A Randomized, Controlled Trial of the Effects of Remote, Intercessory Prayer on Outcomes in Patients Admitted to the Coronary Care Unit

William S. Harris, PhD; Manohar Gowda, MD; Jerry W. Kolb, MDiv; Christopher P. Strychacz, PhD; James L. Vacek, MD; Philip G. Jones, MS; Alan Forker, MD; James H. O’Keefe, MD; Ben D. McCallister, MD

Context: Intercessory prayer (praying for others) has been a common response to sickness for millennia, but it has received little scientific attention. The positive findings of a previous controlled trial of intercessory prayer have yet to be replicated.

Objective: To determine whether remote, intercessory prayer for hospitalized, cardiac patients will reduce overall adverse events and length of stay.

Design: Randomized, controlled, double-blind, prospective, parallel-group trial.

Setting: Private, university-associated hospital.

Patients: Nine hundred ninety consecutive patients who were newly admitted to the coronary care unit (CCU).

Intervention: At the time of admission, patients were randomized to receive remote, intercessory prayer (prayer group) or not (usual care group). The first names of patients in the prayer group were given to a team of outside intercessors who prayed for them daily for 4 weeks. Patients were unaware that they were being prayed for, and the intercessors did not know and never met the patients.

Main Outcome Measures: The medical course from CCU admission to hospital discharge was summarized in a CCU course score derived from blinded, retrospective chart review.

Results: Compared with the usual care group (n = 524), the prayer group (n = 466) had lower mean ± SEM weighted (6.35 ± 0.26 vs 7.13 ± 0.27; \(P = .04\)) and unweighted (2.7 ± 0.1 vs 3.0 ± 0.1; \(P = .04\)) CCU course scores. Lengths of CCU and hospital stays were not different.

Conclusions: Remote, intercessory prayer was associated with lower CCU course scores. This result suggests that prayer may be an effective adjunct to standard medical care.

Arch Intern Med. 1999;159:2273-2278

From time immemorial, prayer for the sick has been a common response to the illness of a loved one. In some societies and among certain religious groups, prayer is believed to be the most important therapy that can be offered to a sick person, superseding even medical intervention. Nevertheless, intercessory prayer (praying for others) has rarely been subjected to scientific scrutiny. In 1988, Byrd published the results of a blinded, controlled trial of 393 patients who had been admitted to a coronary care unit (CCU) at San Francisco General Hospital, San Francisco, Calif. Patients were randomly assigned to either a usual care group, which received no organized prayer, or to an experimental, intercessory prayer group, which received remote (from outside of the hospital) prayer from persons unknown to them. Byrd reported a statistically significant beneficial effect of intercessory prayer as assessed by a summary “hospital course” score. Three recent books on spirituality and healing have noted that the Byrd study is the only published trial of intercessory prayer with clinically significant end points, and that more scientifically valid (prospective, randomized, controlled, blinded, etc) studies of prayer were needed. The purpose of the present study was to attempt to replicate Byrd’s findings by testing the hypothesis that patients who are unknowingly and remotely prayed for by blinded intercessors will experience fewer complications and have a shorter hospital stay than patients not receiving such prayer.

RESULTS

INTERCESSORS

The intercessors represented a variety of Christian traditions, with 35% listing their affiliations as nondenominational, 27% as Episcopalian, and the remainder as other Protestant groups or Roman Catholic.
METHODS

PATIENTS AND PROTOCOL

All patients admitted to the CCU at the Mid America Heart Institute (MAHI), Kansas City, Mo, over a 12-month period were eligible for the trial (Figure). The only exceptions were those admitted for workup and wait-listing prior to cardiac transplantation (because of anticipated prolonged stays). Patients admitted for less than 1 day were subsequently excluded because it took up to 24 hours for intercessors to be contacted and prayer initiated. New admissions were identified in the chaplain’s office on a daily basis via computer. The chaplain’s secretary randomly assigned all new patients to either the usual care or prayer group based on the last digit of the medical record number; even numbers were assigned to the prayer group and odd numbers to the usual care group. This allocation scheme allowed no opportunity for bias because medical record numbers are assigned on a sequential basis to all new patients entering the hospital, regardless of how sick they are. In addition, since some patients were readmitted (having been assigned their numbers months to years previously) and some were newly admitted, no systematic assignment of the sickest patients to the odd (usual care) group was possible. Once assigned, the secretary called an intercessory prayer team leader and gave him/her the first name of the patient to be prayed for. No other information (eg, diagnosis, prognosis, age, race, socioeconomic status, or family situation) was available to the secretary; thus, it was not passed on to the intercessors. The secretary was the only person with knowledge of the assignment code, and she had no contact with the patients, the CCU staff (she did not even know where the unit was located within the hospital), the data collectors, or the statistician, all of whom were blinded throughout. After receiving the call from the secretary, the prayer team leader called the other 4 persons on his/her team and directed that the name of the new patient be entered on a log sheet provided. The intercessors were asked to pray daily for the next 28 days for “a speedy recovery with no complications” and anything else that seemed appropriate to them. A period of 28 days was chosen to ensure that prayer would continue throughout the entire hospitalization of at least 95% of patients. Some CCU patients (typically fewer than 5%) request prayer from the hospital chaplain’s staff upon admission to the hospital. When made, these requests were always honored regardless of and without knowledge of group assignment. This study was approved by the hospital’s institutional review board (IRB) and, in order to keep the study blinded, was exempted from the requirement to obtain informed consent (see the “Comment” section).

INTERCESSORS

The intercessors were recruited by the investigators via contacts in the local community. In order to be an intercessor, an individual did not need to be of any particular denomination, but he/she did need to agree with the following statements: “I believe in God. I believe that He is personal and is concerned with individual lives. I further believe that He is responsive to prayers for healing made on behalf of the sick.” Once identified, the intercessors were organized into 13 teams of 5 members (a total of 75), each with 1 person designated as the team leader. Intercessors were randomly assigned to teams; those within a given team did not know the others in the same team, and prayer was offered individually, not in groups.

DATA COLLECTION

Patient demographics and admission diagnoses were obtained from the hospital computer system. All patient charts were reviewed retrospectively by a blinded physician/investigator to collect information regarding comorbid conditions at the time of admission, length of CCU and hospital stays. Intercessors were predominantly women (87%), and their mean age was 56 years. All reported at least weekly church attendance and daily prayer habits (prior to the study). A review of intercessor log sheets indicated that prayer (by at least 1 intercessor) began within 1.2 ± 0.05 days after admission to the CCU. All intercessors who were ultimately going to pray for a given patient began doing so within 1.6 ± 0.16 days after CCU admission.

PATIENTS

A total of 1019 patients were admitted to the CCU during the period of the trial. After elimination of 6 patients who were waiting for cardiac transplantation, 1013 were randomized (Figure), 484 (48%) to the prayer group and 529 (52%) to the usual care group. This difference in sample sizes was most likely caused by chance (P = .18). After subsequent removal of those patients who spent less than 24 hours in the CCU, 524 remained in the usual care group and 466 in the prayer group. Comorbid conditions upon admission were similar for each group (Table 2). Men and women were equally represented in the usual care and prayer groups (66% vs 61% men, respectively; P = .10), and the mean age was 66 years for both groups.

OUTCOMES

The primary predefined end point in this trial was the weighted MAHI-CCU score (Table 4). We found an 11% reduction in scores in the prayer group (6.35 ± 0.26) compared with the usual care group (7.13 ± 0.27) (P = .04). Using the unweighted MAHI-CCU score, which simply counted elements in the original scoring system without assigning point values, the prayer group had 10% fewer elements (P = .04) than the usual care group. There were no statistically significant differences between groups for any individual component of the MAHI-CCU score (Table 3). Mean lengths of stay in the CCU and in the hospital (after initiation of prayer) were not different (Table 4), and median hospital stay was 4.0 days for both groups. There were 2 patients in the prayer group whose hospital stays were approximately twice as long (137 and 161 days) as those of any other patient in the study. Without these 2 patients, length of hospital stay for the prayer group dropped from 6.48 ± 0.54 days to 5.84 ± 0.31 days.
stay, and clinical outcomes. The latter were defined as all new diagnoses, events, or procedures occurring at least 24 hours after admission to the CCU (to allow time for organized prayer to begin) until discharge or death. Thus, if a patient who presented to the emergency department with an acute myocardial infarction was catheterized, revascularized, and then admitted to the CCU, these events/procedures were not recorded as new CCU events. On the other hand, if, after the first day in the CCU, a patient developed unstable angina, had a coronary angiogram, and had a subsequent revascularization procedure, all of these were recorded as new events.

CLINICAL OUTCOMES

Since prayer was offered for a speedy recovery with no complications, it was anticipated that the effect of prayer was unlikely to be evident in any specific clinical outcome category (eg, the need for antibiotics, the development of pneumonia, or the extension of infarction), but would only be seen in some type of global score. Review of the medical literature revealed no previously validated and standardized statistic to quantitate severity of outcomes in critically ill cardiovascular patients. Severity of illness or comorbidity scales, such as the Acute Physiology and Chronic Health Evaluation (APACHE) score and Charlson scale, do exist, but these are prognostic tools designed to predict major health outcomes for individual patients; they are not designed to summarize a CCU course. Accordingly, before the trial began, 3 experienced cardiologists and 1 internist from MAHI and the University of Missouri–Kansas City School of Medicine developed a weighted and summed scoring system called the MAHI-CCU score (Table 1). The MAHI-CCU score is a continuous variable that attempts to describe outcomes from excellent to catastrophic. For example, if, after the first day in the CCU, a patient developed unstable angina (1 point), was treated with antianginal agents (1 point), was sent for heart catheterization (1 point), underwent unsuccessful revascularization by percutaneous transluminal coronary angioplasty (3 points), and went on to coronary artery bypass graft surgery (4 points), his weighted MAHI-CCU score would be 10. Another patient might have developed a fever and received antibiotic treatment (1 point) but experienced no other problems and been discharged from the hospital with a score of 1. A third patient might have suffered a cardiac arrest (5 points) and died (6 points), for a total weighted score of 11 points. In addition to the weighted MAHI-CCU scores, a nonweighted MAHI-CCU score was calculated that was simply a count of events, procedures, and/or prescriptions after CCU admission. For the examples above, the unweighted MAHI-CCU scores would have been 5, 1, and 2, respectively. To evaluate interrater reproducibility for the MAHI-CCU score, 10 physicians (5 cardiologists and 5 cardiology fellows) blindly scored 11 randomly selected CCU patient charts. The raters were in agreement (mean ± SD) 96% ± 3% of the time. Finally, for comparison, the Hospital Course Score used by Byrd was also calculated. The Byrd score broadly categorizes each patient's progress after CCU admission as good, intermediate, or bad.

STATISTICAL ANALYSIS

Baseline variables and specific medical outcomes were analyzed by χ² analysis and the Fisher exact test for categorical data. Byrd scores were analyzed by the Cochran–Armitage trend test; t tests were used to compare continuous variables (eg, age, length of stay, and MAHI-CCU scores). A difference with a 2-tailed P < .05 was accepted as statistically significant, except for comorbid conditions upon admission (Table 2) and individual events/procedures occurring during the CCU stay (Table 3). For these 2 data sets, P < .005 was required for statistical significance because of the multiple comparisons evaluated. Data are presented as means ± SEs. All analyses were carried out blindly on an intention-to-treat basis using SAS, version 6.12 (SAS Institute, Cary, NC).

Neither was significantly different from the length of stay in the usual care group (5.97 ± 0.29 days). There was no significant difference between groups using Byrd's hospital course score (Table 5).

Using a severity-adjusted outcomes score, we found lower overall adverse outcomes for CCU patients randomized to the prayer group compared with those randomized to the usual care group. Lengths of CCU stay and hospital stay after initiation of prayer were not affected. These findings are consistent with those of Byrd, who reported that intercessory prayer for hospitalized patients lowered the hospital course score but did not significantly affect length of stay.

Although there was a trend toward better outcomes in the prayer group using the Byrd score, the difference between groups was not statistically significant. Other than the fact that it is a categorical instead of a continuous statistic, we have no explanation as to why the Byrd score did not detect a difference between groups and the MAHI-CCU score did. There were, however, several important differences between the 2 study designs that may have contributed to this discrepancy. First, the present study was conducted under completely blinded conditions, with neither patients nor medical staff aware that a study was being conducted. In Byrd's trial, the staff and patients were fully aware that the study was in progress, although nobody knew which patients were receiving “study” prayer. Another difference was in the kinds of patients enrolled. In the present trial, informed consent was not sought and thus patients were not prescreened for their willingness to be prayed for. Of the 450 patients invited to participate in the Byrd study, 57 (12.7%) refused to do so “for personal reasons or religious convictions” or were otherwise unwilling to give consent. This indicates that only “prayer-receptive” patients were included in his final cohort. Finally, in Byrd's study, the intercessors were given a considerable amount of information about the patient (eg, diagnoses, general conditions, and updates as their status changed), and they prayed only until the patient left the unit. These factors could have produced a heightened intensity of or commitment to prayer in Byrd's intercessors. In contrast, our intercessors were asked to pray for 28 days re-
and no known risk for the patients in the usual care group risk associated with receiving remote, intercessory prayer, but to the very existence of the trial. This was possible because the hospital’s IRB granted the study an exemption from the requirement to obtain informed consent. Since this may be viewed as problematic by some, the reasons supporting this decision will be discussed in some detail. First, it was agreed that there was no known risk associated with receiving remote, intercessory prayer, and no known risk for the patients in the usual care group associated with not receiving extra prayer. Second, no additional data were collected on the patients in this study beyond those that are normally collected for all patients in the hospital. Third, and perhaps most important, the very process of obtaining informed consent could conceivably have caused increased anxiety in some patients. For example, had they known about the study, the possibility of not being in the prayer group might have greatly distressed some patients. For nonreligious or antireligious patients, having to accept or reject the offer of prayer (especially considering the gravity of their illness) might have been very challenging. The policy of the US Department of Health and Human Services for the protection of human subjects states that the IRB may waive the requirement for the investigator to obtain a signed consent form for some or all subjects if it finds that the research presents no more than minimal risk of harm to the subjects and involves no procedures for which written consent is normally required outside of the research context.9

Scientifically, a study design with complete blinding was preferred because it eliminated any possibility of bias, and enrolling all patients in the study increased its generalizability. In light of all these factors, an exemption was granted.

