing surgery show considerable variation in demographic characteristics, underlying health and comorbidities, indications for surgery, specific surgery planned, type of anesthesia planned (e.g., general vs. spinal anesthesia), and other factors. Differences among these

factors may result in differences in the benefits of finding abnormalities (e.g., anemia) and in the potential harms of testing (e.g., delayed surgery or unnecessary colonoscopy). Therefore, it is important to look not only at the benefits and harms of preoperative testing in general, but also at specific patient and intervention (surgery-related) factors that might change the balance between the benefits and harms: namely, the risk of the surgical procedure, type of anesthesia planned, indication for surgery, comorbidities, and other patient characteristics.

The two most important factors are likely to be the risk of the procedure and the health status of the patient. The risk of procedural complications varies widely based on the type of surgery planned. It thus follows that the potential benefit of preoperative testing will vary based on the risk of complications related to the planned surgery. Although it has yet to be demonstrated, one could expect that some preoperative tests may be of greater value in predicting and ultimately reducing complications in higher rather than lower risk surgeries.

Similarly, one could expect that the risk of complications, and thus the potential value of preoperative testing, may be greater for patients with worse overall health status. The variation in the characteristics of patients undergoing surgery may lead to considerable differences in how abnormal preoperative test findings are handled, as well as their potential effect on surgery.

### **Clinician- and Setting-Based Differences**

Inefficiencies in the preoperative testing processes or failures in the handoff of test results among primary care physicians, surgeons, and anesthesiologists ultimately affect the clinical utility of preoperative testing. Different hospitals, surgeons, and anesthesiologists have different protocols for obtaining preoperative testing, including, but not limited to, ad hoc testing by the surgeon or anesthesiologist, referral to the patient's primary care physician for testing at his or her discretion, and dedicated clinics with standardized protocols based on a patient's health status and planned surgery. Further, the comparator intervention, ad hoc testing, is by definition variable, depending on the clinician ordering the test, to what degree testing is based on any H&P he or she performs, and each clinician's likelihood of ordering few or many tests, which in part will be based on the local culture. Subsequent to testing, there is an implementation issue, in that any changes to patient outcomes due to testing must be mediated through clinical decisions about how to act on abnormal tests. Again, individual clinicians, different specialties, and different surgical settings are likely to

have different thresholds for when and how to respond to abnormal tests. Examples include decisions about whether to delay or cancel surgery or whether to administer blood components preoperatively. This variability in care practices raises questions about whether ad hoc testing results in underutilization and/or overutilization of tests (balancing benefits and harms) compared with per-protocol testing, as well as whether testing ordered and followed up by different disciplines or types of clinicians has equivalent clinical utility.

### **Timing of Testing**

A final factor that needs to be considered is the timing of the tests. Hospitals or surgical centers may dictate that preoperative testing must be done within a limited period before surgery, such as 30 days or 6 months. It is unknown whether there is adequate evidence to support any particular time threshold for preoperative tests.

### **Assessing Clinical Utility of Preoperative Testing**

Preoperative testing can have a direct impact only on certain outcomes of interest, including emotional and cognitive changes in the patient conferred by testing and its results; any harms associated with the testing procedure (e.g., pain, hemorrhage, or bruising from a blood draw; exposure to ionizing radiation from imaging tests); and costs to the patient (in the form of time spent or copayments) or other types of resource utilization. For the most part, however, testing has indirect effects, including influencing treatment choices, delay or cancellation of the procedure (either appropriately to allow correction of or further treatment due to an abnormal test result or unnecessarily if no further treatment or evaluation was truly needed), and cascade testing (where abnormal tests lead to further appropriate or unnecessary tests).

Comparative studies of different preoperative testing strategies can effectively analyze all outcomes of interest. The range of outcomes that can be meaningfully assessed by noncomparative (cohort) studies, though, is more limited. Complication rates, the most important patient-centered outcome, can be adequately assessed only by comparative studies, since the underlying risk of complications will vary across cohorts of patients and types of surgery. The complication rate in a cohort study of routine testing is difficult to interpret in the absence of an estimate of the expected complication rate without routine testing. The only outcomes from cohort studies that can provide some information to address the Key Questions in this report are those directly related to the testing, such as surgery cancellation or delay due to an abnormal test result. However, this outcome is of somewhat limited

value, since it does not address whether the patient benefited from or was harmed by the surgical cancellation or delay.

### **Statement of Work**

Three professional medical associations nominated this topic for systematic review, citing the wide variation in clinical practice, the need for a guideline for routine preoperative testing, and the likelihood that a comparative effectiveness review on this subject would have broad clinical impact—particularly if such a review included the most commonly ordered tests in healthy patients, as well as those with comorbidities, undergoing a wide variety of high- and low-risk surgeries. The target audience for this review includes surgeons, anesthesiologists, and other clinicians involved in perioperative care of surgical patients; policymakers, including clinical practice guideline developers and surgical clinic administrators involved in determining preoperative testing policies and protocols; health care payers; researchers with an interest in perioperative care; and, ultimately, patients undergoing surgical procedures.

The review focuses on the direct evidence (evidence regarding actual changes in patient outcomes and management) of the comparative value of routine preoperative testing versus not testing (or other protocols for testing). This evidence is derived primarily from studies that directly compare testing protocols. These are the only studies that can demonstrate whether uniformly testing an unselected population prior to surgery leads to better outcomes for those patients. We also included cohort studies that report rates of “process outcomes” (rates of surgery cancellation, changes to planned surgery or anesthesia, etc.) only for patients being tested, since the rate of procedure delay, cancellation, and other changes due to testing is, by definition, zero in patients who do not undergo testing.

The review does not evaluate questions that, while important and related to the topic at hand, do not provide direct evidence of the comparative value of testing versus not testing. The review does not evaluate analyses that would require assumptions about what might have occurred without testing or assumptions about how testing might improve outcomes based on different rates of complications among patients with abnormal and normal preoperative tests. Specifically—

- We do not base assessments of the benefits and harms of preoperative testing on the incidence of perioperative complications (such as major bleeding) in studies that report only on patients who underwent testing (i.e.,

noncomparative studies). While these studies make conclusions regarding the possible value of testing, they do not provide evidence regarding the actual effect of routine preoperative tests, since the complication rates absent routine testing are unknown.

- We do not systematically review the prevalence rates of abnormal test results for different populations of patients undergoing surgery. These data do not provide evidence that ordering the test would alter perioperative outcomes, since the effect of acting on the abnormal test result on perioperative outcomes is unknown.
- We do not systematically review the test performance (e.g., sensitivity and specificity) of any of the tests because, again, the effect on perioperative outcomes of acting on the true or false abnormal test result is unknown.
- We do not assess test results (i.e., abnormal vs. normal) as predictors of outcomes. The goal of this review is to assess whether actually ordering routine preoperative tests alters care and patient outcomes, and association studies do not provide data on how the test performs in different populations or the balance of benefits and harms.

## Key Questions

We address the following Key Questions:

**Key Question 1:** How do routine or per-protocol preoperative testing strategies compare to no testing or alternative testing strategies with respect to outcomes—including perioperative clinical outcomes, quality of life or satisfaction, periprocedural patient management decisions, and resource utilization—among patients undergoing elective surgical procedures? How do outcomes vary by—

- a. The risk of the surgical procedure, the type of anesthesia planned, the indication for surgery, comorbidities, or other patient characteristics?

