Missing the obvious

Themes for Bandolier are sometimes planned, and sometimes they just creep up without one noticing. This month it is the latter, and only when it is written does the theme become clear. Each item demonstrates the way in which we, all of us, including researchers, just fail to see the obvious.

The obvious, of course, varies, and trying to define obvious is not easy, but we all know obvious when we see it. For instance, with inhaled corticosteroids the adverse event of oral candidiasis is obvious, and a review tells us that the risk is twice as great with inhaled steroids. But what is the actual rate? How many people in a hundred using inhaled steroids suffer it? No answer in a review, nor much in the literature, despite anecdotal comments that range from rare to 15% or more.

With methotrexate for ectopic pregnancy, the obvious is missing randomised trials, and we don’t know which of two regimens is right, even though we suspect that doctors are treating the right patient right.

With noncompliance issues, more anecdotal reports suggest that vast proportions of prescribed drugs are never used, but hard evidence is hard to find. If it were true, the costs would be enormous, and it would be an obvious target for change. But not many brownie points for academic pointy heads there, though.

With pharmaceutical company economic submissions, most had serious problems, at least in Australia up to 1997. Most of the flaws were obvious. If pharmaceutical companies are some of the most powerful organisations in the world, with more neurones per square inch than any other, how come they didn’t spot the obvious?

With obesity, a study tells us what we know, that overweight people die younger and lose about seven years of life compared with their thinner brothers and sisters. But it’s obvious, so why are we getting fatter not thinner?

And with diagnostic tests, we know that most reports suffer from massive bias. Yet here the obvious needs to be hammered home. In a review of hysteroscopy, probably only one or two of 65 reports were bias free. Yet 25,000 women were studied, and the result is probably that we just don’t know how good the test actually is. It’s obvious that diagnostic testing needs a new major effort to drag it into the evidence-based age.

Inhaled steroids for COPD

COPD is a big problem, and one that is getting bigger. This is partly because of ageing populations, but is also a reflection of past industry and smoking. The WHO predicts that COPD, by 2020, will be the fifth most prevalent disease and the third most common cause of death.

Inhaled steroids are used to combat airways inflammation. Meta analysis has shown that, in patients with clearly defined moderately severe COPD, relatively high daily doses of inhaled corticosteroids improve FEV1 over two years of treatment [1]. But the issue is not just one of ventilatory outcomes, but of important clinical outcomes, like exacerbations of COPD, or hospital admissions, or death, as well as adverse events. A new systematic review [2] adds a little to this story.

Review

Five electronic databases, including the Cochrane Library, were searched for placebo-controlled trials of inhaled corticosteroids for at least six months in COPD. Experts were also contacted to identify any other studies. The primary outcome sought was the frequency (or risk) of respiratory exacerbations. The definition of exacerbation was that used by the trials, and the frequency of exacerbations was calculated per patient-month of treatment. Other outcomes were the rate of decline in FEV1 and all-cause mortality.

Results

Nine randomised, placebo-controlled studies were found. All those reporting exacerbations were of sufficient quality so as to minimise bias. They differed greatly in size, in duration and in definition of exacerbation. Size varied from 26 to 1277 patients, with 3,926 in total. Duration varied from six to 40 months, with three trials lasting six months and the remainder at least two years.

In this issue

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inhaled steroids for COPD</td>
<td>1</td>
</tr>
<tr>
<td>Methotrexate for ectopic pregnancy</td>
<td>3</td>
</tr>
<tr>
<td>Noncompliance with antihypertensives</td>
<td>4</td>
</tr>
<tr>
<td>Interpreting pharmacoeconomics</td>
<td>5</td>
</tr>
<tr>
<td>Obesity and life expectancy</td>
<td>6</td>
</tr>
<tr>
<td>Hysteroscopy for endometrial disease</td>
<td>7</td>
</tr>
</tbody>
</table>
Exacerbations were defined and reported in six trials. The most common definition used was worsening of symptoms, usually with changing treatment and including use of oral steroids and/or antibiotics. One used cough and phlegm more than usual, and for another hospital admission for a respiratory condition.

The treatment most commonly used was inhaled budesonide at 800-1,600 µg/day (five trials), with fluticasone 1000 µg/day in two, and triamcinolone 1,200 µg/day and beclamethasone 1,500 µg/day each used in one trial. Control in all cases was usual care with placebo.