Overall distribution of patients.

Table 1. Mid America Heart Institute–Cardiac Care Unit (MAHI-CCU) Scoring System

<table>
<thead>
<tr>
<th>MAHI-CCU Score</th>
<th>Comorbid Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Need for antianginal agents, antibiotics, arterial monitoring, or catheterization; development of unstable angina</td>
</tr>
<tr>
<td>2</td>
<td>Need for antiarrhythmic, inotropic, diuretic, or vasodilator drugs; development of pneumonia, atrial fibrillation, supraventricular tachycardia, hypotension, or anemia requiring a transfusion</td>
</tr>
<tr>
<td>3</td>
<td>Need for a temporary pacemaker; Swan-Ganz catheterization, an implanted cardiac defibrillator, an electrophysiology study, radiofrequency ablation, or an interventional coronary procedure (ie, a percutaneous transluminal coronary angioplasty); development of third-degree heart block, extension of infarct, or gastrointestinal bleed; or readmission to the cardiac care unit</td>
</tr>
<tr>
<td>4</td>
<td>Need for a permanent pacemaker, an intra-aortic balloon pump, major surgery (of any kind), percutaneous transluminal coronary angioplasty with stent placement and/or rotablator, or intubation/ventilation; development of congestive heart failure, ventricular tachycardia, ventricular fibrillation, or sepsis</td>
</tr>
<tr>
<td>5</td>
<td>Cardiac arrest</td>
</tr>
<tr>
<td>6</td>
<td>Death</td>
</tr>
</tbody>
</table>

Downloaded from www.archinternmed.com at Oxford University Library Services, on November 29, 2010
©1999 American Medical Association. All rights reserved.
In evaluating the results of this trial, it is important to note that we were most likely studying the effects of supplementary intercessory prayer. Since at least 50% of patients admitted to this hospital state that they have a religious preference, it is probable that many if not most patients in both groups were already receiving intercessory prayer from friends, family, and clergy during their hospitalization. Thus, there was an unknowable and/or direct prayer from friends, family, and clergy to note that we were most likely studying the effects of intercessory prayer to “real” but currently unknown physiological forces that are “generated” by the intercessors and “received” by the patients; the latter explanation would be, by definition, beyond the ken of science. However, this trial was designed to explore not a mechanism but a phenomenon. Clearly, proof of the latter must precede exploration of the former. By analogy, when James Lind, by clinical trial, determined that lemons and limes cured scurvy aboard the HMS Salisbury in 1753, he not only did not know about ascorbic acid, he did not even understand the concept of a “nutrient.” There was a natural explanation for his findings that would be clarified centuries later, but his inability to articulate it did not invalidate his observations.

Although we cannot know why we obtained the results we did, we can comment on what our data do not show. For example, we have not proven that God answers prayer or that God even exists. It was intercessory prayer, not the existence of God, that was tested here. All we have observed is that when individuals outside of the hospital speak (or think) the first names of hospitalized patients with an attitude of prayer, the latter explanation would be, by definition, beyond the ken of science. However, this trial was designed to explore not a mechanism but a phenomenon. Clearly, proof of the latter must precede exploration of the former. By analogy, when James Lind, by clinical trial, determined that lemons and limes cured scurvy aboard the HMS Salisbury in 1753, he not only did not know about ascorbic acid, he did not even understand the concept of a “nutrient.” There was a natural explanation for his findings that would be clarified centuries later, but his inability to articulate it did not invalidate his observations.

Neither this study nor that of Byrd provided any explanation for his findings that would be clarified or understood. There was a natural explanation for his findings that would be clarified or understood. There was a natural explanation for his findings that would be clarified or understood. There was a natural explanation for his findings that would be clarified or understood. There was a natural explanation for his findings that would be clarified or understood. There was a natural explanation for his findings that would be clarified or understood. There was a natural explanation for his findings that would be clarified or understood.

In the absence of any natural or supernatural explanations, the former explanation would attribute the beneficial effects of intercessory prayer to “real” but currently unknown physiological forces that are “generated” by the intercessors and “received” by the patients; the latter explanation would be, by definition, beyond the ken of science. However, this trial was designed to explore not a mechanism but a phenomenon. Clearly, proof of the latter must precede exploration of the former. By analogy, when James Lind, by clinical trial, determined that lemons and limes cured scurvy aboard the HMS Salisbury in 1753, he not only did not know about ascorbic acid, he did not even understand the concept of a “nutrient.” There was a natural explanation for his findings that would be clarified centuries later, but his inability to articulate it did not invalidate his observations.

Although we cannot know why we obtained the results we did, we can comment on what our data do not show. For example, we have not proven that God answers prayer or that God even exists. It was intercessory prayer, not the existence of God, that was tested here. All we have observed is that when individuals outside of the hospital speak (or think) the first names of hospitalized patients with an attitude of prayer, the latter appeared to have a “better” CCU experience. Although our findings would be expected to occur by chance alone only 1 out of 25 times that such an experiment was conducted, chance still remains a possible explanation of our results.

Interest in alternative or complementary medicine is growing rapidly in this country,11,12 and prayer “therapy” falls into this category. Two recent books13,14 have fo-

**Table 3. Effects of Intercessory Prayer on Individual Components of the Mid America Heart Institute–Cardiac Care Unit (MAHI-CCU) Score**

<table>
<thead>
<tr>
<th>MAHI-CCU Score Component</th>
<th>Usual Care Group (n = 524)</th>
<th>Prayer Group (n = 466)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antianginal agents</td>
<td>59 (11.3)</td>
<td>47 (10.1)</td>
<td>.62</td>
</tr>
<tr>
<td>Antibiotics</td>
<td>82 (15.6)</td>
<td>77 (16.5)</td>
<td>.77</td>
</tr>
<tr>
<td>Unstable angina</td>
<td>4 (0.8)</td>
<td>1 (0.2)</td>
<td>.38</td>
</tr>
<tr>
<td>Anterior/ventricular</td>
<td>42 (8.0)</td>
<td>32 (6.9)</td>
<td>.57</td>
</tr>
<tr>
<td>Catheterization</td>
<td>180 (34.4)</td>
<td>162 (34.8)</td>
<td>.94</td>
</tr>
<tr>
<td>Antiarrhythmics</td>
<td>56 (10.7)</td>
<td>50 (10.7)</td>
<td>.94</td>
</tr>
<tr>
<td>Inotropes</td>
<td>78 (14.5)</td>
<td>69 (14.8)</td>
<td>.96</td>
</tr>
<tr>
<td>Vasodilation</td>
<td>78 (14.9)</td>
<td>59 (12.7)</td>
<td>.36</td>
</tr>
<tr>
<td>Diuretics</td>
<td>112 (21.4)</td>
<td>97 (20.8)</td>
<td>.89</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>10 (1.9)</td>
<td>12 (2.6)</td>
<td>.62</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>17 (3.2)</td>
<td>12 (2.6)</td>
<td>.66</td>
</tr>
<tr>
<td>Supraventricular tachycardia</td>
<td>6 (1.1)</td>
<td>2 (0.4)</td>
<td>.29</td>
</tr>
<tr>
<td>Hypotension</td>
<td>7 (1.3)</td>
<td>8 (1.7)</td>
<td>.82</td>
</tr>
<tr>
<td>Anemia/transfusion</td>
<td>66 (12.6)</td>
<td>50 (10.7)</td>
<td>.42</td>
</tr>
<tr>
<td>Temporary pacor</td>
<td>16 (3.0)</td>
<td>13 (2.8)</td>
<td>.95</td>
</tr>
<tr>
<td>Third-degree heart block</td>
<td>1 (0.2)</td>
<td>2 (0.4)</td>
<td>.60</td>
</tr>
<tr>
<td>Readmit to cardiac unit</td>
<td>22 (4.2)</td>
<td>25 (5.4)</td>
<td>.48</td>
</tr>
<tr>
<td>Swan-Ganz catheter</td>
<td>172 (32.8)</td>
<td>123 (26.4)</td>
<td>.03</td>
</tr>
<tr>
<td>Implanted cardiac defibrillator</td>
<td>6 (1.1)</td>
<td>10 (2.1)</td>
<td>.32</td>
</tr>
<tr>
<td>Electrophysiology study</td>
<td>15 (2.9)</td>
<td>10 (2.1)</td>
<td>.61</td>
</tr>
<tr>
<td>Radiofrequency ablation</td>
<td>8 (1.5)</td>
<td>2 (0.4)</td>
<td>.11</td>
</tr>
<tr>
<td>Extension of infarct</td>
<td>19 (3.8)</td>
<td>0 (0.0)</td>
<td>.50</td>
</tr>
<tr>
<td>Gastrointestinal bleed</td>
<td>12 (2.3)</td>
<td>5 (1.1)</td>
<td>.22</td>
</tr>
<tr>
<td>Interventional coronary procedure</td>
<td>155 (29.6)</td>
<td>121 (26.0)</td>
<td>.31</td>
</tr>
<tr>
<td>PTCA alone</td>
<td>69 (13.2)</td>
<td>62 (13.3)</td>
<td>.95</td>
</tr>
<tr>
<td>PTCA with stent and/or rotablator</td>
<td>86 (16.4)</td>
<td>59 (12.7)</td>
<td>.10</td>
</tr>
<tr>
<td>Permanent pacer</td>
<td>21 (4.0)</td>
<td>12 (2.6)</td>
<td>.28</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>17 (3.2)</td>
<td>19 (4.1)</td>
<td>.60</td>
</tr>
<tr>
<td>Ventricular fibrillation/tachycardia</td>
<td>12 (2.3)</td>
<td>10 (2.1)</td>
<td>.95</td>
</tr>
<tr>
<td>Intra-aortic balloon pump</td>
<td>20 (3.8)</td>
<td>12 (2.6)</td>
<td>.36</td>
</tr>
<tr>
<td>Major surgery</td>
<td>76 (14.5)</td>
<td>51 (10.9)</td>
<td>.11</td>
</tr>
<tr>
<td>Sepsis</td>
<td>7 (1.3)</td>
<td>7 (1.5)</td>
<td>.96</td>
</tr>
<tr>
<td>Intubation/ventilation</td>
<td>27 (5.2)</td>
<td>26 (5.6)</td>
<td>.88</td>
</tr>
<tr>
<td>Cardiac arrest</td>
<td>6 (1.1)</td>
<td>5 (1.1)</td>
<td>.84</td>
</tr>
<tr>
<td>Death</td>
<td>46 (8.8)</td>
<td>42 (9.0)</td>
<td>.99</td>
</tr>
</tbody>
</table>

*PTCA indicates percutaneous transluminal coronary angioplasty.

**Table 4. Effects of Intercessory Prayer on Mid America Heart Institute–Cardiac Care Unit (MAHI-CCU) Scores and Length of Stay in the CCU and in the Hospital**

<table>
<thead>
<tr>
<th></th>
<th>Usual Care Group (n = 524)</th>
<th>Prayer Group (n = 466)</th>
<th>Percentage Change</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAHI-CCU score</td>
<td>7.13 ± 0.27</td>
<td>6.35 ± 0.26</td>
<td>−11 .04</td>
<td></td>
</tr>
<tr>
<td>Unweighted MAHI-CCU</td>
<td>3.00 ± 0.10</td>
<td>2.70 ± 0.10</td>
<td>−10 .04</td>
<td></td>
</tr>
<tr>
<td>Length of CCU stay, d†</td>
<td>1.23 ± 0.09</td>
<td>1.12 ± 0.08</td>
<td>−9 .28</td>
<td></td>
</tr>
<tr>
<td>Length of hospital stay, d‡</td>
<td>5.97 ± 0.29</td>
<td>6.48 ± 0.54</td>
<td>+9 .41</td>
<td></td>
</tr>
</tbody>
</table>

‡CCU indicates cardiac care unit.
†A simple count of events (diagnoses, drugs prescribed, and procedures) from the MAHI-CCU score (Table 1), presented as events per patient.
‡Length of stay was determined for a period beginning 1 day after admission to the CCU (ie, the day prayer began) until CCU/hospital discharge.

**Table 5. Effects of Remote, Intercessory Prayer on Byrd Scores**

<table>
<thead>
<tr>
<th></th>
<th>Usual Care Group (n = 524)</th>
<th>Prayer Group (n = 466)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>338 (64.5)</td>
<td>314 (67.4)</td>
<td>.29</td>
</tr>
<tr>
<td>Intermediate</td>
<td>71 (13.5)</td>
<td>63 (13.5)</td>
<td>.11</td>
</tr>
<tr>
<td>Bad</td>
<td>115 (21.9)</td>
<td>89 (19.1)</td>
<td>.29</td>
</tr>
</tbody>
</table>

*P = .29 by Cochran-Armitage trend test.
cused on the health benefits of a patient's own spiritual orientation. Each has documented that church membership/attendance is associated with improved medical outcomes.\textsuperscript{13-15} People who believe in God and pray during illness have been reported to have better health outcomes than people who do not.\textsuperscript{16-18} For some, faith is an effective means of stress reduction, which has itself been shown to reduce cardiac morbidity.\textsuperscript{19} Some of these benefits may derive from favorable hormonal, autonomic, and immunologic responses to the emotional reassurance that belief can provide. Nevertheless, the present trial was designed to study the impact not of personal spirituality, but of prayer offered for patients regardless of their spiritual orientation.

Other studies besides Byrd's have explored the impact of intercessory prayer on health outcomes. O’Leari\textsuperscript{22} examined the effects of intercessory prayer on self-esteem, anxiety, and depression in 406 subjects (who received either no prayer, directed prayer, or nondirected prayer) and in the 90 intercessors. There were no specific benefits detected for the prayer groups. A pilot study of the effects of intercessory prayer on 40 recovering alcoholics likewise reported no clinical benefit.\textsuperscript{23} Finally, in a 6-month trial of “distant healing” in patients with acquired immune deficiency syndrome, Sicher et al\textsuperscript{24} found statistically significant benefits for the intervention group (fewer new illnesses, physician visits, hospitalizations, and days of hospitalization; lower illness severity scores; and improved mood scores). These studies illustrate the broadening scope of interest in remote therapies and suggest that scientifically valid, properly controlled studies can be carried out in this emerging arena.