- b. The structure of testing (e.g., routine for everyone vs. per protocol, whether testing is conducted in a specialized preoperative clinic) or who orders the tests (e.g., surgeon vs. anesthesiologist vs. primary care physician)?
- c. The length of time prior to the procedure that the tests are conducted?

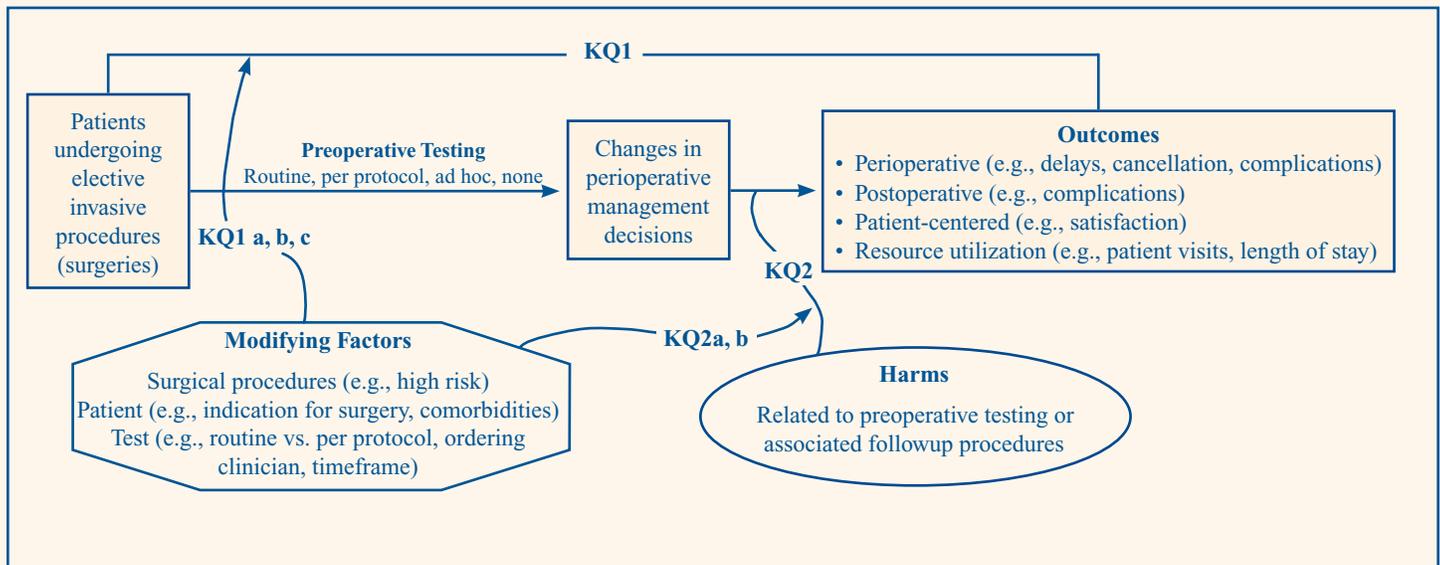
**Key Question 2:** What are the harms of routine or per-protocol preoperative testing strategies compared to no testing or to alternative testing strategies? How do outcomes vary by—

- a. The risk of the surgical procedure, the type of anesthesia planned, the indication for surgery, comorbidities, or other patient characteristics?
- b. The structure of testing (e.g., routine for everyone vs. per protocol, whether testing is conducted in a specialized preoperative clinic) or who orders the tests (e.g., surgeon vs. anesthesiologist vs. primary care physician)?

## Analytic Framework

To guide the development of the Key Questions for the evaluation of preoperative testing, we developed an analytic framework (Figure A) that maps the specific linkages associating the populations of interest, the interventions, the outcomes of interest (including harms), and the potential modifying factors. Specifically, this analytic framework depicts the chain of logic that the evidence must support to link the interventions to improved health outcomes.

**Figure A. Analytic framework for routine preoperative testing**



KQ = Key Question.

## Methods

During a phase of topic refinement, in preparation for conducting this comparative effectiveness review, we convened a panel of Key Informants (including domain experts in anesthesia, general and breast surgery, and cardiology; health care payers with an interest in preoperative testing; a patient advocate; and representatives from the three topic nominators) and local domain experts (including an epidemiologist, internist, anesthesiologist, ophthalmologist, radiologist, and a thoracic and general surgeon). These individuals helped the team develop the Key Questions and the scope of work. We convened a Technical Expert Panel (TEP), which included experts in anesthesia, general surgery, urology, cardiology, internal medicine, and family medicine. The TEP provided input to help refine the protocol, identify important issues, and define parameters for the review of evidence. The TEP was also asked to suggest additional studies.

We conducted literature searches of studies in MEDLINE® and Ovid Healthstar® (from inception to July 22, 2013), as well as the Cochrane Central Trials Registry and Cochrane Database of Systematic Reviews (through the second quarter of 2013). The reference lists of prior systematic reviews and relevant guidelines were hand-searched. All citations were screened to identify articles relevant to each Key Question. The search included terms for surgical procedures, preoperative care, and diagnostic tests, including the specific tests ECG, chest radiography, blood counts, coagulation tests, biochemistry, glucose, urinalysis, kidney function tests, liver function tests, pregnancy tests, hemoglobinopathies, and pulmonary function tests.

Three team members double-screened all abstracts after an iterative training period to ensure that all screeners agreed upon the eligibility criteria. Full-text articles were retrieved for all potentially relevant articles. These were rescreened for eligibility. All rejected articles were confirmed by the team leader.

## Population and Condition of Interest

We included studies conducted in both adults ( $\geq 18$  years) and children undergoing surgical procedures requiring either anesthesia or sedation, including—

- Patients undergoing any elective or ambulatory surgical or other invasive procedure that commonly requires anesthesia or sedation of any type or approach that is administered by an anesthesia team member. Cataract surgery was included regardless of local practice regarding anesthesia or sedation.
- Procedures in any setting, including inpatient, outpatient, and office based.
- Any category of risk for surgical or anesthetic complications.
- Surgical procedures in any risk category, ranging from minor and minimally invasive through high-risk, maximally invasive surgeries (e.g., vascular, neurologic, thoracic, abdominal, and pelvic surgeries).

Patients undergoing nonsurgical diagnostic procedures that may require anesthesia or sedation (e.g., biopsy, colonoscopy) were excluded.

## Interventions of Interest

We included all preoperative tests that we, our local expert, and the TEP agreed were likely to be conducted routinely or on a per-protocol basis. These included basic laboratory tests, simple radiography, and selected other relatively simple diagnostic tests.

The tests had to have been conducted in the preoperative period for the purpose of assessing the patient's risk and status prior to the planned procedure. We excluded tests performed for the purpose of diagnosis or staging of the disease for which the surgery was being performed or for specific surgical planning. We also excluded patient factors other than tests, including patient history, symptoms, physical examination signs or findings, and demographic features, or panels of "tests" that included any of these factors. While patient symptoms, such as decompensated congestive heart failure, may be important reasons for altering, delaying, or canceling surgery, they should be routinely assessed as part of an appropriate standard of care. In addition, for a given surgical procedure or set of procedures, the tests had to have been conducted either routinely (i.e., in all patients undergoing the procedure, regardless of age, sex, or medical condition) or based on a standard protocol (i.e., in all patients who met certain predetermined criteria based on age, sex, medical condition, or other factors).