The average age of patients in the trials was 54-66 years, with variable percentages of current smokers, and with baseline FEV1 of about 1 to 2.5 litres.

Exacerbations

In six trials, exacerbations occurred at an average rate of 0.07 per month (0.8 a year) with placebo, with a reduced rate of 0.05 per month (0.6 a year) with inhaled corticosteroid (Figure 1). The relative risk for exacerbation was 0.68 (0.64 to 0.72). On an annual basis, the number needed to treat was 4.8 (4.0 to 5.9). This means that treating five patients with COPD with inhaled corticosteroids will prevent one exacerbation.

Death

Mortality in six trials averaged 4.1% with placebo, and 3.4% with inhaled corticosteroids (Figure 2). The difference was not statistically significant, with a relative risk of 0.84 (0.6 to 1.2).

FEV1

The rate of decline in FEV1 with placebo was very variable, with average reductions in trials of 12 to 180 mL. Over the nine trials, the mean difference with inhaled corticosteroid was to effect a weighted average reduction of 28 mL in the decline of FEV1.

Adverse events

These were not reported in detail, and absolute rates were not given. We are told that the frequency of oropharyngeal candidiasis was increased with inhaled corticosteroids, with a relative risk of 2.1 (1.5 to 3.1), and that a similar increase in risk was seen for skin bruising, with a relative risk of 2.1 (1.6 to 2.8).

Comment

Interesting stuff, and generally well done. The review highlights some important issues, though. First was that the definition of those things that affect patients healthcare providers, exacerbations in respiratory symptoms, were neither well nor consistently defined. Perhaps we need two definitions, relating to worsening in condition, and another more serious outcome of hospital admission because of worsening respiratory function.

Frustratingly little detail was given about adverse events like candidiasis. Knowing that it occurs twice as often with inhaled corticosteroids is much less useful than knowing how often it occurs. Recent attempts to elicit a number from knowledgeable folk has elicited responses ranging from rare to 15%.

What we do have here is a splendid example of systematic reviews telling us something about a treatment (though not nearly enough in terms of the balance of benefit and harm). More importantly, it helps to define the research agenda. The review almost tells us how to do better and more informative studies in future. Let’s hope that someone takes notice.

References:

METHOTREXATE FOR ECTOPIC PREGNANCY

What can you do when you want to compare two treatments, but there are no randomised trials, no comparative trials, and only observational data? A meta-analysis of two treatment regimens for methotrexate in ectopic pregnancy [1] reminds us of some of the problems. These are likely to be in the type of patient treated, and what the treatment actually is.

Review

The focus for the review was two main medical treatments that use methotrexate for ectopic pregnancy. One involves the use of a single dose of methotrexate, usually based on an injected dose of about 50 mg/sq metre. The alternative is to use 0.1 mg/kg intramuscularly on alternate days for up to four doses.

Searching used a single database for English-only studies. Included were those reporting the number of patients treated with either of the two standard protocols. Nonstandard dosing protocols, tiny studies with fewer than 10 cases, and studies examining interstitial, cervical or ovarian pregnancy were excluded. Studies were rated according to quality and completeness of data reporting, design, diagnosis reporting, and inclusion and exclusion criteria.

Information was available on individual patients. The outcome was failure of treatment defined as abandonment of medical management in favour of surgical management. Information on failure or success was collected together with human chorionic gonadotrophin (hCG) level, the presence of embryonic cardiac activity, the protocol, the number of doses, adverse events and hospital admission. Logistic regression was used to determine associations between outcome and factors.

Figure 1: Number of doses of methotrexate used in the two regimens

<table>
<thead>
<tr>
<th>Number of doses given</th>
<th>Single dose</th>
<th>Multidose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20%</td>
<td>80%</td>
</tr>
<tr>
<td>2</td>
<td>40%</td>
<td>60%</td>
</tr>
<tr>
<td>3</td>
<td>20%</td>
<td>80%</td>
</tr>
<tr>
<td>≥4</td>
<td>40%</td>
<td>60%</td>
</tr>
</tbody>
</table>

Results

Information was available on 1,327 cases of women diagnosed with ectopic pregnancy from 26 studies, treated with methotrexate, mostly (80%) with a single dose regimen. Individual studies reported on 12 to over 300 women. Five of the studies included duplicated information on at least some women, and this duplicate information was not used.