The principal limitation of this study was defining the end point measure (ie, determining how to quantify how well a patient did during a CCU stay). The score we devised, although intuitive and evenly applied to both groups, has not been validated. (It should be noted that the Byrd score is likewise an unvalidated measure of CCU outcomes.) It is not immediately obvious how any score could be validated given the fact that there is no known criterion standard summary statistic with which we could compare the MAHI-CCU score. The fact that there were significantly fewer total events in the prayer group suggests that the observed difference between groups was not an artifact of the scoring system. Another limitation lies in interpreting the clinical significance of a 10% difference in MAHI-CCU scores. Since the score itself is only an estimate of overall CCU course, there is no known way to ascribe a clinical significance to it, other than to say that as a group, the patients in the prayer group “did 10% better.” The score should be viewed only as a summary statistic designed to detect the impact of a mild global intervention on overall health in large groups, not in individual patients.

In conclusion, using the MAHI-CCU scoring system, we found that supplementary, remote, blinded, intercessory prayer produced a measurable improvement in the medical outcomes of critically ill patients. Our findings support Byrd's conclusions despite the fact that we could not document an effect of prayer using his scoring method. With 2 randomized, controlled trials now suggesting the possible benefits of intercessory prayer, further studies using validated and standardized outcome measures and variations in prayer strategy are warranted to explore the potential role of prayer as an adjunct to standard medical care.

Accepted for publication May 3, 1999.

The authors are indebted to Tracy Isaacs, the secretary to Chaplain Koll, for assigning patients to study groups and contacting the intercessors and to Shubha Gowda, MD, for helping with chart review. The support of the associate chaplains at Saint Luke's Hospital was also much appreciated. The encouragement and support of Richard G. Hastings, Douglas Willhoite, MD, and Robert and Marie Evans were greatly appreciated, as was the critical review provided by John Speritis, MD. Finally, we are indebted to the intercessors for faithfully praying for the patients in this study.

Corresponding author: William S. Harris, PhD, Lipid Research Laboratory, Saint Luke's Hospital, 4401 Wornall Rd, Kansas City, MO 64111 (e-mail: wharris@st-lukes.org).

REFERENCES

cluded as an event in the score calculation but was omitted from Tables 1 and 3 in our article.

See Correction below

Smith and Fisher are correct in noting that an even-odd medical record number randomization scheme is less than optimal; in future trials, we would use, as they suggest, a system that is more impervious to detection. Nevertheless, there is little room for subjectivity in a chart review method that simply records the presence or absence of a set of predetermined events. Thus, we do not believe that our findings were biased by this approach. These writers also raise the issue of “file-drawer bias,” ie, the reluctance of some investigators to publish no-effect studies. We clearly have no control over what others may have done, and while this charge can be leveled at any field of inquiry, the fact that in this very young field several studies with negative findings have been published argues against such bias. We hope that most investigators, in addressing an important question and having designed their study to the best of their abilities, would make (as we did) an a priori commitment to publish their results regardless of outcome for the good of the overall scientific enterprise.

Several letters raised questions regarding the theological implications of our study. As we noted in our article, we cannot draw any conclusions regarding the existence or nature of God from this trial.

A critically important attribute of any scientist is open-mindedness, the willingness to objectively consider new or alternative concepts and hypotheses. There is a growing demand among patients that we acknowledge their need to be treated as whole persons who have not only physical but emotional and spiritual needs as well. Practicing as we do in a large metropolitan hospital among a wide variety of religious traditions, we are acutely sensitive to the need for a nonsectarian approach to addressing spiritual issues. This diversity is mirrored in the spectrum of religious practices among our authors, which ranged from a variety of Protestant and Roman Catholic traditions to Hinduism. Since spiritual factors may play some role in healing, additional studies are needed to clarify the place of intercessory prayer in maintaining and restoring health.

William S. Harris, PhD
Manohar Gowda, MD
Jerry W. Kolb, MDiv
Christopher P. Strychacz, PhD
James L. Vacch, MD
Philip G. Jones, MS
Alan Forker, MD
James H. O’Keefe, Jr, MD
Ben D. McCallister, MD
Kansas City, Mo

Errors in Results. In the Original Investigation titled “A Randomized, Controlled Trial of the Effects of Remote, Intercessory Prayer on Outcomes in Patients Admitted to the Coronary Care Unit,” published in the October 25, 1999, issue of the ARCHIVES (1999;159:2273-2278), the authors, Harris et al, were prompted by questions raised in postpublication correspondence to reevaluate their calculations and feel that 2 points need to be clarified. In Table 3 of their article, a percutaneous transluminal coronary angioplasty procedure (PTCA) with a stent and/or a rotablator appeared to count as one event. However, when they calculated the unweighted score, they gave one point for PTCA and an additional point for stent and one for rotablator when these occurred in the same patient. Thus, a patient receiving all 3 procedures was given 3 points, not 1, as was implied in Table 3. Second, the need for a cardiovascular stress test (such as a thallium test or an echocardiogram) was included in the calculation of the Mid American Heart Institute—Cardiac Care Unit (MAHI-CCU) scores but was omitted from Tables 1 and 3 of their article. There were 4+ of these events in the usual care group (8.4%) and 26 (5.6%) in the prayer group (P=.11). The following tabulation clarifies how Harris et al arrived at the scores reported in Table 4:

<table>
<thead>
<tr>
<th>Usual Care Group</th>
<th>Prayer Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum of points from Table 3 as published</td>
<td>1436</td>
</tr>
<tr>
<td>Extra points for PTCA + stent</td>
<td>79</td>
</tr>
<tr>
<td>Extra points for PTCA + rotablator</td>
<td>5</td>
</tr>
<tr>
<td>Extra points for PTCA + stent + rotablator</td>
<td>4*</td>
</tr>
<tr>
<td>Cardiovascular stress test</td>
<td>44</td>
</tr>
<tr>
<td>Total events</td>
<td>1568</td>
</tr>
<tr>
<td>No. of patients</td>
<td>524</td>
</tr>
<tr>
<td>Unweighted MAHI-CCU score as published</td>
<td>3.0</td>
</tr>
</tbody>
</table>

*Two patients × 2 extra points each.
†P=.04.

In the calculation of the weighted MAHI-CCU score, the need for cardiovascular stress tests was ranked as a category 4 event; if reclassified as a category 2 event, the mean ± SEM scores become 6.97 ± 0.26 for the usual care group and 6.24 ± 0.26 for the prayer group (P = .05); the effect size remains 10% to 11%.

In Table 4, the number of patients in the Usual Care Group was incorrectly reported as “(n = 52)”; it should have been “(n = 524).”

In Table 4, the number of patients in the Usual Care Group was incorrectly reported as “(n = 52)”; it should have been “(n = 524).”

ABSTRACT
Objective To determine whether the author’s 20.9 lb (9.5 kg) carbon frame bicycle reduced commuting time compared with his 29.75 lb (13.5 kg) steel frame bicycle.

Design Randomised trial.

Setting Sheffield and Chesterfield, United Kingdom, between mid-January 2010 and mid-July 2010.

Participants One consultant in anaesthesia and intensive care.

Main outcome measure Total time to complete the 27 mile (43.5 kilometre) journey from Sheffield to Chesterfield Royal Hospital and back.

Results The total distance travelled on the steel frame bicycle during the study period was 809 miles (1302 km) and on the carbon frame bicycle was 711 miles (1144 km). The difference in the mean journey time between the steel and carbon bicycles was 00:00:32 (h:min:sec; 95% CI –00:03:34 to 00:02:30; P=0.72).

Conclusions A lighter bicycle did not lead to a detectable difference in commuting time. Cyclists may find it more cost effective to reduce their own weight rather than to purchase a lighter bicycle.

INTRODUCTION
I have always been keen on cycling. As a child in the 1970s, a student in the 1980s, and a junior doctor in the 1990s, my prime means of local transport was a bicycle. An accident, shortly after taking up a registrar job in Sheffield, wrote off my bike and led me to turn to the internal combustion engine. However, after a number of years as a consultant, peer pressure and the desire to improve my fitness led to a decision to return to the saddle.

I acquired a second hand steel frame bike for £50, spruced it up, and set off. I soon got into the swing of cycling the 27 miles (43.5 kilometres) from home in Sheffield, United Kingdom, to work in Chesterfield and back, managing it most days when I wasn’t on call and didn’t have commitments off site. After about six months of commuting I began to wonder whether the one way journey time of about 55 minutes could be reduced. Those in the know suggested a new bike could knock 10% off it.

Evidence based cycling is not high on the bicycle salesman’s agenda. No one will tell you how much more efficient one bicycle is over another; they just say it is better. Making a decision on what was perceived to be best and dreaming of extra time in bed, I looked into the UK government’s Cycle to Work scheme. This scheme allows an employee to purchase a bicycle (up to a cost of £1000 (£1180; $1560)) at a significant discount by using tax incentives, provided the bicycle is used for commuting to and from work. The initiative aims to “promote healthier journeys to work and reduce environmental pollution.” However, doubt has been expressed in the popular press regarding whether the new generation of middle aged men in lycra (MAMILs) are actually using their scheme funded bikes to commute or just to gum up the roads (particularly hills) at weekends. The benefits are debatable but attractive, and the scheme has encouraged a lot of people to spend a lot of money on high end bicycles. I purchased a bike at the top end of the cost allowed by the scheme and opted for a carbon frame because it was significantly lighter than my existing bicycle’s steel frame. The wheels were lighter and tyres narrower too. All were factors that made me believe that the extra £950 I had spent would get me to work in a trice.

My new bike seemed wonderful, if somewhat uncomfortable. I didn’t notice a dramatic decrease in commuting time, nor did the cycle computer I had fitted to my new bicycle to record any notably swift journeys. But, one sunny morning, I got to work in 43 minutes, the fastest I could recall. My steel bike was consigned to a corner of the garage to gather dust—until I had a puncture. The next day I was back on my old steel bike. I fitted the cycle computer, set off . . . and discovered I had got to work in 44 minutes. “Hang on,” I thought, “was that minute worth £950 or was it a fluke?” There was only one answer: a randomised trial. I toyed with the idea of blinding it but, in the interest of self preservation and other road users, decided against it.

METHODS
This was a single centre, randomised, non-blinded trial; n=1. Both bicycles were of traditional “road” construction with drop handlebars, although the frame of one was made of steel and the second carbon (table 1; fig 1). Identical lights and fittings were used on each bike.

Between mid-January 2010 and mid-July 2010, either the steel frame bicycle (29.75 lb (13.5 kg)) or the carbon bicycle (20.9 lb (9.5 kg)) was randomly

The initiative aims to “promote healthier journeys to work and reduce environmental pollution.” However, doubt has been expressed in the popular press regarding whether the new generation of middle aged men in lycra (MAMILs) are actually using their scheme funded bikes to commute or just to gum up the roads (particularly hills) at weekends. The benefits are debatable but attractive, and the scheme has encouraged a lot of people to spend a lot of money on high end bicycles. I purchased a bike at the top end of the cost allowed by the scheme and opted for a carbon frame because it was significantly lighter than my existing bicycle’s steel frame. The wheels were lighter and tyres narrower too. All were factors that made me believe that the extra £950 I had spent would get me to work in a trice.

My new bike seemed wonderful, if somewhat uncomfortable. I didn’t notice a dramatic decrease in commuting time, nor did the cycle computer I had fitted to my new bicycle to record any notably swift journeys. But, one sunny morning, I got to work in 43 minutes, the fastest I could recall. My steel bike was consigned to a corner of the garage to gather dust—until I had a puncture. The next day I was back on my old steel bike. I fitted the cycle computer, set off . . . and discovered I had got to work in 44 minutes. “Hang on,” I thought, “was that minute worth £950 or was it a fluke?” There was only one answer: a randomised trial. I toyed with the idea of blinding it but, in the interest of self preservation and other road users, decided against it.

METHODS
This was a single centre, randomised, non-blinded trial; n=1. Both bicycles were of traditional “road” construction with drop handlebars, although the frame of one was made of steel and the second carbon (table 1; fig 1). Identical lights and fittings were used on each bike.

Between mid-January 2010 and mid-July 2010, either the steel frame bicycle (29.75 lb (13.5 kg)) or the carbon bicycle (20.9 lb (9.5 kg)) was randomly
allocated for my daily commute according to the toss of a £1 coin. The time the bicycle was moving for the 27 mile (43.5 km) round trip was recorded with a Sigma BC906 bicycle computer. The clothing worn was determined by the weather conditions on the morning of travel. Water was not carried.

The journey, predominantly on urban A roads, included 0.62 miles (1 km) of dual carriageway, 1.86 miles (3 km) of country lanes, and 328 feet (100 metres) of farm track. The total ascent for the round trip was 2766 feet (843 metres; fig 2).

The journey times were entered into Calc, a spreadsheet in the Open Office Suite. Times were compared using a two tailed Student’s t test.

RESULTS
A total of 30 journeys and 809 miles (1302 km) were travelled on the steel frame bicycle during the six month study period, compared with 26 journeys and 711 miles (1144 km) on the carbon frame bicycle (table 2). Two journeys on the steel bike were excluded owing to punctures. One journey on the carbon bike was excluded after an offer of a lift home with a colleague.

The top speed achieved was 36 mph (58 kph) on both bicycles. The slowest journey was on the carbon bike in heavy snow (2:03:20 hours:minutes:seconds). The fastest journey was on the steel bike (1:37:40) and was as a direct result of chasing one of my fitter cycling colleagues to work (fig 3). The average journey time on the steel frame bicycle was 1:47:48, and the average journey time on the carbon frame bicycle was 1:48:21. The difference in the mean journey time was 00:00:32 (95% CI –00:03:34 to 00:02:30; P=0.72).

Forces acting against the cyclist
Gravity
The difference in weight between the two bicycles is 8.85 lb (4 kg), whereas the rider weighs the same at 167.6 lb (76 kg). The energy expended on lifting the steel bike and rider through 2766 feet (843 metres) is about 740 kilojoules, compared with about 706 kilojoules for the carbon bike (see web appendix A). The additional energy expended on lifting the steel bike compared with the carbon bike was 34 kilojoules (5% extra).