Intervention and comparator arms were sorted into four categories: routine (everyone was scheduled to have all tests), per protocol (a protocol was used to determine who had which tests), ad hoc (testing was done at a clinician's discretion), or no testing. The distinction between routine and per-protocol testing was not always clear. If a study did not report sufficient information to distinguish the two, we assumed that routine testing was conducted. In a few instances, when a large number of tests were done routinely and a single test (e.g., ECG) was done per protocol, we also categorized this as routine testing.

## Comparators of Interest

Comparators of interest included no preoperative testing (of a panel of tests or an individual test); ad hoc testing (i.e., the tests were conducted at the discretion of the ordering clinician, regardless of the reason); per-protocol testing (as a comparator to routine testing); a different panel of routine tests; testing conducted in a different setting or by a different type of clinician (e.g., in a specialized preoperative testing clinic vs. by the patient's primary care physician); and testing done at different presurgery time points (e.g., within 30 days vs. within 6 months).

## Outcomes of Interest

For Key Question 1, outcomes were confined to those related to the conduct of the surgical procedures and anesthesia, perioperative events, patient satisfaction, and resource utilization. Specifically, they included clinical and other patient-centered outcomes (procedure or anesthesia delay, procedure cancellation, perioperative outcomes, including mortality and surgical complications); quality of life; satisfaction; patient resources; unplanned hospital readmission; change in disposition of care after surgery; length of hospital stay; other resource utilization, such as additional testing induced by a positive test or treatments for perioperative complications; and an intermediate outcome (changes to perioperative patient management other than procedure delay or cancellation). For Key Question 2, outcomes of interest included adverse events or harms related to testing, including complications of followup testing or treatment of abnormal test results, or poor outcomes related to delaying or canceling a procedure.

## Eligible Study Designs

We included published peer-reviewed articles. We included studies that covered any timeframe, although they had to be longitudinal in design to the extent that testing was done prior to the planned procedure and followup occurred at least up to the time of the procedure.

We included comparative studies (in which one or more protocols for testing were compared with other protocols for testing, including protocols for no testing), whether randomized controlled trials (RCTs) or nonrandomized studies. We included both prospective and retrospective studies.

Because we expected the comparative studies to be limited in quantity and quality, we also evaluated cohort (noncomparative single-group) studies in which all study participants had the same testing battery or protocol. However, we limited these studies to those that reported "process" outcomes in which the process of care was altered, including procedure or anesthesia delay; procedure cancellation; and other resource utilization, such as unplanned followup tests or procedures and changes to perioperative patient management. As discussed above in the Statement of Work section, rates of other outcomes without a comparator would not provide interpretable data about the true benefits or harms of routine testing.

## Data Extraction

Data from each study were extracted by one experienced methodologist. The extraction was reviewed and confirmed by at least one other methodologist. Data were extracted into customized forms in the Systematic Review Data Repository™ at [sdr.ahrq.gov](http://sdr.ahrq.gov).

## Quality Assessment

We assessed the methodological quality of studies based on predefined criteria. We used a three-category grading system (low, medium, or high risk of bias) to denote the methodological quality of each study. This system defines a generic grading scheme that is applicable to varying study designs, including RCTs, nonrandomized studies, and cohort studies.

**Low risk of bias.** These studies have the least apparent bias, and their results are considered valid. They generally possess the following: a clear description of the population, setting, interventions, and comparison groups; appropriate measurement of outcomes; appropriate statistical and analytic methods and reporting; no reporting errors; clear reporting of dropouts and a dropout rate less than 20 percent; and no obvious bias.

**Medium risk of bias.** These studies are susceptible to some bias, but it is not sufficient to invalidate the results. They do not meet all the criteria for low risk of bias due to some deficiencies, but none are likely to introduce major bias. They may be missing information, making it difficult to assess limitations, including risk of bias per se, and potential problems.

**High risk of bias.** These studies have been judged to carry a significant risk of bias that may invalidate the reported findings. These studies have serious errors in design, analysis, or reporting and contain discrepancies in reporting or have large amounts of missing information.

## Minimal Important Difference

With input from the TEP, we made a priori definitions of minimal important differences (MIDs). The MID is a clearly defined clinical threshold, below which the evidence (effect estimates and corresponding confidence intervals [CIs]) shows no meaningful difference and above which the evidence shows a benefit or harm of one intervention over another. For mortality and major or severe life- or health-altering morbidities and complications (such as stroke, myocardial infarction, or life-threatening hemorrhage), the MID is 0 percent because any difference is of concern to patients and clinicians for this low-risk (generally low-cost) intervention

(preoperative testing). However, to make the determination that there is evidence of no difference, we used a threshold of 20 percent on the relative risk (RR) scale. For other, noncritical outcomes, we also used an MID of 20 percent, based on agreement that smaller differences would not be clinically important.

## Grading the Body of Evidence

We graded the strength of the body of evidence, in accordance with the AHRQ “Methods Guide for Effectiveness and Comparative Effectiveness Reviews,”<sup>9</sup> based on risk of bias, consistency across studies, directness of the evidence, precision (based on the MID), and risk of reporting bias. The strength of evidence was ranked as either high, moderate, low, or insufficient. Ratings were assigned based on our level of confidence that the evidence reflected the true effect for the major comparisons of interest. We further assessed the body of evidence regarding its applicability to the U.S. population of patients undergoing surgical procedures.

## Results

The literature search yielded 4,581 citations. From these, 220 articles were provisionally accepted for review based on abstracts and titles. After screening the full text, 57 studies (in 58 articles) were found to have met the inclusion criteria. Fourteen of the 57 were comparative, and the remainder were single-group studies. Three RCTs focused on cataract surgery, two RCTs and six nonrandomized studies focused on general or various surgeries, one RCT focused on vascular surgery, and one nonrandomized study each focused on tonsillectomy and orthopedics. Overall, the studies evaluated the preoperative tests for the following procedures: general or various surgeries (37 studies), tonsillectomy (5 studies), cataract surgery (4 studies), orthopedic surgery (4 studies), vascular surgery (3 studies), head and neck/ear, nose, throat surgery (2 studies), and 1 study each for neurosurgery and electroconvulsive therapy. Seventeen of the studies were conducted in children, 25 in adults, and 15 in a mixed population of adults and children. Forty studies were published before 2000, including 7 of the 14 comparative studies; 17 studies were published after 2000. Thirteen studies had a high risk of bias, 10 had a medium risk of bias, and 34 had a low risk of bias.

The preoperative tests evaluated in the studies fall into the following categories: basic metabolic panel (electrolytes, kidney function, glucose); extended metabolic panel (liver function tests and other serum tests); blood counts

(including hemoglobin, hematocrit, white blood cells, and platelets); hemostasis tests (including prothrombin time, partial thromboplastin time, and bleeding time); urinalysis; pregnancy tests; ECG; chest x ray (CXR); pulmonary function testing; and echocardiography.