The actual number of doses used was not always in accord with the protocol (Figure 1). More than a single dose was used in 15% of women in whom a single dose was planned, and a single dose only was used in 10% of women in whom it was planned to use multiple doses.

Failure of medical management occurred in 11% of women treated with a single dose and 7% of women treated with multiple doses. Women treated with multiple doses had a significantly higher hCG level, and increasing hCG was significantly associated with failed treatment. The presence of embryonic cardiac activity was significantly associated with the failure of medical treatment. A statistically significantly higher failure rate was seen for single dose compared to multiple dose regimen, particularly when the analysis was controlled for the hCG value and the presence of embryonic cardiac activity.

There was no difference in hospital admission (12%), abdominal pain (22%) or other adverse event rates (33%) when the analysis was adjusted for the hCG level.

Comment

Overall, the success rate was 89%, with adverse events were minor and self-limiting. Only 1 woman in 10 treated with methotrexate needed surgery.

Women treated with single dose methotrexate were more likely to fail medical management of ectopic pregnancy than those treated with multiple doses. But women treated with single doses had lower hCG levels, and lower hCG was independently associated with lower failure rates. Clinicians may have been using the single dose regimen for women with good prognosis and using the multiple dose regimen preferentially for women with poorer prognosis, or who were more advanced in their pregnancy. They may have been using the right treatment for the right patient.

Determining whether there is a true difference in efficacy or harm proved to be extremely difficult, even with a good analysis of a reasonable number of individual patients. Perhaps what we learn most from this interesting meta-analysis, apart from confirming the utility of randomised trials from their absence, is that this treatment is effective and that clinicians seem to be making the best choice for their patients. But the optimum dosing may be more than one, and not more than four doses.

References:
NONCOMPLIANCE WITH ANTIHYPERTENSIVES

Bandolier is interested in whether patients actually take the medicines they are prescribed. This used to be called compliance, though concordance is now the favoured term. Anecdotally, not taking medicines is thought to be common, and one experienced pharmacist recently told Bandolier that in the UK up to 40% of medicines prescribed are not actually taken when the prescriptions are filled.

There are several steps, though:

1. Having a prescription.
2. Having the prescription filled.
3. Taking the medicine.
4. Reporting how often the medicine prescribed is actually taken.

A new US study from Havard [1] suggests that, for antihypertensive medicines, there is little “concordance between step 1 and steps 2 and 4.

Study

Patients with diagnosed hypertension treated in a health maintenance or veterans organisation in New England formed the population. Medicines were either free, or had a minimal co-payment to be made by the patient.

Computer systems captured diagnoses, prescriptions, physician and hospital visits. Records for patients with a diagnosis of hypertension during 1996 and who were in the system for a full year after the first antihypertensive medication prescription were collected, and a random sample asked to participate. Of the 500 of about 1,000 patients who were willing to participate, 200 patients taking a single antihypertensive agent were selected.

A telephone survey instrument was developed after literature review and discussion with physicians, and tested for feasibility. It assessed self-reported compliance and other factors. Data on prescriptions filled for days of supply over the 365 days following the first prescription was collected.

Results

The 200 patients were mostly (80%) over 55 years, about 60% were men, and had been hypertensive for an average of 12 years. Almost all of them reported a high level of compliance, with only five of the 200 reporting using their medicines on fewer than 300 of 365 days.

In fact, over half of them filled prescriptions for fewer than 300 days (Figure 1), with 11% collecting less than half the medicine they should have taken, and six (3%) filling no prescriptions at all while claiming a very high degree of compliance.

Those who filled prescriptions for more than 80% of days were no different from those filling prescriptions for fewer than 80% of days, in terms of age, sex, education, or any other of a range of variables, though the number of provider contacts in the previous six months may have been related.

Comment

No surprise there, then. But there is a quantification of unfilled prescriptions, and excellent evidence that, at least for hypertension, that self-report of compliance by patients is likely to be unhelpful. In other areas, like anticoagulation with warfarin where INR can be used as a check, things may be different. What this study does not address is prescriptions filled but medicines not taken.

Compliance (concordance) is an important issue to which Bandolier would like to return. Readers who are aware of good studies can help by reporting them, as searching is not straightforward.

Is this at all important? It may have major economic implications. Bandolier hears of the amount of returned medicines being weighed in tons in just one area. Yet we have constraints on effective medicines. Waste in healthcare should not be allowed.