Friction (rolling resistance)
The difference in friction [rolling resistance] between bicycles was 0.2 Newtons. The extra power necessary on the steel bike to overcome this difference was 1.2 watts.

Drag
The power required to overcome drag on a touring bike—steel, carbon or chocolate framed—at 15 mph (24 kph) is about 170 watts.

Winter versus summer
The difference between the mean journey time in winter (20 January to 19 April 2010) and summer (21 April to 22 July 2010) was 00:06:50 (95% CI 00:04:39 to 00:08:59; P<0.01).

DISCUSSION
The results show that there was no measurable difference in commuting time over 27 miles (43 km) on the carbon frame bicycle compared with the steel frame bicycle. This is at variance to the intuitive assumption that less weight means more speed. Why might this be the case?

Though a 30% reduction in bicycle weight may seem large, the reduction in total weight (bicycle + rider) of 4% is much less impressive. The effect this weight reduction has on the forces acting against the cyclist (gravity, friction [rolling resistance], drag [wind resistance], and the force to accelerate bicycle and rider), as

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Bicycle specification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Steel frame bicycle</strong></td>
<td><strong>Carbon frame bicycle</strong></td>
</tr>
<tr>
<td>Frame</td>
<td>Steel 321 Alloy</td>
</tr>
<tr>
<td>Wheels</td>
<td>36 spoke 700C wheel of standard alloy rim construction</td>
</tr>
<tr>
<td>Tyres</td>
<td>32 mm Schwalbe Marathon</td>
</tr>
<tr>
<td>Pedals</td>
<td>Non-clip</td>
</tr>
<tr>
<td>Weight</td>
<td>29.75 lb (13.5 kg)</td>
</tr>
</tbody>
</table>
Table 2 | Speed, distance, and journey times

<table>
<thead>
<tr>
<th></th>
<th>Steel frame bicycle</th>
<th>Carbon frame bicycle</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of journeys</td>
<td>30</td>
<td>26</td>
<td>4</td>
</tr>
<tr>
<td>Total distance</td>
<td>809 miles (1302 km)</td>
<td>711 miles (1144 km)</td>
<td>98 miles (158 km)</td>
</tr>
<tr>
<td>Top speed</td>
<td>36 mph (58 kph)</td>
<td>36 mph (58 kph)</td>
<td>0</td>
</tr>
<tr>
<td>Fastest journey time (hr:min:sec)</td>
<td>1:37:40</td>
<td>1:40:50</td>
<td>00:03:10</td>
</tr>
<tr>
<td>Slowest journey time (hr:min:sec)</td>
<td>1:57:44</td>
<td>2:03:20</td>
<td>00:05:26</td>
</tr>
<tr>
<td>Average journey time (hr:min:sec)</td>
<td>1:47:48</td>
<td>1:48:21</td>
<td>-0.00:03:34 to 00:02:30*</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>00:04:54</td>
<td>00:06:15</td>
<td>—</td>
</tr>
</tbody>
</table>

*P=0.72.

well as the effect of the road conditions, need to be considered.

Forces acting against the cyclist

Gravity

Perhaps the most obvious benefit of a lighter bike is that it is easier to cycle up hill. Though the additional energy expended on lifting the steel bike was an extra 5%, the overall effect is less as energy is conserved. Gravitational potential energy gained going up will be converted into kinetic energy going down.

Friction (rolling resistance)

Friction (rolling resistance) is relatively small for a bicycle on tarmac and is dependent on the tyre contact area and side wall flex. The manufacturers’ literature implied that both sets of tyres had similar resistance. The extra power necessary on the steel bike to overcome the difference between bikes was 1.2 watts. My brightest bicycle light has a 1 watt light emitting diode.

Drag

Drag is a factor of considerable importance. It is independent of mass and proportional to the cube of the velocity. The power required to overcome drag on the steel touring bike is seven times that required to overcome rolling resistance. The exponential increase in drag with increase in velocity has the perverse effect of counteracting anything else that may increase the speed of the bike.

Acceleration

Acceleration is a little more complex. There are two factors to consider, the force necessary to accelerate the wheel as it rotates and the force required to accelerate the cyclist and the rest of the bike. There is a very good explanation of acceleration on Wikipedia, particularly with respect to wheels, where lighter rims can confer a significant advantage, but only if there are a significant number of points of speed change on the journey. There were not enough on mine.

Winter versus summer

There was a statistically significant difference between times in the first (winter) and second (summer) halves of the trial. My summer clothing was relatively tight fitting. Looser shell winter clothing may increase drag by as much as 30%. Another factor that might increase journey time in winter is fear of falling off. When the road is wet or there is the possibility of ice, then the cyclist is more cautious. Winter is also associated with higher winds, and, as all cyclists know, the wind is always against you!

Traffic

Regardless of whether the bike is carbon or steel, you still have to stop at junctions and red lights.

Implications

Given these findings, why then do so many of us buy “performance” bicycles? Marketing must shoulder some of the responsibility. Many of us respond to “new” pharmaceuticals in a similar way to how cyclists respond to “new” bicycles. The industry invests significantly in marketing products of marginal benefit and we, as medical consumers, frequently buy into the panacea rather than objectively considering the evidence. We must excuse consumerism, particularly at this time of year, because without it our capitalist society would collapse.

The purchase of the carbon bike made me feel good, and even though the ride is “harsher” (less comfortable), I still commute on it, especially in good weather. I haven’t compared the brakes but they seem better. Which do I enjoy riding most? Well, after the trial I have to go for the steel bike. I get there as quickly, and it is more comfortable, better value, and has more “character.” If the carbon bike were stolen would I replace it? I’d have to say no. I’d spend the money on high visibility low drag clothing and better lights.

Lance Armstrong, seven times winner of the Tour de France, said, “It’s not about the bike.” One wonders whether Jane Austen had us cycling enthusiasts, rather than young ladies, in mind when she suggested that conversation should be confined to the “weather and the state of the roads.”

Conclusions

A 30% reduction in bicycle weight did not reduce commuting time over a distance of 27 miles (43.5 km). A new lightweight bicycle may have many attractions, but if the bicycle is used to commute, a reduction in
the weight of the cyclist rather than that of the bicycle may deliver greater benefit and at reduced cost.

I thank C Cooper and H Spencer for their helpful comments and R Groves for proof reading the manuscript and checking the maths.

Funding: The study was entirely funded by the author and the author has no commercial relationship with any bicycle manufacturer or commercial cycling enterprise.

Competing interests: The author has completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declares: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; and no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Ethical approval was not obtained as the sole investigator and subject was the author. The research was conducted on his regular journey to and from work using his normal mode of transport.

Critical Appraisal Skills Programme (CASP)
making sense of evidence

10 questions to help you make sense of randomised controlled trials

How to use this appraisal tool
Three broad issues need to be considered when appraising the report of a randomised controlled trial:

- Is the trial valid?
- What are the results?
- Will the results help locally?

The 10 questions on the following pages are designed to help you think about these issues systematically.

The first two questions are screening questions and can be answered quickly. If the answer to both is “yes”, it is worth proceeding with the remaining questions.

You are asked to record a “yes”, “no” or “can’t tell” to most of the questions. A number of italicised prompts are given after each question.

These are designed to remind you why the question is important. Record your reasons for your answers in the spaces provided.

The 10 questions are adapted from Guyatt GH, Sackett DL, and Cook DJ, Users’ guides to the medical literature. II. How to use an article about therapy or prevention. JAMA 1993; 270 (21): 2598-2601 and JAMA 1994; 271(1): 59-63

© Public Health Resource Unit, England (2006). All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the Public Health Resource Unit. If permission is given, then copies must include this statement together with the words “© Public Health Resource Unit, England 2006”. However, NHS organisations may reproduce or use the publication for non-commercial educational purposes provided the source is acknowledged.
Screening Questions

1. Did the study ask a clearly-focused question?  
   ☐ Yes  ☐ Can’t tell  ☐ No
   Consider if the question is ‘focused’ in terms of:
   – the population studied
   – the intervention given
   – the outcomes considered

2. Was this a randomised controlled trial (RCT) and was it appropriately so?  
   ☐ Yes  ☐ Can’t tell  ☐ No
   Consider:
   – why this study was carried out as an RCT
   – if this was the right research approach for the question being asked

Is it worth continuing?

Detailed Questions

3. Were participants appropriately allocated to intervention and control groups?  
   ☐ Yes  ☐ Can’t tell  ☐ No
   Consider:
   – how participants were allocated to intervention and control groups. Was the process truly random?
   – whether the method of allocation was described. Was a method used to balance the randomization, e.g. stratification?
   – how the randomization schedule was generated and how a participant was allocated to a study group
   – if the groups were well balanced. Are any differences between the groups at entry to the trial reported?
   – if there were differences reported that might have explained any outcome(s) (confounding)
4. Were participants, staff and study personnel ‘blind’ to participants’ study group?  

Consider:
– the fact that blinding is not always possible
– if every effort was made to achieve blinding
– if you think it matters in this study
– the fact that we are looking for ‘observer bias’

5. Were all of the participants who entered the trial accounted for at its conclusion?  

Consider:
– if any intervention-group participants got a control-group option or vice versa
– if all participants were followed up in each study group (was there loss-to-follow-up?)
– if all the participants’ outcomes were analysed by the groups to which they were originally allocated (intention-to-treat analysis)
– what additional information would you liked to have seen to make you feel better about this

6. Were the participants in all groups followed up and data collected in the same way?  

Consider:
– if, for example, they were reviewed at the same time intervals and if they received the same amount of attention from researchers and health workers. Any differences may introduce performance bias.

7. Did the study have enough participants to minimise the play of chance?  

Consider:
– if there is a power calculation. This will estimate how many participants are needed to be reasonably sure of finding something important (if it really exists and for a given level of uncertainty about the final result).
8. How are the results presented and what is the main result?

   Consider:
   – if, for example, the results are presented as a proportion of people experiencing an outcome, such as risks, or as a measurement, such as mean or median differences, or as survival curves and hazards
   – how large this size of result is and how meaningful it is
   – how you would sum up the bottom-line result of the trial in one sentence

9. How precise are these results?

   Consider:
   – if the result is precise enough to make a decision
   – if a confidence interval were reported. Would your decision about whether or not to use this intervention be the same at the upper confidence limit as at the lower confidence limit?
   – if a p-value is reported where confidence intervals are unavailable

10. Were all important outcomes considered so the results can be applied? □ Yes □ Can’t tell □ No

   Consider whether:
   – the people included in the trail could be different from your population in ways that would produce different results
   – your local setting differs much from that of the trial
   – you can provide the same treatment in your setting

   Consider outcomes from the point of view of the:
   – individual
   – policy maker and professionals
   – family/carers
   – wider community

   Consider whether:
   – any benefit reported outweighs any harm and/or cost. If this information is not reported can it be filled in from elsewhere?
   – policy or practice should change as a result of the evidence contained in this trial

© Public Health Resource Unit, England (2006). All rights reserved.
Effect of tranexamic acid on surgical bleeding: systematic review and cumulative meta-analysis

Katharine Ker research fellow, Phil Edwards senior lecturer, Pablo Perel clinical senior lecturer, Haleema Shakur senior lecturer, Ian Roberts professor of epidemiology

Clinical Trials Unit, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

Abstract
Objective To assess the effect of tranexamic acid on blood transfusion, thromboembolic events, and mortality in surgical patients.

Design Systematic review and meta-analysis.

Data sources Cochrane central register of controlled trials, Medline, and Embase, from inception to September 2011, the World Health Organization International Clinical Trials Registry Platform, and the reference lists of relevant articles.

Study selection Randomised controlled trials comparing tranexamic acid with no tranexamic acid or placebo in surgical patients. Outcome measures of interest were the number of patients receiving a blood transfusion; the number of patients with a thromboembolic event (myocardial infarction, stroke, deep vein thrombosis, and pulmonary embolism); and the number of deaths. Trials were included irrespective of language or publication status.

Results 129 trials, totalling 10 488 patients, carried out between 1972 and 2011 were included. Tranexamic acid reduced the probability of receiving a blood transfusion by a third (risk ratio 0.62, 95% confidence interval 0.58 to 0.65; P<0.001). This effect remained when the analysis was restricted to trials using adequate allocation concealment (0.68, 0.62 to 0.74; P<0.001). The effect of tranexamic acid on myocardial infarction (0.68, 0.43 to 1.09; P=0.11), stroke (1.14, 0.65 to 2.00; P=0.65), deep vein thrombosis (0.86, 0.53 to 1.39; P=0.54), and pulmonary embolism (0.61, 0.25 to 1.47; P=0.27) was uncertain. Fewer deaths occurred in the tranexamic acid group (0.61, 0.38 to 0.98; P=0.04), although when the analysis was restricted to trials using adequate concealment there was considerable uncertainty (0.67, 0.33 to 1.34; P=0.25). Cumulative meta-analysis showed that reliable evidence that tranexamic acid reduces the need for transfusion has been available for over 10 years.

Conclusions Strong evidence that tranexamic acid reduces blood transfusion in surgery has been available for many years. Further trials on the effect of tranexamic acid on blood transfusion are unlikely to add useful new information. However, the effect of tranexamic acid on thromboembolic events and mortality remains uncertain. Surgical patients should be made aware of this evidence so that they can make an informed choice.

Introduction
In October 2011 the BMJ published a randomised controlled trial on the effect of tranexamic acid on blood transfusion in patients undergoing radical retropubic prostatectomy. The authors pointed out that this was the first trial to assess the effect of tranexamic acid on blood transfusion in this particular operation. While this may be the case, it was not the first trial to examine the effect of tranexamic acid on blood transfusion in surgery more generally. A systematic review published in 2001 presented data from 18 clinical trials and showed that tranexamic acid reduces the probability of blood transfusion in elective surgery by 34%. We assessed the current evidence for the effect of tranexamic acid on blood transfusion, thromboembolic events, and mortality in surgical patients. To examine how the evidence has changed over time we used cumulative meta-analyses.