## **Comparative Studies**

### **Cataract Surgery**

Three RCTs—two with low, one with moderate risk of bias—compared routine versus no (or ad hoc) preoperative testing with ECG, basic metabolic panel, and complete blood count (CBC) for patients undergoing cataract surgery. The studies were clinically similar to each other and consistent; there is a high strength of evidence of no clinically important difference in complication rates. By meta-analysis, for total complications, the RR is 0.99 (95% CI, 0.86 to 1.14). There is also a high strength of evidence suggesting that routine testing does not affect rates of procedure cancellation, but the confidence intervals were too wide to definitely exclude a clinically important difference: RR=1.00 (95% CI, 0.42 to 2.38) and 0.97 (95% CI, 0.78 to 1.21). No other outcomes were reported. The evidence is inadequate to evaluate potential differences based on subgroups of patients. Overall, there is no evidence of different outcomes related to routine preoperative testing.

### **General or Various Surgeries, Adults**

One RCT with low risk of bias and four nonrandomized studies with high risk of bias compared routine testing (two studies) or per-protocol testing (three studies) with ad hoc testing, using ECG, CXR, basic and extended metabolic panels, CBC, hemostasis tests, and urinalysis in adult patients undergoing a broad range of elective surgeries. A sixth study compared time periods when patients were to receive either routine testing, during a retrospective period, or per-protocol testing, during a prospective period, with a large number of tests. None of the nonrandomized studies adjusted for baseline differences in patient characteristics, types of surgery, surgeons or anesthesiologists, their experience, or other confounders. They also did not analyze how or whether the routine or per-protocol tests were linked to resulting outcomes (complications). The RCT reported only on complications, of which there were only a small number; therefore, this trial was underpowered to provide any reliable estimate of relative differences in complications. We have no confidence in the estimate of effects across these studies due to these methodological deficiencies,

the important clinical heterogeneity (differences) across all studies, and the high risk of bias of the nonrandomized studies (particularly related to lack of necessary adjustments). Therefore, there is insufficient evidence regarding perioperative complications. There is also insufficient evidence of a clinically significant difference in the rate of perioperative death. The clinical heterogeneity of studies, without reporting of subgroup analyses of patients or procedures within studies, further precludes a conclusion about which patients would benefit from routine testing. There is also insufficient evidence regarding other specific outcomes, including return to the operating room, prolonged hospital stay, or surgical cancellation or delay. No trial reported on quality of life or satisfaction, change in anesthesia or procedure plan, or resource utilization. A single nonrandomized study with high risk of bias provided insufficient evidence regarding the comparison of routine and per-protocol testing. Given the deficiencies in the evidence across studies, it was not possible to compare the effects of routine and per-protocol testing. No trial addressed Key Question 2 regarding harms of routine preoperative testing. The evidence is inadequate to evaluate potential differences based on subgroups of interest.

### **Orthopedic Surgery, Adults**

There is insufficient evidence regarding the comparison of routine versus per-protocol preoperative testing in adults undergoing orthopedic surgery. A single retrospective nonrandomized study with high risk of bias found no difference in the rate of unplanned hospital admissions within 30 days of surgery.

### **Vascular Surgery, Adults**

There is insufficient evidence regarding the comparison of routine versus per-protocol preoperative testing in adults undergoing vascular surgery. A single RCT with low risk of bias failed to find differences in rates of perioperative death or cardiac complications.

### **General or Various Surgeries, Children**

One RCT from 1975 with medium risk of bias reported limited outcome data. A retrospective nonrandomized study with high risk of bias failed to provide sufficient evidence regarding the effect on patient and resource outcomes of routine or per-protocol preoperative testing. The limited data suggest no difference in length of hospital stay related to routine testing with basic and extended metabolic panels and a counterintuitive increase in minor perioperative complications with routine preoperative

testing. The age of the studies (38 and 15 years) further calls into question the applicability of their findings to modern pediatric surgical management. No study reported on quality of life, satisfaction, surgical delay, change in anesthesia or procedure plan, resource utilization, or harms of routine testing. The evidence is inadequate to evaluate potential differences based on subgroups of interest.

### **Tonsillectomy and/or Adenoidectomy, Children**

There is insufficient evidence regarding routine or per-protocol preoperative testing in children undergoing tonsillectomy and/or adenoidectomy. A single flawed retrospective nonrandomized study that is 16 years old found significantly higher rates of perioperative bleeding among patients of less experienced surgeons who routinely conducted hemostasis tests than those of more experienced surgeons who performed per-protocol testing. However, none of the bleeding episodes were related to clinically significant abnormal coagulation tests, and the difference in bleeding rates was more likely to have been related to the experience and surgical volume of the surgeons.

### **Cohort Studies**

Given how few comparative studies were available, we looked at cohort studies to test the indirect link between testing and outcomes, since if tests can be shown not to affect management, they cannot affect outcomes. The weaknesses with this approach are that it is not possible to determine if the change in management led to better or worse outcomes and that the implicit comparison can be made only with no testing. No implicit comparison can be made with ad hoc testing based on H&P, since there are no data on management changes based on the ad hoc testing. For the purposes of this section, we combined data from the true cohort studies and the routine or per-protocol arms from the comparative studies. This section focuses on the rates of specific outcomes, and the data from the comparative studies are equivalent to those from the cohort studies. Among the 57 studies eligible for this review, the 47 with relevant outcomes are summarized in this section.

The 47 studies report a total of five “process” outcomes of interest: change in patient management (4 studies conducted in adults); change in surgical technique (3 studies conducted in adults, 1 study conducted in children); change in anesthetic management (10 studies conducted in adults, 6 studies conducted in children); procedure cancellation (19 studies conducted in adults, 11 studies conducted in children); and procedure or anesthetic delay (19 studies conducted in adults, 7 studies conducted in children). Thirty-three (70%) of the studies were published

before 2000. Except for a 5.1-percent rate of procedure delays in one study from 2005, all patient management changes that occurred in 2 percent or more of patients were in older studies. Thirty-nine (83%) of the studies evaluated routine preoperative testing; the other eight evaluated per-protocol testing. An important caveat for the analysis of these studies is that, in general, it is only implied that procedure changes or cancellations were truly due to abnormal test results as opposed to changes that may have occurred for reasons separate from testing. While this caveat also applies to the comparative studies, in these analyses there is no reference group for comparison.

With these caveats, the following conclusions can be made from the cohort studies. In all preoperative testing scenarios for which more than a single study was available (i.e., approaching a sufficient evidence base to form a conclusion), testing resulted in some changes in management. In other words, the evidence suggests that in most situations, routine preoperative testing will result in some delay or cancellation of the procedure (in most studies, <2%) or some changes to anesthetic management (up to 11%) or surgical procedure (<1%). However, it is not possible to say whether the changes led to benefit or harm for patients because, without a comparator group, one cannot assess how the changes in management may have been associated with perioperative outcomes. Two studies suggest that change in management from CXR is more common for older patients (primarily >60 years). Two other studies looked at CXR and ECG by sex and other factors. One of these studies suggests that the effect of ECG is similar in men and women, but the second study suggests that CXR results in change in management in more men, those in a higher ASA risk category, those with respiratory disease, and those with “major” surgeries planned (as opposed to “minor” or “standard” surgeries), particularly in patients undergoing thoracic, cardiac, and vascular surgeries. The studies were too clinically heterogeneous to ascertain whether there were any patterns suggesting a difference in process outcomes based on whether preoperative testing was conducted routinely or per protocol.