Methods
Although we specified and documented the methods of the analysis and inclusion criteria for this systematic review in advance, the protocol was not registered. We searched for all randomised controlled trials that compared tranexamic acid with no tranexamic acid or placebo in elective and emergency surgery. No age restriction was applied. Potentially eligible trials were identified by searching the Cochrane central register of controlled trials (2011, issue 3), Medline (1950 to September 2011), and Embase (1980 to September 2011), using a combination of subject headings and text words to identify...
randomised controlled trials of any antifibrinolytic drug (see supplementary file for Medline search strategy). Searches were not restricted by language or publication status. To identify ongoing or unpublished trials we searched the WHO International Clinical Trials Registry Platform. We also examined the reference lists of eligible trials and reviews. Two authors independently screened the search output to identify records of potentially eligible trials, the full texts of which were retrieved and assessed for inclusion.

**Outcome data**

Outcome measures of interest were the number of patients receiving a blood transfusion; the number of patients with a thromboembolic event (myocardial infarction, stroke, deep vein thrombosis, and pulmonary embolism); and the number of deaths. We contacted trial authors to obtain any missing outcome data.

**Data extraction and risk of bias assessment**

We extracted data on the age and sex of trial participants, type of surgery, dose and timing of tranexamic acid, type of comparator, and outcome data. We also collected information on whether a systematic review had been conducted to support the trial rationale and whether a systematic review was cited in the trial report. We assessed the risk of bias associated with the method of sequence generation, allocation concealment, blinding, and the completeness of outcome data. As the risk of bias for blinding may vary according to outcome, we assessed this separately for each outcome. We rated the risk of bias as being low, unclear, or high according to established criteria.

**Statistical analysis**

For each outcome we calculated risk ratios and 95% confidence intervals. We pooled these using a fixed effect model. Subgroup analyses were carried out to examine whether the effect of tranexamic acid on blood transfusion varied by type of surgery. Sensitivity analyses were done to quantify the effect of tranexamic acid on all outcomes when restricted to trials with adequate allocation concealment and blinded outcome assessment. We carried out a cumulative meta-analysis of the effect of tranexamic acid on blood transfusion based on the date of publication, and, when restricted to trials with adequate concealment, cumulative meta-analyses of the effect of tranexamic acid on blood transfusion, myocardial infarction, and mortality. Heterogeneity was examined by visual inspection of forest plots, the I² statistic, and the χ² test. We inspected funnel plots for the presence of small study effects. Statistical analyses were carried out using Stata version 11 and RevMan version 5.

**Results**

Overall, 127 articles describing 129 randomised controlled trials and totalling 10 488 patients were included; 5484 of these patients were allocated to tranexamic acid and 5004 to a control group (fig 1). The median sample size was 60 (range 10-660) patients. In total, 126 (98%) trials were in elective surgery and three (2%) in emergency surgery. Eleven (8%) trials involved children. The authors of 86 trials were contacted for missing data, 39 of whom provided additional information. Data were available on blood transfusion from 95 (74%) trials, on myocardial infarction from 73 (56%), on stroke from 71 (55%), on deep vein thrombosis from 72 (56%), on pulmonary embolism from 66 (51%), and on mortality from 72 (56%). Seven (5%) trials did not report any data on the outcome measures of interest to this review or reported data in a format that was unsuitable for inclusion in the analyses.

A further 14 ongoing trials were identified, with a median planned sample size of 130 patients. The 14 trials were in orthopaedic (n=5), cardiac (n=4), cranial (n=2), hepatic (n=1), ear, nose, and throat (n=1), and gynaecological (n=1) surgery. In 12 of the 14 trials blood transfusion was a main outcome measure.

**Risk of bias**

Overall, 44 (34%) trials were judged to be at low risk of bias for sequence generation and five (4%) to be at high risk (see the supplementary file for the risk of bias judgments for each methodological quality item for the included trials). The risk of bias in the remaining 80 (62%) trials was unclear owing to lack of information. Allocation was adequately concealed in 36 trials (28%) and inadequately concealed in six (5%), with the other 87 (67%) presenting insufficient information to allow judgment. Of the 95 trials with data on blood transfusion, 69 (73%) were judged to be at low risk of bias, 22 (23%) were unclear. The risk of bias for blinding was similar for thromboembolic outcomes (myocardial infarction, stroke, deep vein thrombosis, and pulmonary embolism), with about 70% judged to be at low risk, 5% at high risk, and 25% at unclear risk. All 72 trials with mortality outcomes were judged to be at low risk of bias for blinding. Of 115 trials reporting eligible outcome data, 72 (63%) were at low risk of bias for incomplete outcome data, 17 (15%) at high risk, and 26 (23%) did not describe adequate information to permit judgment.

**Quantitative data synthesis**

Table 1 presents the results of the meta-analysis.

**Risk of blood transfusion**

Data on blood transfusion were available for 95 trials, including a total of 7838 patients. Tranexamic acid reduced the probability of receiving a blood transfusion by 38% (pooled risk ratio 0.62, 95% confidence interval 0.58 to 0.65; P<0.001). When the analysis was restricted to the 32 adequately concealed trials involving 3408 patients, tranexamic acid reduced the risk of receiving a blood transfusion by 32% (0.68, 0.62 to 0.74; P<0.001). When the analysis was restricted to the 69 trials involving 5968 patients with adequate blinding for this outcome, tranexamic acid reduced the risk of blood transfusion by 37% (0.63, 0.59 to 0.68; P<0.001).

The trials with blood transfusion data involved cardiac (n=42), orthopaedic (n=36), cranial and orthognathic (n=7), gynaecological (n=5), hepatic (n=2), urological (n=2), and vascular (n=1) surgery. Blood transfusion was statistically significantly reduced in cardiac, orthopaedic, cranial and orthognathic, hepatic, and urological surgery (table 2). The pooled estimates for blood transfusion were consistent with a reduction in the tranexamic acid group among trials in vascular and gynaecological surgery, although the results were imprecise. There was moderate heterogeneity in magnitude of the effects of tranexamic acid by type of surgery, although the direction of the effects was largely consistent.

**Thromboembolic events**

There was uncertainty about the effect of tranexamic acid on myocardial infarction (risk ratio 0.68, 95% confidence interval...
0.43 to 1.09; P=0.11), stroke (1.14, 0.65 to 2.00; P=0.65), deep vein thrombosis (0.86, 0.53 to 1.39; P=0.54), and pulmonary embolism (0.61, 0.25 to 1.47; P=0.27). The results were similar when the analyses were restricted to trials with adequate allocation concealment and those with blinded outcome assessment.

Mortality
Fewer deaths occurred in the tranexamic acid group (risk ratio 0.61, 95% confidence interval 0.38 to 0.98; P=0.044), although there was uncertainty about this effect, particularly when the analysis was restricted to the 28 trials with adequate concealment (0.67, 0.33 to 1.34; P=0.25).

Cumulative meta-analyses
The supplementary file shows the results of the cumulative meta-analysis of the 95 trials with data on blood transfusion. A statistically significant effect of tranexamic acid on blood transfusion was first observed after publication of the third trial in 1993 (0.59, 0.43 to 0.80; P=0.001). Although subsequent trials have increased the precision of the point estimate, no substantive change has occurred in the direction or magnitude of the treatment effect.

Figures 2-4 show the cumulative meta-analyses of the effect of tranexamic acid on blood transfusion, myocardial infarction, and mortality among the trials with adequate allocation concealment. A statistically significant effect of tranexamic acid on blood transfusion was consistently observed after publication of the 10th trial in 2001.

Small study effects
Inspection of the funnel plot (fig 5) for the outcome blood transfusion suggested the presence of small study effects favouring tranexamic acid. The other outcomes showed no clear asymmetry in the funnel plots.

Citation of previous systematic reviews
Between 1994 and 2011, 30 systematic reviews have been published on the effects of tranexamic acid in surgery. Assuming a 12 month publication time lag, 98 of the 116 (84%) included trial reports published as full journal articles were published when at least one systematic review was available. Examination of the reference lists of these reports indicated that 45 (46%) did not cite any of the available systematic reviews. The authors of two of the 116 trial reports had carried out a systematic review and presented the findings within the final trial publication.

Discussion
Reliable evidence that tranexamic acid reduces blood transfusion in surgical patients has been available for many years. The treatment effect varies somewhat according to the type of surgery, but the effect is consistently large and remains so when the analysis is restricted to trials with adequate allocation concealment. The effect of tranexamic acid on thromboembolic events and mortality has not been adequately assessed by clinical trials in surgery and remains uncertain. In view of the evidence, those planning further placebo controlled trials should explain why they think that tranexamic acid might not reduce the risk of blood transfusion in the particular group of surgical patients under consideration and focus their efforts on resolving the uncertainties about the effect of tranexamic acid on thromboembolic events and mortality.

Strengths and weaknesses of the review
The inferences that can be made from the included trials depend on their quality, and many had methodological limitations. However, the large and statistically significant effect on blood transfusion remained when the analysis was restricted to trials with adequate allocation concealment and with adequate blinding.

We systematically searched a range of databases for published and unpublished trials. However, we cannot exclude the possibility that some were missed. Indeed, the observed asymmetry in the funnel plot could be explained by publication bias. If many unpublished trials show little or no effect of tranexamic acid on blood transfusion, then this meta-analysis may have overestimated the treatment effect. Although some degree of overestimation is likely, it seems improbable that publication bias could account for all of the observed effect.

Although mortality and thromboembolic outcomes showed no obvious asymmetry in the funnel plots, publication and other reporting biases remain a potential threat to the validity of the effect estimates. Mortality data were reported in only a third of the included trials, and less than half reported data on myocardial infarction, stroke, deep vein thrombosis, and pulmonary embolism. Inadequate reporting of adverse events is not unusual in reports of clinical trials and hinders the reliable estimation of treatment effects.\(^\text{106–117}\) After contacting the trial authors we obtained some missing data and were able to include mortality data for three quarters of the included trials and data on myocardial infarction, stroke, deep vein thrombosis, and pulmonary embolism for about half of the trials. However, the effect of outcome reporting bias in this review remains open to question. Even if there was no significant bias, the precision of the estimates is low and the data are compatible with either a moderate increase or a moderate decrease in the risk of thromboembolic events.

Implications of the findings
The evidence in this review suggests that the uncertainty about the effect of tranexamic acid on blood transfusion in surgical patients was resolved over a decade ago; however, uncertainties about its effect on thromboembolic events and mortality persist. Despite this, trials of tranexamic acid continue to assess the effect on blood transfusion. One reason may be a reluctance to generalise the evidence across surgery types, although there is no evidence that the relative effect of tranexamic acid on blood transfusion varies by type of surgery. A second reason may be that trialists are unaware of the existing evidence when initiating a new trial. Our observation that only half of the trials cited one or more of the available systematic reviews and just two carried out their own systematic review, does suggest that many trialists are indeed failing to adequately consider the existing evidence.

Blood is a scarce and costly resource and blood transfusion is not without risk. The cost of a unit of red cells to the National Health Service has increased from £78 (€96; $126) in 2000 to £125 in 2011, and blood transfusion has several rare but serious adverse effects. Worldwide, most people do not have access to safe blood. Globally the most important transfusion related risks are HIV, hepatitis B virus, and hepatitis C virus, due to their high prevalence. That tranexamic acid safely reduces the need for blood transfusion in surgery has important health and economic implications in high, middle, and low income countries. The evidence that tranexamic acid reduces the need for blood transfusion is strong but the safety of routine use of tranexamic acid in surgical patients remains uncertain. A modest increase in the risk of thromboembolic effects could outweigh...
the benefits of reduced blood use. Although some increased risk might be expected on theoretical grounds, recent evidence from the CRASH-2 (Clinical Randomisation of an Antifibrinolytic in Significant Haemorrhage) trial of tranexamic acid in bleeding trauma patients showed a statistically significant reduction in mortality with no increase in thromboembolic effects. Indeed, there was a statistically significant reduction in the risk of myocardial infarction in trauma patients who received tranexamic acid.179

Further small trials of tranexamic acid in surgical patients considered in isolation will not resolve the uncertainties about the effects on thromboembolic events and mortality. Because thromboembolic events are relatively rare, such trials lack sufficient power to detect clinically important increases in risk, and a meta-analysis of small trials remains vulnerable to publication bias. The ongoing Aspirin and Tranexamic Acid for Coronary Artery Surgery trial180 with a planned sample size of 4300 high risk patients undergoing coronary artery surgery, should contribute importantly to resolving the uncertainty about the effect of tranexamic acid on mortality and thromboembolic events in this specific group. We urge investors involved in all ongoing trials of tranexamic acid in surgery to collect data on thromboembolic events and mortality for inclusion in a prospective meta-analysis until the uncertainties are resolved.

However, a need remains for a larger pragmatic clinical trial of the effect of routine use of tranexamic acid in a heterogeneous group of surgical patients. The possibility that tranexamic acid might reduce mortality without any increase in the risk of thromboembolic events would justify the effort and expenditure involved.

We thank Karen Blackhall for designing and performing the bibliographic database searches; Mercedes Hernandez for help in contacting authors and data extraction; Joey Kwong, Gabi Meineke, Daniella Manno, Ayumi Naito, and Emma Sydnerham for assisting with the translation of non-English language papers; the following trial authors who responded to our requests for further information: J Alvarez, G Angelini, G Benoni, C Chen, W Choi, C Dadure, P Duran de la Fuente, A Ejaazi, E Elwadity, J Engell, L Eslaminia, J Field, B Garg, J Gill, S Goobie, G Greiff, J Guay, J Hardy, H Hustud, J Jimenez Rivera, T Johansson, P Kakkar, A Later, J McConnell, F Moret, D Neillipovit, R Niskanen, P Patel, H Pleym, R Rannikkio, M Seer, L Trappepe, T Tsusumimoto, T Vanek, G Wang, N Webster, J Wong, C Wu, and J Yepes.

Contributors: KK and IR conceived the study. KK is guarantor. IR is a National Institute for Health Research senior investigator. KK, IR, PE, PP, and HS screened the search output. KK extracted data and carried out the analyses. KK and IR wrote the manuscript with contributions from PP, PE, and HS. The final version was approved by all authors.

Funding: This study received no external funding.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; and no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Data sharing: No additional data available.

Evidence that TXA reduces blood transfusion in surgical patients has been available for over a decade, yet the effect on thromboembolic events and mortality remains uncertain.

Further trials on the effect of TXA on blood transfusion are unlikely to add useful new information.

A large pragmatic clinical trial of TXA in a heterogeneous group of surgical patients is needed to resolve the uncertainties about the effects on thromboembolic events and mortality.


132. Aspirin and tranexamic acid for coronary artery surgery: ATACAS


135. Tranexamic acid for coronary artery surgery: NCT01094977

136. Fibin glue or tranexamic acid for total knee arthroplasty-ATRESHMS (NCT01363070)

137. Hemostatic and anti-inflammatory effects of ustipintin and tranexamic acid in cardiopulmonary bypass surgery (NCT01060189) http://clinicaltrials.gov/show/NCT01060189

138. Impact of tranexamic acid on red blood cell transfusion in surgical (NCT01258015) http://clinicaltrials.gov/show/NCT01258015

139. Tranexamic acid (TXA) versus epsilon amino caproic acid (EACA) versus placebo for spine surgery (NCT00958581) http://clinicaltrials.gov/show/NCT00958581

140. Intravenous tranexamic acid and intraoperative visualization during functional endoscopic sinus surgery (NCT01116698) http://clinicaltrials.gov/show/NCT01116698

141. Multicenter, randomized placebo-controlled trial to evaluate the effect of perioperative use of tranexamic acid on treatment results and surgical bleeding in major spine surgery (NCT01165695) http://clinicaltrials.gov/show/NCT01165695

142. Non-diabetic scoliosis treated with tranexamic acid (NCT01089140) http://clinicaltrials.gov/show/NCT01089140

143. Search for optimum dose and timing of tranexaminic acid administration in cardiac surgery with cardiopulmonary bypass (JPRN UMIN000000327) http://apps.jprn.pts北海/sheartrial/trial_3.asp?id=JPRN-U000000327

144. Tranexamic acid for craniofacial surgery (NCT01722436) http://clinicaltrials.gov/show/NCT01722436

145. Tranexamic acid in surgery of advanced ovarian cancer (NCT07401611) http://clinicaltrials.gov/show/NCT07401611

146. Tranexamic acid plus Neosynephrine and tranexamic acid in cardiac surgery (NCT01060176) http://clinicaltrials.gov/show/NCT01060176

147. Tranexamic acid versus placebo to reduce perioperative bleeding after major hepatic resection (NCT00675394) http://clinicaltrials.gov/show/NCT00675394


Liu JM, Peng HM, Shen JX, Qiu GX. [A meta-analysis of the effectiveness and safety of using tranexamic acid in spine surgery]. Zhonghua Wu Ke Za Zhi 2010;48:937-42.


Accepted: 26 March 2012

Cite this as: BMJ 2012;344:e3054

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non-commercial License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non-commercial and is otherwise in compliance with the license. See: http://creativecommons.org/licenses/by-nc/2.0/ and http://creativecommons.org/licenses/by-nc/2.0/legalcode.

No commercial reuse: See rights and reprints http://www.bmj.com/permissions
Subscribe: http://www.bmj.com/subscribe
### Tables

**Table 1 | Meta-analysis of effect of tranexamic acid on blood transfusion, thromboembolic events, and mortality**

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Events (tranexamic acid/control)</th>
<th>Pooled risk ratio (95% CI)</th>
<th>P value*</th>
<th>I^2 (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood transfusion:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All trials</td>
<td>1067/1520</td>
<td>0.62 (0.58 to 0.65)</td>
<td>&lt;0.001</td>
<td>69</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Well concealed trials</td>
<td>459/609</td>
<td>0.68 (0.62 to 0.74)</td>
<td>&lt;0.001</td>
<td>55</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Adequate blinding</td>
<td>847/1182</td>
<td>0.63 (0.59 to 0.68)</td>
<td>&lt;0.001</td>
<td>54</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Myocardial infarction:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All trials</td>
<td>23/35</td>
<td>0.68 (0.42 to 1.09)</td>
<td>0.11</td>
<td>0</td>
<td>0.90</td>
</tr>
<tr>
<td>Well concealed trials</td>
<td>16/25</td>
<td>0.70 (0.39 to 1.25)</td>
<td>0.22</td>
<td>0</td>
<td>0.82</td>
</tr>
<tr>
<td>Adequate blinding</td>
<td>18/33</td>
<td>0.59 (0.36 to 0.98)</td>
<td>0.04</td>
<td>0</td>
<td>0.81</td>
</tr>
<tr>
<td>Stroke:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All trials</td>
<td>23/16</td>
<td>1.14 (0.65 to 2.00)</td>
<td>0.65</td>
<td>0</td>
<td>0.92</td>
</tr>
<tr>
<td>Well concealed trials</td>
<td>5/4</td>
<td>1.18 (0.36 to 3.83)</td>
<td>0.78</td>
<td>0</td>
<td>0.92</td>
</tr>
<tr>
<td>Adequate blinding</td>
<td>23/16</td>
<td>1.14 (0.65 to 2.00)</td>
<td>0.65</td>
<td>0</td>
<td>0.92</td>
</tr>
<tr>
<td>Deep vein thrombosis:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All trials</td>
<td>25/29</td>
<td>0.86 (0.53 to 1.39)</td>
<td>0.54</td>
<td>0</td>
<td>0.96</td>
</tr>
<tr>
<td>Well concealed trials</td>
<td>13/14</td>
<td>0.92 (0.45 to 1.85)</td>
<td>0.81</td>
<td>0</td>
<td>0.81</td>
</tr>
<tr>
<td>Adequate blinding</td>
<td>18/22</td>
<td>0.82 (0.46 to 1.44)</td>
<td>0.49</td>
<td>0</td>
<td>0.98</td>
</tr>
<tr>
<td>Pulmonary embolism:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All trials</td>
<td>4/8</td>
<td>0.61 (0.25 to 1.47)</td>
<td>0.27</td>
<td>0</td>
<td>0.96</td>
</tr>
<tr>
<td>Well concealed trials</td>
<td>1/3</td>
<td>0.52 (0.10 to 2.75)</td>
<td>0.44</td>
<td>0</td>
<td>0.80</td>
</tr>
<tr>
<td>Adequate blinding</td>
<td>4/6</td>
<td>0.70 (0.26 to 1.87)</td>
<td>0.48</td>
<td>0</td>
<td>0.91</td>
</tr>
<tr>
<td>Mortality:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All trials</td>
<td>20/34</td>
<td>0.61 (0.38 to 0.98)</td>
<td>0.04</td>
<td>0</td>
<td>0.97</td>
</tr>
<tr>
<td>Well concealed trials</td>
<td>9/15</td>
<td>0.67 (0.33 to 1.34)</td>
<td>0.25</td>
<td>0</td>
<td>0.85</td>
</tr>
<tr>
<td>Adequate blinding</td>
<td>20/34</td>
<td>0.61 (0.38 to 0.98)</td>
<td>0.04</td>
<td>0</td>
<td>0.97</td>
</tr>
</tbody>
</table>

*Test for effect.
Table 2 | Meta-analysis of effect of tranexamic acid on risk of blood transfusion, stratified by type of surgery

<table>
<thead>
<tr>
<th>Type of surgery</th>
<th>No of events (tranexamic acid/control)</th>
<th>Pooled risk ratio (95% CI)</th>
<th>P value*</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac</td>
<td>622/835</td>
<td>0.65 (0.60 to 0.70)</td>
<td>&lt;0.001</td>
<td>60</td>
</tr>
<tr>
<td>Orthopaedic</td>
<td>296/462</td>
<td>0.55 (0.49 to 0.61)</td>
<td>&lt;0.001</td>
<td>83</td>
</tr>
<tr>
<td>Hepatic</td>
<td>29/54</td>
<td>0.52 (0.39 to 0.68)</td>
<td>&lt;0.001</td>
<td>93</td>
</tr>
<tr>
<td>Urological</td>
<td>40/60</td>
<td>0.66 (0.48 to 0.91)</td>
<td>0.01</td>
<td>2</td>
</tr>
<tr>
<td>Vascular</td>
<td>11/19</td>
<td>0.58 (0.34 to 0.99)</td>
<td>0.05</td>
<td>—</td>
</tr>
<tr>
<td>Gynaecological</td>
<td>17/50</td>
<td>0.86 (0.48 to 1.54)</td>
<td>0.61</td>
<td>65</td>
</tr>
<tr>
<td>Cranial and orthognathic</td>
<td>52/76</td>
<td>0.63 (0.45 to 0.86)</td>
<td>0.004</td>
<td>46</td>
</tr>
</tbody>
</table>

*Test for effect.
Fig 1 Selection of trials for review
Fig 2 Cumulative meta-analysis of the effect of tranexamic acid in surgery on risk of blood transfusion in adequately concealed trials
**Fig 3** Cumulative meta-analysis of the effect of tranexamic acid in surgery on risk of myocardial infarction in adequately concealed trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>Tranexamic acid</th>
<th>Control</th>
<th>Risk ratio (95% CI)</th>
<th>Risk ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horrow 1990</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Horrow 1991</td>
<td>0/37</td>
<td>0/44</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Coffey 1995</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Horrow 1995</td>
<td>0/121</td>
<td>0/24</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Karski 1996</td>
<td>0/99</td>
<td>0/48</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Benonl 1996</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Boylan 1996</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Hardy 1998</td>
<td>1/43</td>
<td>2/45</td>
<td>0.52 (0.05 to 5.56)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Benoni 2000</td>
<td>0/20</td>
<td>0/19</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Tanaka 2001</td>
<td>0/73</td>
<td>0/26</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Casali 2001</td>
<td>0/20</td>
<td>0/20</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Benoni 2001</td>
<td>1/18</td>
<td>0/20</td>
<td>1.07 (0.19 to 5.94)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Casali 2002</td>
<td>1/29</td>
<td>1/29</td>
<td>1.05 (0.25 to 4.47)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Husted 2003</td>
<td>0/20</td>
<td>0/20</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Casali 2004</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Diprose 2005</td>
<td>5/60</td>
<td>4/60</td>
<td>1.16 (0.45 to 3.00)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Johansson 2005</td>
<td>0/47</td>
<td>0/53</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Karski 2005</td>
<td>2/147</td>
<td>3/165</td>
<td>1.04 (0.45 to 2.41)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Kultunen 2005</td>
<td>1/20</td>
<td>1/20</td>
<td>1.04 (0.47 to 2.31)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Vanek 2006</td>
<td>0/32</td>
<td>0/30</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Open 2006</td>
<td>0/15</td>
<td>0/14</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Murphy 2006</td>
<td>0/50</td>
<td>1/50</td>
<td>0.96 (0.44 to 2.06)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Maddali 2007</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Jimenez 2007</td>
<td>0/24</td>
<td>0/26</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Sadeghi 2007</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Chen 2008</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Elwadly 2008</td>
<td>0/32</td>
<td>0/32</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Wang 2008</td>
<td>1/73</td>
<td>0/74</td>
<td>1.04 (0.49 to 2.17)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Later 2009</td>
<td>0/99</td>
<td>8/103</td>
<td>0.66 (0.34 to 1.27)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Taghaddam 2009</td>
<td>0/50</td>
<td>0/50</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Zufferey 2010</td>
<td>0/57</td>
<td>0/53</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Gungorduk 2010</td>
<td>0/330</td>
<td>0/330</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Dadure 2011</td>
<td>0/19</td>
<td>0/20</td>
<td>Not estimable</td>
<td>Not estimable</td>
</tr>
<tr>
<td>McConnell 2011</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>Greiff 2011</td>
<td>3/30</td>
<td>4/33</td>
<td>0.68 (0.68 to 1.24)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Crescenti 2011</td>
<td>1/100</td>
<td>1/100</td>
<td>0.70 (0.39 to 1.25)</td>
<td>Not estimable</td>
</tr>
<tr>
<td>Trial</td>
<td>No with outcome/No in group</td>
<td>Risk ratio (95% CI)</td>
<td>Risk ratio (95% CI)</td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td>Horrow 1990</td>
<td>Not reported/Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
</tr>
<tr>
<td>Horrow 1991</td>
<td>Not reported/Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
</tr>
<tr>
<td>Coffey 1995</td>
<td>0/16</td>
<td>0.29 (0.01 to 6.69)</td>
<td>Not reported</td>
<td></td>
</tr>
<tr>
<td>Horrow 1995</td>
<td>Not reported/Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
</tr>
<tr>
<td>Karski 1996</td>
<td>Not reported/Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
</tr>
<tr>
<td>Benonl 1996</td>
<td>0/43</td>
<td>0.17 (0.02 to 1.36)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Boylan 1996</td>
<td>0/25</td>
<td>0.45 (0.17 to 1.19)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Hardy 1998</td>
<td>0/43</td>
<td>0.46 (0.15 to 1.44)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Banonl 2000</td>
<td>0/20</td>
<td>0.63 (0.30 to 1.32)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Tanea 2001</td>
<td>Not reported/Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
</tr>
<tr>
<td>Casati 2001</td>
<td>Not reported/Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
</tr>
<tr>
<td>Banonl 2001</td>
<td>0/8</td>
<td>0.66 (0.29 to 1.52)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Husted 2003</td>
<td>0/20</td>
<td>0.47 (0.17 to 1.30)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Casati 2004</td>
<td>1/52</td>
<td>0.45 (0.17 to 1.19)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Diprose 2005</td>
<td>0/60</td>
<td>0.46 (0.15 to 1.44)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Johansson 2005</td>
<td>0/47</td>
<td>0.47 (0.17 to 1.30)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Karski 2005</td>
<td>3/147</td>
<td>0.63 (0.30 to 1.32)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Kultunen 2005</td>
<td>0/20</td>
<td>0.69 (0.34 to 1.41)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Vanek 2006</td>
<td>0/32</td>
<td>0.69 (0.34 to 1.41)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Open 2006</td>
<td>0/50</td>
<td>0.67 (0.33 to 1.34)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Murphy 2006</td>
<td>0/111</td>
<td>0.67 (0.33 to 1.34)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Maddall 2007</td>
<td>0/111</td>
<td>0.67 (0.33 to 1.34)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Jimenez 2007</td>
<td>0/24</td>
<td>0.67 (0.33 to 1.34)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Sadeghi 2007</td>
<td>0/32</td>
<td>0.67 (0.33 to 1.34)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Chen 2008</td>
<td>0/26</td>
<td>0.67 (0.33 to 1.34)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Elwadi 2008</td>
<td>0/32</td>
<td>0.67 (0.33 to 1.34)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Wang 2008</td>
<td>0/73</td>
<td>0.67 (0.33 to 1.34)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Later 2009</td>
<td>1/99</td>
<td>0.67 (0.33 to 1.34)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Taghaddan 2009</td>
<td>Not reported/Not reported</td>
<td>Not reported</td>
<td>Not reported</td>
<td></td>
</tr>
<tr>
<td>Zuffere 2010</td>
<td>1/57</td>
<td>0.67 (0.33 to 1.34)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Gungorduk 2010</td>
<td>0/330</td>
<td>0.67 (0.33 to 1.34)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Dadure 2011</td>
<td>0/19</td>
<td>0.67 (0.33 to 1.34)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>McConnell 2011</td>
<td>0/22</td>
<td>0.67 (0.33 to 1.34)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Grie 2011</td>
<td>0/30</td>
<td>0.67 (0.33 to 1.34)</td>
<td>Not estimable</td>
<td></td>
</tr>
<tr>
<td>Crescenti 2011</td>
<td>0/100</td>
<td>0.67 (0.33 to 1.34)</td>
<td>Not estimable</td>
<td></td>
</tr>
</tbody>
</table>