## **Discussion**

### **Key Findings and Strength of Evidence**

We identified 57 studies that reported clinically pertinent outcomes in patients who had routine or per-protocol preoperative testing performed. However, only 14 of the studies provided direct comparisons between routine or per-protocol testing and ad hoc or no testing, and only

two studies compared routine with per-protocol testing. Furthermore, only seven of the comparative studies were RCTs, three of which were conducted in patients undergoing cataract surgery. The large majority of data come from cohort studies that provided evidence only about how frequently procedures or anesthesia were canceled, delayed, or altered in response to preoperative testing.

In summary, there is a high strength of evidence from three well-conducted RCTs that consistently found that, for patients scheduled for cataract surgery, preoperative ECG, metabolic panel (or glucose), and CBC have no effect on total perioperative complications or procedure cancellation (Table A). In contrast, there is insufficient evidence for the effect of routine preoperative testing in all other surgeries and populations. There is also insufficient evidence to estimate a difference in outcomes based on whether preoperative testing was conducted routinely or per protocol. There are one RCT and five nonrandomized studies of routine or per-protocol testing in adults undergoing various elective surgeries; however, the studies were highly heterogeneous in populations, elective surgeries, and tests used. Furthermore, the nonrandomized studies were all fundamentally flawed in that they failed to adjust for differences among study groups in the patients, surgeries, surgeons, anesthetics used, anesthesiologists, or other possible confounders. These studies generally found lower rates of postoperative complications and deaths among patients undergoing routine or per-protocol testing, but the heterogeneity and flaws in the studies preclude any confidence in the accuracy or validity of the findings. However, while there is no evidence regarding minimally invasive surgeries similar to cataract surgery, it may be valid to conclude that routine preoperative testing in these other low-risk surgeries would also have no effect.

There is insufficient evidence for all other categories of procedures and patients, for all other outcomes of interest, and regarding more detailed analyses of differences in how testing is performed. In particular, there is no comparative evidence regarding quality of life or satisfaction, resource utilization, or harms. Among comparative studies, there is insufficient reported evidence regarding how outcomes may differ in different subgroups of patients, or how the effect of preoperative testing may vary based on the risk of the surgical procedure or other factors.

The apparent difference in the effect of routine or per-protocol testing in patients undergoing cataract and general elective surgery is arguably not surprising. Cataract surgery is a very low-risk procedure, safe enough to be done in an ophthalmologist's office, that is minimally

invasive and usually requires only local anesthesia with sedation. Other than increases in vagal tone, there is little reason to expect cardiac strain in the typical patient undergoing cataract surgery. While the patients are typically elderly, and thus have a relatively high rate of comorbidities, they are generally not suffering from any acute illnesses. In contrast, general elective surgeries in adults encompass a wide range of patients and surgeries, including many with acute or serious medical conditions requiring surgery and highly invasive cardiothoracic, abdominal, and vascular surgeries. These patients are intrinsically at higher risk of perioperative complications and thus, conceptually, may benefit most from preoperative tests that pick up correctable abnormalities that may be associated with complications.

Most of the evidence was from cohort studies. However, the nature of the intervention under consideration (preoperative testing) makes the lack of a direct comparator (ad hoc testing) among these studies particularly problematic in terms of interpreting the findings. Regardless of the specific preoperative tests used or how they are implemented, the rate of perioperative complications due to either the procedure or the anesthesia will always depend primarily on the underlying risks of the surgical procedure, the type of anesthesia used, the skill and experience of the surgeons and anesthesiologists, the medical condition of the patients, and the quality of perioperative care. The risk of perioperative complications when preoperative testing was conducted, without information about the risk of complications without testing (or only ad hoc testing), does not provide information on the effect of the testing on those risks. An adequate comparator that controls for the myriad factors that also impact perioperative complications is needed.

## Study Limitations

Across nonrandomized studies, there was a lack of adjustment for possible confounders. All of the nonrandomized studies failed to control for cluster effects, particularly those related to individual surgeons or surgical experience. Six nonrandomized studies compared different time periods within an institution before or after implementation or removal of a preoperative testing policy. However, institutional differences between the time periods (such as incremental improvements in surgical techniques, anesthesia, or nursing care) were not accounted for. The bias that can result from the lack of adjustment (e.g., by propensity score) was best exemplified in the nonrandomized study that compared concurrent surgeries. In one of the two comparative studies comparing routine versus per-protocol testing with hemostasis tests on

children undergoing tonsillectomy and/or adenoidectomy, the comparison was really between the bleeding complication rates of the 2 most experienced surgeons (who used a testing protocol in 2,624 children) and those of the 11 less experienced surgeons (who did routine testing in 1,750 children total). Arguably, the finding that perioperative bleeding was more common in the latter group provides evidence that surgical experience and skill are predictors of complications and says little or nothing about whether preoperative testing may (or may not) have prevented any bleeding episodes.

### **Intrinsic Limitations of Research on Preoperative Testing**

Another limitation of the evidence that would be difficult to overcome also relates to the nature of the intervention. Preoperative testing does not in and of itself affect the outcomes of interest (except resource utilization and possibly quality of life/satisfaction, although there are no data on these outcomes). Instead, the preoperative tests potentially cause the health care providers to alter a patient's management—by implementing an intervention to correct or account for the abnormal test; by delaying, canceling, or changing the procedure or anesthesia; or by making changes to postoperative care. Additionally, the preoperative test may be useful for perioperative management to use as a reference (e.g., to know whether a measure has changed in a postoperative test compared with the preoperative test—for example, whether an ECG abnormality is new or not). Thus, the value of any preoperative test is fully dependent on the health care providers and their responses to abnormal tests. One could expect responses to vary among surgeons, anesthesiologists, primary care physicians, nurse practitioners, and other providers. One could also expect them to vary among individual providers across hospitals, settings (e.g., urban vs. rural), geographic regions, and a myriad of other health care provider variables. However, none of these factors were accounted for in the studies. This limitation further hampers the interpretation of the evidence, particularly from the cohort studies, but also arguably from the unadjusted nonrandomized studies.

Interpretation of the evidence is further complicated by the wide variability in clinical practice in the thoroughness of preoperative H&P (and whether it is done) and the general lack of reporting regarding H&P in the studies. This could have an important impact on what tests are conducted ad hoc (i.e., in the comparator arms of the studies). Rather than leading to more or less testing, it can lead to more appropriate testing, since the tendency to order tests based on a “shotgun” approach will be reduced. But H&P could be considered equivalent to a

“test” performed by the clinician (instead of the laboratory or radiology technician), which may or may not have value independent of true preoperative tests. Furthermore, H&P is intrinsically nonstandardized and heterogeneous, depending on the specific questions asked and the details of the examination. Traditionally, H&Ps have been completed in the surgical clinics and on the day of surgery by the anesthesiology teams. More recently, preoperative assessment clinics staffed by perioperative medicine specialists are becoming more common. These clinics focus on optimizing patients for their perioperative course, and a thorough H&P is the cornerstone of that process. However, none of the studies specifically investigated testing in this setting.

Any management changes due to abnormal test results (and presumably any subsequent changes in perioperative outcomes) would logically be the same regardless of whether testing was done routinely, per protocol, or at the clinician's discretion. Therefore, the variability in ad hoc testing could have an important impact on the comparison of outcomes between ad hoc and routine or per-protocol testing. Without good descriptions in studies of typical H&P or the triggers to order ad hoc tests, it is difficult to interpret the applicability of the studies to the general (or any specific) population and the comparison among different testing regimens.