**Fig 4** Cumulative meta-analysis of the effect of tranexamic acid in surgery on risk of death in adequately concealed trials

**Fig 5** Funnel plot with pseudo 95% confidence limits for meta-analysis on effect of tranexamic acid on risk of blood transfusion
Diagnostic Accuracy of Pulmonary Embolism Rule-Out Criteria: A Systematic Review and Meta-analysis

Balwinder Singh, MBBS, Ajay K. Parsaik, MBBS, Dipti Agarwal, MBBS, Alok Surana, MBBS, Soniya S. Mascarenhas, RN, Subhash Chandra, MBBS

From the Division of Pulmonary and Critical Care Medicine, Mayo Clinic, Rochester, MN (Singh); the Department of Neurology (Parsaik) and Department of Emergency Medicine (Agarwal), Mayo Clinic, Rochester, MN; the Department of General Medicine, S.N. Medical College, Jodhpur, India (Surana); Health Care Management, London Training College, London, UK (Mascarenhas); and the Department of Internal Medicine, Greater Baltimore Medicine Center, Towson, MD (Chandra).

Study objective: To perform a systematic review and meta-analysis to define the diagnostic performance of pulmonary embolism rule-out criteria (PERC) in deferring the need for D-dimer testing to rule out pulmonary embolism in the emergency department (ED).

Methods: We searched EMBASE, MEDLINE, Scopus, Web of Knowledge, and all the evidence-based medicine reviews that included the Cochrane Database of Systematic Reviews through August 14, 2011, and hand searched references in potentially eligible articles and conference proceedings of major emergency medicine organizations for the previous 2 years. We selected studies that reported diagnostic performance of PERC, reported original research, and were conducted in the ED, with no language restrictions. Two investigators independently identified eligible studies and extracted data. We used contingency tables to calculate sensitivity, specificity, and likelihood ratios.

Results: We found 12 qualifying cohorts (studying 13,885 patients with 1,391 pulmonary embolism diagnoses), 10 prospective and 2 retrospective, from 6 countries. Pooled sensitivity, specificity, positive likelihood ratios, and negative likelihood ratios for 10 included studies were 0.97 (95% confidence interval [CI] 0.96 to 0.98), 0.23 (95% CI 0.22 to 0.24), 1.24 (95% CI 1.18 to 1.30), and 0.17 (95% CI 0.13 to 0.23), respectively. Significant heterogeneity was observed in specificity ($I^2=97.2\%$) and positive likelihood ratio ($I^2=84.2\%$).

Conclusion: The existing literature suggests consistently high sensitivity and low but acceptable specificity of the PERC to rule out pulmonary embolism in patients with low pretest probability. [Ann Emerg Med. 2012;59:517-520.]

Please see page 518 for the Editor’s Capsule Summary of this article.

See Editorial, P. 524.

INTRODUCTION

Pulmonary embolism often has a nonspecific clinical presentation. Emergency physicians have been increasing their use of diagnostic testing in an attempt to avoid missing this potentially life-threatening diagnosis, increasing both cost and use of medical resources.

To try to limit such diagnostic testing, Kline et al. developed a clinical decision rule (pulmonary embolism rule-out criteria [PERC]) from parameters available at initial emergency department (ED) assessment. Patients meeting all 8 PERC (younger than 50 years, pulse rate <100 beats/min, SpO2 >94%, no unilateral leg swelling, no hemoptysis, no surgery or trauma within 4 weeks, no previous deep venous thrombosis or pulmonary embolism, and no oral hormone use) would appear to have a pretest probability low enough to defer D-dimer testing, thus removing any possibility of subsequent imaging. However, a recent systematic review of clinical decision rules for pulmonary embolism did not include PERC. Therefore, we performed a systematic review and meta-analysis to summarize the diagnostic accuracy of PERC.

MATERIALS AND METHODS

Data Collection and Processing

We performed a comprehensive search of the following biomedical databases through August 14, 2011: EMBASE, MEDLINE, SCOPUS, Web of Knowledge, and all the EBM reviews that included the Cochrane Database of Systematic Reviews. The search strategy is detailed in Appendix E1, available online at http://www.annemergmed.com. We hand searched references cited in potentially eligible articles and the previous 2 years’
conference proceedings of major emergency medicine organizations (Society for Academic Emergency Medicine and American College of Emergency Physicians, Canadian Association of Emergency Physicians). We performed PubMed searches of authors of identified abstracts to locate full articles otherwise missed.

Two investigators (B.S. and A.K.P.) independently screened first titles and abstracts and then full texts of potentially eligible articles. With no language restrictions, we selected studies that reported diagnostic performance of PERC to rule out pulmonary embolism, reported original research, and were conducted in the ED setting. We assessed interobserver agreement for study selection with Cohen’s weighted $\kappa$. Disagreements were resolved by consensus in the presence of third investigator (S.C.).

Because there is no single well-validated and widely accepted quality assessment tool for assessing the study methodology of clinical decision rule studies, we adopted a previously used checklist, which was developed after reviewing the QUADAS tool, recommendations by Laupacis et al, and recommendations by Stiell et al, among others. The checklist included 7 criteria: (1) patients selected in an unbiased fashion (consecutive or random sample); (2) the study sample included a wide-spectrum pulmonary embolism pretest probability for which PERC was designed; (3) predictor variables were assessed without knowledge of the outcome; (4) outcomes were assessed without knowledge of the predictor variables; (5) outcomes were accurately defined; (6) loss-to-follow-up rate of less than 10%; and (7) explicit interpretation of the risk score by clinicians in practice without knowledge of the outcome.

Responses for each criterion were dichotomized to yes and no/unclear. Two investigators (B.S., A.K.P.) independently graded study methodology quality with this checklist as yes and no/unclear. Interreviewer agreement was assessed and disagreements were resolved as above for study selection.

Two reviewers (B.S. and S.C.) then independently extracted data from the included articles, using a predesigned form, and assessed the reported quality of the methods. Data points were study characteristics (author, country, publication year, number of patients, study settings, study design, description of study participants, and duration of follow-up), subject selection (inclusion and exclusion criteria), PERC classification, outcome definition and measurement, outcomes in PERC positives and negatives, and follow-up. We determined the number of outcomes as reported by the study or calculated from reported sensitivity and specificity and cohort size. If a study reported more than 1 cohort, each cohort was included separately. Our primary outcome of interest was the diagnosis of pulmonary embolism or venous thromboembolism or death caused by venous thromboembolism within 90 days of initial ED evaluation. All the disagreements in data extraction were resolved by consensus in the presence of third investigator (S.C.). We excluded studies reported in abstract form only and those scoring less than 50% on the methodology checklist from quantitative data synthesis.

Primary Data Analysis

We describe continuous variables with either means with SD or medians with interquartile range as reported in the included studies. Categorical variables are expressed as frequency of occurrence and proportions. We used contingency tables to calculate the pooled sensitivity and specificity. A random-effects model was used to calculate pooled likelihood ratios and diagnostic odds ratios. We quantified the statistical heterogeneity between the studies with $I^2$ statistic, which indicates the proportion of variability in study estimate.

Because PERC was originally developed for patients with low clinical suspicion of pulmonary embolism, we performed a subgroup analysis based on pulmonary embolism prevalence. We further performed meta-regression with a generalization of Littenberg and Moses linear model to determine association between pulmonary embolism prevalence and PERC diagnostic accuracy. All analyses were performed with MetaDiSc software.

**Editor’s Capsule Summary**

*What is already known on this topic*

The pulmonary embolism rule-out criteria (PERC) are commonly used to identify patients for whom D-dimer or other testing can be deferred.

*What question this study addressed*

Are the PERC reliable?

*What this study adds to our knowledge*

In this meta-analysis of 11 studies from 6 countries, the PERC were highly sensitive (97%) in excluding pulmonary embolism but were nonspecific (23%).

*How this is relevant to clinical practice*

This pooled analysis strongly corroborates the safety of using PERC to defer D-dimer testing.

---

*Editor’s Capsule Summary*

*What is already known on this topic*

The pulmonary embolism rule-out criteria (PERC) are commonly used to identify patients for whom D-dimer or other testing can be deferred.

*What question this study addressed*

Are the PERC reliable?

*What this study adds to our knowledge*

In this meta-analysis of 11 studies from 6 countries, the PERC were highly sensitive (97%) in excluding pulmonary embolism but were nonspecific (23%).

*How this is relevant to clinical practice*

This pooled analysis strongly corroborates the safety of using PERC to defer D-dimer testing.
RESULTS

The results of study search and screening are displayed in Figure 1. Investigator agreement for screening study abstracts and then full text was excellent (k = 0.94 and 0.80, respectively).

The 11 final studies (Table E1, available online at http://www.annemergmed.com) included 12 cohorts comprising 13,885 patients from 6 countries (United States, United Kingdom, Switzerland, France, Belgium, and New Zealand). Eleven cohorts were urban and 1 was rural. Two cohorts were derived retrospectively, with the rest prospective. Included patients were 56% women, with a mean age of 52.9 years (SD 8.5 years). Follow-up ranged from 14 to 90 days. Results of the methodological quality checklist are shown in Table E2 (available online at http://www.annemergmed.com), with an investigator agreement of 0.66. Study populations appeared unbiased in all cohorts, and none had reported implementation of PERC in clinical practice.

Test performance of the included studies is shown in Figure 2. The pooled sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio were 0.97 (95% confidence interval [CI] 0.96 to 0.98), 0.23 (95% CI 0.22 to 0.24), 1.24 (95% CI 1.18 to 1.30), and 0.18 (95% CI 0.13 to 0.23), respectively. The overall proportion of missed pulmonary embolisms was 0.32% (95% CI 0.20% to 0.44%) (44 of 13,885 total cases). The pooled diagnostic odds ratio was 7.3 (95% CI 5.4% to 9.8%). Significant heterogeneity was observed in specificity (I^2 = 97.2%) and positive likelihood ratio (I^2 = 84.2%).

To perform the preplanned subset analysis, we divided studies into 2 groups according to pulmonary embolism prevalence above or below 10%. The pooled specificity was 0.16 (95% CI 0.14 to 0.17) in higher-prevalence group and 0.24 (95% CI 0.24 to 0.25) in lower-prevalence group.

We found no significant association between pulmonary embolism prevalence and PERC diagnostic performance on meta-regression analysis (coefficient of –0.038 [SE 0.023]; P = .14) or on relative diagnostic odds ratios (0.92; 95% CI 0.91 to 1.01).

LIMITATIONS

A major limitation of this meta-analysis is the small number of studies available for data synthesis. We could not assess the possibility of publication bias because the meta-analysis included fewer than 20 studies. Further, this analysis is limited by specificity heterogeneity.

DISCUSSION

We conducted a systematic review of the literature to assess the diagnostic performance of PERC in deferring the need for D-dimer when considering the diagnosis of pulmonary embolism in the ED. We found that when the pretest probability is low, PERC are highly sensitive in predicting pulmonary embolism, and D-dimer testing is thus unnecessary. These findings are at a confidence of what is considered “level 2 evidence,” ie, demonstrated accuracy in either 1 large prospective study including a broad spectrum of patients and clinicians or validated in several smaller settings that differed from one another.

Our meta-analysis observed consistently high PERC sensitivity across cohorts from 6 countries and both rural and
urban settings. Use of PERC could thus avoid the frequent expensive diagnostic imaging that typically results when a D-dimer result is positive.

Application of a well-validated clinical decision rule such as PERC adds objectivity to the diagnostic evaluation of pulmonary embolism and should decrease the excessive testing that has resulted from physician fears of missing pulmonary embolism and of litigation. Our meta-analysis reports consistent high sensitivity and negative predictive value of PERC, with missed pulmonary embolism in just 0.5% of patients. Two of the included studies (Hugli et al.\textsuperscript{12} and Righini et al.\textsuperscript{13}) report a higher frequency of missed pulmonary embolism and have raised concern about the reliability of PERC. However, their higher failure rate likely results from the higher pulmonary embolism prevalence observed in their European settings. The threshold for pulmonary embolism diagnostic imaging in the United States is substantially lower than that in Europe, presumably because of the higher litigation risk.\textsuperscript{14,19,20} The PERC rule was developed for use in low-probability settings.\textsuperscript{2,12,14}

The PERC rule is limited by its low specificity. We did not find a significant difference in PERC performance based on pulmonary embolism prevalence.

In summary, our meta-analysis has demonstrated high sensitivity for the PERC rule and evidence that the rule can be used in settings of low pretest probability with confidence. The major limitation of PERC is its low but acceptable specificity.

Supervising editor: Steven M. Green, MD

Author contributions: AKP and BS contributed equally to the article and collected data. AKP, BS, and SC were responsible for study design. AKP, DA, AS, SSM, and SC analyzed the data. BS was responsible for study selection. All authors participated in writing the article. SC takes responsibility for the paper as a whole.