### **Limitations of Cohort Studies**

Because of the underlying lack of interpretability of the complication rates in these studies, we restricted analyses to “process” outcomes related to decisions about whether the procedure or anesthesia was altered based on testing. These included cancellation or delay of surgery, changes in either the planned surgery or anesthesia, and overall changes in patient management. To the extent possible, based on the reported data, we focused on decisions that were made specifically because of test results (presumably abnormal results), but most studies did not clearly define their outcomes, requiring us to assume this was the case. However, the information to be gleaned from most of these studies was limited. When no procedures were canceled or delayed and no changes were made to either the planned procedure or anesthesia, it may be reasonable to conclude that the testing was of no value, at least up to the time that the procedure was performed. However, the assumption that the testing was of no value overall requires that the postoperative course also be unaffected by the availability of the preoperative tests. In reality, it is likely that some abnormal preoperative tests, such as an elevated glucose, would alter perioperative management, such as more intensive glucose monitoring.

Interpreting the finding that a certain (nonzero) percentage of procedures were canceled, delayed, or changed is not straightforward. First, one must make a conclusion as to whether the cancellations, delays, or changes were warranted. Second, one must make assumptions about whether the patients' outcomes were changed. If a procedure was canceled or delayed, at a certain level the patient's immediate health care was worsened, assuming the planned surgery was necessary. However, it is unknowable whether the delay or cancellation may have prevented a complication that would have been worse than the prolongation of the disease state necessitating surgery. Third, one must make a determination as to whether the testing led to changes in care sufficiently rarely (below some percentage threshold) that the testing is of sufficiently limited value to safely forgo it, or whether the changes in care occur frequently enough that they can be assumed to be an important tool or predictor regarding surgical management.

With these caveats, the following conclusions can be made from the cohort studies. In all cases where there are at least two studies (i.e., approaching a sufficient evidence base to form a conclusion), there was no test or set of tests used routinely for a similar population (adults or children) prior to a similar set of procedures for which the testing consistently resulted in no changes in management. In

other words, the evidence suggests that in most situations, routine preoperative testing will result in some delay or cancellation of the procedure or some change to anesthetic management or surgical procedure. Again, whether these changes benefit or harm patients is unknown from these data. That said, the only studies that directly compared outcomes in subsets of patients were cohort studies that evaluated changes in patient management, including specialty consultations or nonsurgery-related changes in patient care. Two studies suggest that change in management from CXR is more common for older patients (primarily >60 years). Two other studies also looked at CXR and ECG by sex and other factors. One of these studies suggests that the effect of ECG is similar in men and women, but the second study suggests that CXR results in change in management in more men, those in a higher ASA risk category, those with respiratory disease, and those with "major" surgeries planned (as opposed to "minor" or "standard" surgeries), particularly in patients undergoing thoracic, cardiac, and vascular surgeries. However, given the small number of studies that compared outcomes in different subgroups of patients, together with the unknown connection between changing patient management and true patient outcomes, it is premature to conclude that the differences found are clinically important.

**Table A. Routine or per-protocol preoperative testing: Findings and strength of evidence**

<b>Outcome</b>	<b>Surgery</b>	<b>Tests</b>	<b>Study Design (Risk of Bias)</b>	<b>Finding</b>	<b>Strength of Evidence</b>
Perioperative complications, total	Cataract surgery	ECG, metabolic panel, CBC	RCT (2 low, 1 medium)	No effect of testing (summary RR = 0.99; 95% CI, 0.86 to 1.14).	High
	Various, adults (comparison: routine vs. ad hoc testing)	Multiple*	RCT (1 low) NRS(4 high)	In most studies, fewer complications occurred with testing, but studies were highly heterogeneous and underpowered; not a clinically important difference.	Insufficient
	Various, adults (comparison: routine vs. per-protocol testing)	Multiple*	NRS (1 high)	No events in either group.	Insufficient
	Various, children	Multiple†	NRS (1 high)	More complications occurred with testing, but not a clinically important difference.	Insufficient
Perioperative death	Vascular, adults	Stress echo	RCT (1 high)	No significant difference in cardiac events.	Insufficient
	Various, adults (comparison: routine vs. ad hoc testing)	Multiple*	NRS (4 high)	In most studies, fewer deaths occurred with testing, but studies were highly heterogeneous and underpowered.	Insufficient
	Various, adults (comparison: routine vs. per-protocol testing)	Multiple*	NRS (1 high)	No events in either group.	Insufficient
	Vascular, adults	Stress echo	RCT (1 high)	Cardiac and respiratory deaths were rare; no difference between groups.	Insufficient
Perioperative complications, major (total)	Various, children	Multiple†	NRS (1 high)	Imprecise estimate failing to support a difference.	Insufficient

**Table A. Routine or per-protocol preoperative testing: Findings and strength of evidence (continued)**

<b>Outcome</b>	<b>Surgery</b>	<b>Tests</b>	<b>Study Design (Risk of Bias)</b>	<b>Finding</b>	<b>Strength of Evidence</b>
Perioperative complications, specific (selected)	Various, adults (comparison: routine vs. ad hoc testing)	Multiple*	RCT (1 low) NRS (3 high)	Clinically important difference: fewer episodes of renal failure with testing (0.9% vs. 0%; 1 study). Significant but not clinically important difference: fewer episodes of pneumonia with testing (1 study). No significant differences for other complications, including any outcome from RCT	Insufficient
	Various, adults (comparison: routine vs. per-protocol testing)	Multiple*	NRS (1 high)	No difference between groups, but only rare events.	Insufficient
Return to operating room	Various, children	Multiple†	NRS (1 high)	Clinically important difference: more episodes of persistent vomiting with testing (RR = 1.76; 95% CI, 1.22 to 2.54). Clinically important difference: more episodes of restlessness with testing (RR = 3.91; 95% CI, 2.19 to 6.97). No significant differences were found for other complications.	Insufficient
	Tonsillectomy, children (comparison: routine vs. ad hoc testing)	Coagulation tests	NRS (1 high)	No significant difference in bleeding complications.	Insufficient
Unplanned hospital admission	Orthopedic, adults	Multiple*	NRS (1 high)	No significant difference in rate of return to operating room. No significant difference in rate of unplanned hospital admissions.	Insufficient

**Table A. Routine or per-protocol preoperative testing: Findings and strength of evidence (continued)**

Outcome	Surgery	Tests	Study Design (Risk of Bias)	Finding	Strength of Evidence
Procedure cancellation	Cataract surgery	ECG, metabolic panel, CBC	RCT (1 low, 1 medium)	Likely no effect of testing‡ (summary RR = 0.97; 95% CI, 0.79 to 1.20)	High
	Various, adults	Multiple*	NRS (1 high)	Possibly no effect of testing (RR = 0.93; 95% CI, 0.76 to 1.14).	Insufficient
	Various, children	Multiple†	NRS (1 high)	No effect of testing (no surgeries canceled).	Insufficient
Procedure delay	Various, adults	Multiple*	NRS (1 high)	No significant difference in procedure delay.	Insufficient
	Various, children	Multiple†	NRS (1 high)	No significant difference in procedure delay.	Insufficient
Length of stay	Various, adults	Multiple*	NRS (1 high) RCT (1 medium)	No significant difference in length of stay.	Insufficient
	Various, children	Multiple†	NRS (1 high)	No significant difference in length of stay.	Insufficient
Quality of life/satisfaction, anesthesia change, surgery change, resource utilization, or harms	None	Not applicable	No studies	None	Insufficient
	None	Not applicable	No studies	None	Insufficient

CBC = complete blood count, CI = confidence interval, ECG = electrocardiogram, NRS = nonrandomized comparative study,

RCT = randomized controlled trial, RR = relative risk, Stress echo = dobutamine stress echocardiogram.

\*ECG, chest x ray, basic and extended metabolic panels, CBC, coagulation tests, and urinalysis.

†Hemoglobin, urinalysis, creatine phosphokinase, and cholinesterase.

‡Just fails to meet 20% minimal important difference threshold for evidence of no difference.

## **Limitations of Systematic Review**

We relied mainly on electronic database searches and perusal of reference lists to identify relevant studies. Unpublished relevant studies may have been missed. We also kept the review focused on the evidence that most directly addresses the comparative effect of routine or per-protocol preoperative testing versus ad hoc or no testing. Thus, we did not review the wide range of indirect evidence from which conclusions about whether testing might be of value might be inferred. The Statement of Work section in the Introduction spells out the broader research questions that were not addressed here. The decision to narrow the scope of the review was made in part due to time and resource constraints. Future updates of this review may be able to broaden the scope of the research questions, particularly if it remains the case that there are few eligible comparative studies.

The conclusions, to a large extent, reflect the limitations of the underlying evidence base. Our ability to address most of the issues raised by the Key Questions was hampered by a paucity or complete lack of data, particularly from comparative studies.

## **Applicability**

In general, the applicability of the evidence is limited, with the exception of the studies of cataract surgery. The cataract RCTs all had similar findings, despite being conducted in different settings, in different countries, and with somewhat different eligibility criteria and study designs. Furthermore, the first trial was conducted in nearly 20,000 patients. This implies that the conclusion that there is no effect of routine testing with ECG, a basic metabolic panel, and blood counts for cataract surgery is likely to be broadly applicable. The applicability of the findings for adults undergoing a range of elective surgeries is less clear. The studies evaluated different tests in different populations receiving different surgical procedures and did not adequately report the conditions under which ad hoc testing was done (i.e., the extent of H&P or the triggers to order testing).

## **Evidence Gaps and Future Research**

Table B summarizes the evidence gaps with regard to the two Key Questions and subquestions of this systematic review.

**Table B. Evidence gaps**

<b>Key Question</b>	<b>Category</b>	<b>Evidence Gap</b>
Beneficial effects of routine or per-protocol preoperative testing	General	<ul style="list-style-type: none"> <li>For all procedures and surgeries requiring more than local anesthesia except cataract surgery, there is a paucity or lack of comparative studies to assess the value of the intervention.</li> </ul>
	Population	<ul style="list-style-type: none"> <li>Evidence is needed to evaluate the effect of testing for—                             <ul style="list-style-type: none"> <li>All elective procedures except cataract surgery</li> <li>Specific procedures</li> <li>Different types of anesthesia</li> <li>Different aged populations—children, adults, and older adults</li> <li>Different preoperative health status, including comorbidities</li> <li>Different categories of anesthesia risk</li> </ul> </li> <li>Existing studies generally provide poor descriptions of the patient populations—specific procedures planned, disease conditions, comorbidities, surgical and anesthesia risk categories, race, and other factors.</li> </ul>
	Interventions and comparators	<ul style="list-style-type: none"> <li>Difference in effect of routine testing (in all patients) vs. per-protocol testing (in selected patients).</li> <li>Effect of individual tests (within panels of tests) compared with effect of other individual tests.</li> <li>Different effects based on who ordered the test or the structure of testing (e.g., if done through a preanesthesia clinic or internist’s office). These data are generally not reported.</li> <li>How long prior to the planned procedure tests can be performed (e.g., within 1 week or 6-12 months) and still provide a benefit (assuming the preoperative testing is beneficial).</li> </ul>
	Outcomes	<ul style="list-style-type: none"> <li>Major perioperative complications (to some degree in contrast with total complications).</li> <li>Quality of life or satisfaction.</li> <li>Resource utilization.</li> <li>Postoperative management.</li> <li>Perioperative complications: improved standardization is needed regarding which perioperative complications should be reported; however, the list of complications will vary depending on the procedure.</li> </ul>
Harms of routine or per-protocol preoperative testing	General/outcomes	<ul style="list-style-type: none"> <li>There is no evidence regarding harms of testing.</li> </ul>
Subgroup analyses	General	<ul style="list-style-type: none"> <li>No comparative studies provided subgroup analyses based on any baseline patient characteristics, procedures, anesthesia type, or other factors listed above under Population or Interventions and comparators.</li> </ul>

For all procedures and surgeries requiring more than local anesthesia except cataract surgery, there is a paucity or lack of comparative studies to assess the value of the intervention. Evidence is needed to evaluate specific procedures and types of anesthesia, and specific populations, including patients at different surgical risk. Evidence is needed to compare routine testing versus per-protocol testing, the effect of individual tests, who orders and manages tests, and the timing of tests. Evidence is needed for all clinical outcomes, but it is particularly lacking for quality of life and satisfaction, resource utilization, and harms.

A large series of RCTs would best address the important research questions regarding routine and per-protocol preoperative testing. Focused studies evaluating specific tests or panels of tests in well-defined patients undergoing a narrow set of procedures will be of greater value to clinicians and decisionmakers deciding who should be routinely tested preoperatively than less focused studies. Conducting a series of such trials appears to be quite feasible, given the large number of elective procedures performed at many hospitals or surgical clinics; the low cost of the intervention (since in many situations the trial will primarily involve randomizing patients to either receive tests that are already available to them or to withhold those tests, as opposed to requiring resources to cover the costs of additional interventions); and the short term of the postoperative followup that is required (during hospitalization or up to 1 to 3 months). Trials should collect sufficient data to effectively stratify patients based on the major variables of interest (procedures, tests, comorbidities, etc.), or alternatively, multiple trials should each focus on a specific aspect of the research question. In particular, since it is likely that the effect of preoperative testing will vary substantially based on the specific surgery (as suggested by the different effects found between cataract trials and general surgery studies), trials should either focus on a single type of surgery or, at a minimum, stratify their results by surgery or surgery risk class. Furthermore, studies should stratify their results based on patient risk category, such as ASA category and comorbidities. Studies should capture the full range of perioperative outcomes, including patient quality of life/satisfaction and resource utilization. Studies should be sufficiently powered to evaluate, at a minimum, total major perioperative complications. Preferably they should be sufficiently powered to cover specific major complications, such as death. Also, preferably they should be sufficiently powered to allow for a priori subgroup analyses and analyses specific to at least some individual procedures and tests.