Funding and support: By Annals policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist.

Publication dates: Received for publication July 9, 2011. Revisions received September 8, 2011, and October 10, 2011. Accepted for publication October 25, 2011. Available online December 15, 2011.

Address for correspondence: Subhash Chandra, MBBS, E-mail subhash.budania@gmail.com.

REFERENCES
APPENDIX E1.

Search strategy.
Ovid MEDLINE(R) 1948 to August Week 2 2011 # Searches

Results Search Type
1 *pulmonary embolism/di 3318 Advanced
2 venous thrombosis/ or venous thromboembolism/ 17149 Advanced
3 2 and ("pe" or (pulmonary adj embolism)).mp. [mp=protocol supplementary concept, rare disease supplementary concept, title, original title, abstract, name of substance word, subject heading word, unique identifier] 4300 Advanced
4 1 or 3 7306 Advanced
5 4 and (perc or (rule adj "out")).mp. [mp=protocol supplementary concept, rare disease supplementary concept, title, original title, abstract, name of substance word, subject heading word, unique identifier] 124 Advanced
6 4 and (exp emergency medical services/ or triag*.mp. or emergencies.mp. or emergency medicine/) [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword] 113 Advanced
7 4 and (decision support systems, clinical/ or diagnosis, differential/ or decision support techniques/ or diagnosis, computer assisted/ or algorithms/) 842 Advanced
8 7 and (risk factors/ or risk assessment/) 150 Advanced
9 5 or 6 or 8 399 Advanced
10 7 and ("sensitivity and specificity"/ or validat*.mp. or probability/ or likelihood*.mp. or low.mp.) [mp=protocol supplementary concept, rare disease supplementary concept, title, original title, abstract, name of substance word, subject heading word, unique identifier] 293 Advanced
11 9 or 10 588 Advanced
12 exp case control study/ or exp case study/ or exp clinical trial/ or exp intervention study/ or exp longitudinal study/ or exp major clinical study/ or exp prospective study/ or exp retrospective study/ 2207353 Advanced
13 follow up/ 493436 Advanced
14 comparative study/ or systematic review/ or meta-analysis/ or cohort*.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword] 744561 Advanced
15 11 and (12 or 13 or 14) 261 Advanced
16 11 and diagnostic accuracy/

WoS/Scopus
Topic=("pulmonary embolism*" AND (PERC OR "rule out" OR "clinical rule" OR "clinical predict* rule" OR "clinical probability" OR "low risk" OR "low probability" OR "no risk")) AND Topic=(ed OR emergenc* OR triage*)

Singh et al

Pulmonary Embolism Rule-Out Criteria

Embase 1988 to 2011 Week 33 # Searches Results Search Type
1 *pulmonary embolism/di 6447 Advanced
2 venous thrombosis/ or venous thromboembolism/ 28731 Advanced
3 2 and ("pe" or (pulmonary adj embolism)).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword] 3402 Advanced
4 1 or 3 9390 Advanced
5 4 and (perc or (rule adj "out")).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword] 162 Advanced
6 4 and (exp emergency medical services/ or triag*.mp. or emergencies.mp. or emergency medicine/) [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword] 113 Advanced
7 4 and (decision support systems, clinical/ or diagnosis, differential/ or decision support techniques/ or diagnosis, computer assisted/ or algorithms/) 842 Advanced
8 7 and (risk factors/ or risk assessment/) 150 Advanced
9 5 or 6 or 8 399 Advanced
10 7 and ("sensitivity and specificity"/ or validat*.mp. or probability/ or likelihood*.mp. or low.mp.) [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword] 293 Advanced
11 9 or 10 588 Advanced
12 exp case control study/ or exp case study/ or exp clinical trial/ or exp intervention study/ or exp longitudinal study/ or exp major clinical study/ or exp prospective study/ or exp retrospective study/ 2207353 Advanced
13 follow up/ 493436 Advanced
14 comparative study/ or systematic review/ or meta-analysis/ or cohort*.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword] 744561 Advanced
15 11 and (12 or 13 or 14) 261 Advanced
16 11 and diagnostic accuracy/

WoS/Scopus
Topic=("pulmonary embolism*" AND (PERC OR "rule out" OR "clinical rule" OR "clinical predict* rule" OR "clinical probability" OR "low risk" OR "low probability" OR "no risk")) AND Topic=(ed OR emergenc* OR triage*)
Table E1. Characteristics of included cohorts.

<table>
<thead>
<tr>
<th>Author Year Country</th>
<th>N (% Women)</th>
<th>Age, Mean (SD) or Median (IQR)</th>
<th>Subject Selection</th>
<th>Outcome Definition</th>
<th>PE Prevalence</th>
<th>Duration of Follow-up, days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wolf 2008 United States</td>
<td>134 (54.0)</td>
<td>58 (43–72)</td>
<td>Adults aged 18–85 y with clinically suspected PE, chest radiograph, and ECG included Exclusion: pregnant, hemodynamically unstable, known D-dimer level in recent past</td>
<td>Combination of: High-probability V/Q scan using modified Prospective Investigation of Pulmonary Embolism Diagnosis criteria Intermediate-probability V/Q scan with a high pretest clinical suspicion Contrast-enhanced CT scan chest Pulmonary angiogram Diagnosis of VTE</td>
<td>11.9</td>
<td>90</td>
</tr>
<tr>
<td>Hogg 2005 UK</td>
<td>425 (51.1)</td>
<td>38.3 (15.0)</td>
<td>Adults (&gt;18 y) presenting to ED with pleuritic chest pain Exclusion: pneumothorax, ECG changes of myocardial infarction, ischemia or pericarditis, pregnancy or trauma within 4 wk</td>
<td>Combination of: High probability V/Q scan with high clinical probability CT-pulmonary angiography Digital subtraction angiography</td>
<td>5.4</td>
<td>90</td>
</tr>
<tr>
<td>Kline 2004 United States, LR</td>
<td>1,427 (60.0)</td>
<td>47 (17)</td>
<td>Adults (&gt;18 y) with clinical suspicion for PE whom emergency physicians believed were at low risk to justify exclusion of PE on the basis of a negative D-dimer result</td>
<td>Combination of: CT angiography CT angiography-venography V/Q scan (followed by duplex ultrasonography of the extremities)</td>
<td>8.0</td>
<td>90</td>
</tr>
<tr>
<td>Kline 2004 United States, VLR</td>
<td>382 (56.0)</td>
<td>56 (18)</td>
<td>Adults (&gt;18 y) presenting with shortness of breath but emergency physician stated pulmonary embolism not the most likely diagnosis</td>
<td>Combination of: CT angiography CT angiography-venography V/Q scan (followed by duplex ultrasonography of the extremities)</td>
<td>2.4</td>
<td>90</td>
</tr>
<tr>
<td>Dachs 2010 United States</td>
<td>213</td>
<td></td>
<td>All the ED patients who underwent a CT scan to rule out PE</td>
<td>CT chest</td>
<td>8.5</td>
<td>90</td>
</tr>
<tr>
<td>Hugli 2011 Switzerland, France, Belgium</td>
<td>1,675 (56.7)</td>
<td>61 (45–76)</td>
<td>Adult outpatients treated in the ED, with a clinical suspicion of PE. Exclusion: contraindication to multidetector CT MDCT (ie, allergy to iodine contrast agents, creatinine clearance &lt;30 ml/min, or pregnancy), a terminal illness with an expected survival of &lt;3 mo, a previous documented diagnosis of PE or were receiving anticoagulant therapy at presentation.</td>
<td>Combination of: Positive MDCT or pulmonary angiography result High-probability V/Q scan Proximal deep venous thrombosis documented by compression ultrasonography.</td>
<td>21.3</td>
<td>90</td>
</tr>
<tr>
<td>Author Year Country</td>
<td>N (% Women)</td>
<td>Age, Mean (SD) or Median (IQR)</td>
<td>Subject Selection</td>
<td>Outcome Definition</td>
<td>PE Prevalence</td>
<td>Duration of Follow-up, days</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------</td>
<td>-------------------------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Beam 2007 United States</td>
<td>189</td>
<td>Adults (&gt;18 y) with clinical suspicion for PE for whom emergency physicians considered formal PE evaluation necessary</td>
<td>Combination of: CT scan and V/Q scan</td>
<td>4.2</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Righini 2005 Switzerland</td>
<td>762 (58.0)</td>
<td>61 (19)</td>
<td>Consecutive outpatients suspected of having PE</td>
<td>Combination of: Clinical probability assessment, D-dimer measurement, Venous ultrasonography, Helical CT, Pulmonary angiogram</td>
<td>25.7</td>
<td></td>
</tr>
<tr>
<td>Kline 2008 United States, New Zealand</td>
<td>8,138 (67.0)</td>
<td>49.1</td>
<td>Adults (≥18 f) with clinical suspicion for PE on emergency physician’s evaluation. Exclusion: (1) positive pulmonary vascular imaging study result in last 7 days, (2) patient indicated that the enrollment hospital was not his or her choice for follow-up, (3) patient would be lost to follow-up (eg, homeless, psychiatric disorders, international travelers, person arrested for felonies)</td>
<td>Combination of: A high-probability V/Q scan, CT angiogram, Conventional pulmonary angiogram, PE on autopsy</td>
<td>7.7</td>
<td>45</td>
</tr>
<tr>
<td>Kline 2010 United States</td>
<td>115</td>
<td>ED patients (&gt;17 y) admitted with chief complaints: chest pain, shortness of breath, respiratory distress, syncope, hypotension, palpitations, cough, altered mental status, or syntax indicating that the patient was sent from outside facility for PE evaluation</td>
<td>Combination of: D-Dimer, Pulmonary vasculature imaging, Venous ultrasonography</td>
<td>1.74</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Courtney 2006 United States</td>
<td>315</td>
<td>ED patients with any testing (V/Q scan, CT scan or D-dimer test) to evaluate for PE</td>
<td>Combination of: V/Q scan, CT scan, D-dimer test</td>
<td>4.44</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Crichlow 2011 United States</td>
<td>110 (74)</td>
<td>46.4 (30.8–62.0)</td>
<td>Patients who received CT pulmonary angiography or lower-extremity duplex ultrasonography</td>
<td>Combination of: CT pulmonary angiography, Lower-extremity duplex ultrasonography</td>
<td>5.26</td>
<td>90</td>
</tr>
</tbody>
</table>

IQR, Interquartile range; PE, pulmonary embolism; V/Q, ventilation-perfusion scan; CT, computed tomography; VTE, venous thromboembolism; LR, low risk; VLR, very low risk.
Table E2. Quality assessment study methodology score obtained by each study on the checklist.*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Were the patients selected in an unbiased fashion (consecutive or random sampling)?</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2) Do they represent a spectrum of pretest probability the PERC is used for?</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0/1*</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3) Were the predictor variables assessed without knowledge of the outcome?</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4) Were the outcomes assessed without knowledge of the predictor variables?</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5) Were the outcomes defined accurately (especially PE)?</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6) Was follow-up adequate (&lt;10% lost to follow-up)?</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7) Was there an explicit interpretation of PERC by clinicians in practice without knowledge of the outcome?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Yes = 1, no = 0.  †The study had 2 cohorts.  ‡Scores represent the score for the low-risk and very low-risk cohort, respectively.
10 questions to help you make sense of reviews

How to use this appraisal tool

Three broad issues need to be considered when appraising the report of a systematic review:

- Is the study valid?
- What are the results?
- Will the results help locally?

The 10 questions on the following pages are designed to help you think about these issues systematically.

The first two questions are screening questions and can be answered quickly. If the answer to both is “yes”, it is worth proceeding with the remaining questions.

You are asked to record a “yes”, “no” or “can’t tell” to most of the questions. A number of italicised prompts are given after each question.

These are designed to remind you why the question is important. Record your reasons for your answers in the spaces provided.

The 10 questions are adapted from Oxman AD, Cook DJ, Guyatt GH, Users’ guides to the medical literature. VI. How to use an overview. JAMA 1994; 272 (17): 1367-1371

© Public Health Resource Unit, England (2006). All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the Public Health Resource Unit. If permission is given, then copies must include this statement together with the words “© Public Health Resource Unit, England 2006”. However, NHS organisations may reproduce or use the publication for non-commercial educational purposes provided the source is acknowledged.
Screening Questions

1. Did the review ask a clearly-focused question?
   Consider if the question is ‘focused’ in terms of:
   – the population studied
   – the intervention given or exposure
   – the outcomes considered

2. Did the review include the right type of study?
   Consider if the included studies:
   – address the review’s question
   – have an appropriate study design

Is it worth continuing?

Detailed Questions

3. Did the reviewers try to identify all relevant studies?
   Consider:
   – which bibliographic databases were used
   – if there was follow-up from reference lists
   – if there was personal contact with experts
   – if the reviewers searched for unpublished studies
   – if the reviewers searched for non-English-language studies

4. Did the reviewers assess the quality of the included studies?
   Consider:
   – if a clear, pre-determined strategy was used to determine which studies were included. Look for:
     – a scoring system
     – more than one assessor
5. If the results of the studies have been combined, was it reasonable to do so?

Consider whether:

– the results of each study are clearly displayed
– the results were similar from study to study (look for tests of heterogeneity)
– the reasons for any variations in results are discussed

6. How are the results presented and what is the main result?

Consider:

– how the results are expressed (e.g. odds ratio, relative risk, etc.)
– how large this size of result is and how meaningful it is
– how you would sum up the bottom-line result of the review in one sentence

7. How precise are these results?

Consider:

– if a confidence interval were reported. Would your decision about whether or not to use this intervention be the same at the upper confidence limit as at the lower confidence limit?
– if a p-value is reported where confidence intervals are unavailable
8. Can the results be applied to the local population?

Consider whether:

– the population sample covered by the review could be different from your population in ways that would produce different results

– your local setting differs much from that of the review

– you can provide the same intervention in your setting

9. Were all important outcomes considered?

Consider outcomes from the point of view of the:

– individual

– policy makers and professionals

– family/carers

– wider community

10. Should policy or practice change as a result of the evidence contained in this review?

Consider:

– whether any benefit reported outweighs any harm and/or cost. If this information is not reported can it be filled in from elsewhere?