Observational studies can provide a lesser level of evidence to provide information on the comparative effectiveness of alternative preoperative testing strategies. However, the intrinsic heterogeneity and risk of confounding require that great care and attention be given to how the data are analyzed (e.g., with a priori subgroup analyses) and whether it is possible to adequately adjust for fundamental differences between nonrandomized cohorts of patients having or not having testing done. At a minimum, observational studies need to be adjusted for differences in patient and surgical characteristics and to control for cluster effects for individual surgeons or based on surgical experience. To be of use, observational studies should include concurrent patients who do or do not receive testing and who are as similar as possible. Even then, it will be important to use strong statistical methods to adjust analyses for differences in the cohorts unrelated to testing and confounders (e.g., propensity score or instrumental variable methods). All the suggestions made for RCTs regarding focusing or stratifying analyses based on surgical, patient, and other study characteristics also apply to observational studies.

In the face of a paucity of reliable evidence regarding the benefits, harms, and resources used with routine or per-protocol preoperative testing, decision analyses may be of value to delineate plausible estimates of the range of how beneficial or harmful and resource intensive preoperative testing could be. Such analyses could be useful to rank tests and procedures by likely benefit and thus help to prioritize research for specific tests and procedures. Such models will require direct evidence of the comparative effect of testing, as reviewed here, along with other indirect evidence, including the likelihood of specific perioperative complications for specific procedures, the likelihood that specific tests would diagnose conditions that would impact the rate of complications, the effects of correcting or ameliorating any such conditions, whether a test result could be acted on to impact the rate of complications, the likelihood of true- and false-positive test results, and the effects of delaying or canceling the procedures.

Regardless of the design of future studies, to allow answers about the value of routine or per-protocol preoperative testing, it is important that a large number of studies be conducted covering a wide range of scenarios, but that they be specific enough to allow applicability to decisionmaking for particular patients undergoing particular procedures in a given setting. Alternative prioritization approaches may be reasonable. Initially focusing on people who are most likely to have life-threatening perioperative complications, including older patients, those in higher ASA categories, those with important comorbidities, and those undergoing

higher risk surgeries, would allow for relatively small, low-resource studies that would be adequately powered. In these cases, complications would be more common and test abnormalities may also be more common. Not only would studies of these groups have the greatest potential to affect patients most likely to have complications, but the studies would also be better powered due to the higher complication rates than in lower risk populations. Further studies of patients at high risk of surgical bleeding (for example, children undergoing tonsillectomy and/or adenoidectomy) are also warranted. Alternatively, one could argue that future research should focus on lower risk populations and surgeries. While these studies would need to be relatively large due to low complication rates, the findings of these studies may have the greatest impact since they would address more common surgeries and more typical patients. Furthermore, hospitals, clinicians, and patients may be more willing to forgo preoperative testing in low- rather than high-risk settings. We believe it is likely that higher risk patients undergoing higher risk procedures would continue to have preoperative testing done regardless of evidence showing the testing to be ineffective. Given the different arguments that could be made about who to include in future studies and limited resources to conduct such research, this topic may be worthy of undergoing a formal value-of-information analysis.<sup>10</sup>

## Conclusions

With the exception of cataract surgery, there is a paucity of reliable evidence regarding the benefits, harms, and resource utilization associated with routine or per-protocol preoperative testing for all tests used for all procedures. There is a high strength of evidence, which is broadly applicable, that ECG, basic metabolic panel (biochemistry), and CBC have no effect on important clinical outcomes in patients scheduled for cataract surgery, including total perioperative complications and procedure cancellations. But despite several nonrandomized studies, there is insufficient evidence regarding the value of routine or per-protocol preoperative testing for other procedures and populations. Based on studies with a high risk of bias, there is a possibility that complications and deaths occurred more commonly among patients undergoing ad hoc as opposed to routine or per-protocol testing. This raises a caution against extrapolating the cataract findings to other surgeries and populations who may be at higher risk of complications due to the nature of the procedures or underlying illnesses and comorbidities. The evidence is insufficient to clarify specifically which

routinely conducted or per-protocol tests may be of benefit or no benefit for which patients undergoing which procedures. There is no evidence regarding quality of life or satisfaction, resource utilization, or harms of testing. There is also no evidence regarding how the value of testing may differ based on the risks of a specific surgical procedure; the type of anesthesia planned; the indication for surgery; comorbidities or other patient characteristics; the structure of testing (e.g., routine for everyone vs. per protocol, whether testing is conducted in a specialized preoperative clinic); who orders the tests (e.g., surgeon vs. anesthesiologist vs. primary care physician); or the length of time prior to the procedure that the tests are conducted. Given the large number of patients undergoing elective surgery, there is a clear need to develop better evidence for when routine or per-protocol testing improves patient outcomes and what the harms may be.

## References

1. Apfelbaum JL, Connis RT, Nickinovich DG, et al. Practice advisory for preanesthesia evaluation: an updated report by the American Society of Anesthesiologists Task Force on Preanesthesia Evaluation. *Anesthesiology*. 2012 Mar;116(3):522-38. PMID: 22273990.
2. Kumar A, Srivastava U. Role of routine laboratory investigations in preoperative evaluation. *J Anaesthesiol Clin Pharmacol*. 2011 Apr;27(2):174-9. PMID: 21772675.
3. Bryson GL. Has preoperative testing become a habit? *Can J Anaesth*. 2005 Jun;52(6):557-61. PMID: 15983138.
4. Kaplan EB, Sheiner LB, Boeckmann AJ, et al. The usefulness of preoperative laboratory screening. *JAMA*. 1985 Jun 28;253(24):3576-81. PMID: 3999339.
5. Johnson RK, Mortimer AJ. Routine pre-operative blood testing: is it necessary? *Anaesthesia*. 2002 Sep;57(9):914-7. PMID: 12190758.
6. Pasternak LR. Preoperative testing: moving from individual testing to risk management. *Anesth Analg*. 2009 Feb;108(2):393-4. PMID: 19151262.
7. MacPherson RD, Reeve SA, Stewart TV, et al. Effective strategy to guide pathology test ordering in surgical patients. *ANZ J Surg*. 2005 Mar;75(3):138-43. PMID: 15777393.
8. Klein AA, Arrowsmith JE. Should routine pre-operative testing be abandoned? *Anaesthesia*. 2010 Oct;65(10):974-6. PMID: 21198466.
9. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(11)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality; March 2011. Chapters available at [www.effectivehealthcare.ahrq.gov](http://www.effectivehealthcare.ahrq.gov).

10. Myers E, Sanders GD, Ravi D, et al. Evaluating the Potential Use of Modeling and Value-of-Information Analysis for Future Research Prioritization Within the Evidence-based Practice Center Program. (Prepared by the Duke Evidence-based Practice Center under Contract No. 290-2007-10066-I.) AHRQ Publication No. 11-EHC030-EF. Rockville, MD: Agency for Healthcare Research and Quality. June 2011. [www.effectivehealthcare.ahrq.gov/reports/final.cfm](http://www.effectivehealthcare.ahrq.gov/reports/final.cfm).

## **Full Report**

This executive summary is part of the following document:  
Balk EM, Earley A, Hadar N, Shah N, Trikalinos TA. Benefits and Harms of Routine Preoperative Testing: Comparative Effectiveness. Comparative Effectiveness Review No. 130. (Prepared by Brown Evidence-based Practice Center under Contract No. 290-2012-0012-I.) AHRQ Publication No. 14-EHC009-EF. Rockville, MD: Agency for Healthcare Research and Quality; January 2014. [www.effectivehealthcare.ahrq.gov/reports/final.cfm](http://www.effectivehealthcare.ahrq.gov/reports/final.cfm).