References:
INTERPRETING PHARMACOECONOMICS

A Bandolier reader wondered why it had not examined a study of pharmacoeconomic submissions to an Australian variety of NICE [1]. The simple answer was that it had failed to swim into our ken, despite our vigilance. So a few years on an opportunity to put right our oversight. What it does is review submissions over the period 1994 to 1997, and highlight errors that were found at quite a high rate.

To some extent the information is out of date, because it partly predated new submission rules. Since 1997 submission of pharmacoeconomic data has become common in many countries (including the UK), and there is better understanding of evidence rules. But there are lessons to be learned, especially by those who have to examine economic analyses and make decisions based on them.

Study

The study was based on 326 major applications to the Australian Pharmaceutical Benefits Scheme between 1994 and 1997. Pharmaceutical companies submit information which is used to determine reimbursement issues. The applications were reviewed in detail, including checking literature, and rerunning searches, validating key assumptions and checking computer or other models. A technical subcommittee then reviews the submission and makes a final recommendation to the federal health minister.

Problems with submissions were regarded as significant if both the evaluators and the technical subcommittee considered that the problem could have a serious bearing on the decisions made.

Results

Most applications involved new drugs, or major changes to indication, conditions of use, or price. Of the 326 submissions, 279 (86%) were economic analyses based on randomised trials, with 238 (73%) containing direct comparisons of a new agent and a chosen comparator. Indirect comparisons with a common comparator were used in 41 (13%). Twenty-six (8%) were based on quasi-experimental designs and 21 (6%) were based on uncontrolled data (Table 1). Meta-analyses were used in 64 of the submissions.

Serious problems of interpretation were found in 218 (67%) submissions, and 31 had more than one problem, giving 249 serious problems in total. The main problems in these 218 submissions are shown in Table 2. Examples of problems encountered included:

- No randomised trials
- Identification of additional trials contradicting claims
- Trials of poor quality
- Trials too small
- Trials too short
- Trials not appropriate for indication
- Inappropriate sub group analysis
- Surrogate rather than actual outcomes
- Choice of comparator
- Economic models based on inadequate information
- Calculation errors

Comment

None of these problems should be any surprise to regular readers of Bandolier, because they feature regularly in these pages as problems in assessing evidence. That pharmaceutical companies in Australia failed to recognise these issues in 1994-1997 should concern us for two reasons. The first is that most companies are now multinational and the second is that there was no improvement observed over the period of the study.

Table 1: Types of studies used in economic evaluations

<table>
<thead>
<tr>
<th>Design</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCT direct comparison</td>
<td>238</td>
<td>73</td>
</tr>
<tr>
<td>RCT indirect comparison</td>
<td>41</td>
<td>13</td>
</tr>
<tr>
<td>Quasi-experimental</td>
<td>26</td>
<td>8</td>
</tr>
<tr>
<td>Uncontrolled</td>
<td>21</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>326</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Table 2: Types of problem found in the 218 submissions that had problems

<table>
<thead>
<tr>
<th>Problem</th>
<th>Submissions</th>
<th>Details</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial efficacy issues</td>
<td>154</td>
<td>Availability of trials</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Poor quality trials</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interpretation of results</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Use of surrogate outcomes</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Determining therapeutic equivalence</td>
<td>26</td>
</tr>
<tr>
<td>Comparator issues</td>
<td>15</td>
<td>Uncertainty about choice of comparator or inappropriate comparator</td>
<td>6</td>
</tr>
<tr>
<td>Modelling issues</td>
<td>71</td>
<td>Technical aspects of the model</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unsubstantiated assumptions</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Uncertainty about costs</td>
<td>13</td>
</tr>
<tr>
<td>Calculation errors</td>
<td>9</td>
<td>Errors introducing serious inaccuracies in estimation of cost-effectiveness ratios</td>
<td>4</td>
</tr>
</tbody>
</table>

Problems found with 218 submissions
The authors are really quite gentle with the sponsors of the submissions, rightly identifying that pharmacoeconomic analysis is often a difficult and complex process. They do not believe the problems arise from any deliberate intent to deceive, but rather arose from a failure to take on board the requirements of quality evidence, process, and transparency.

Are things better now? It is difficult to know, without another analysis like this from Australia or elsewhere. Bandolier suspects that there remain huge lacunae of ignorance in the pharmaceutical world. A test is to use key words or phrases, like Cochrane, systematic review, NNT, or even Bandolier.

Will things change? You bet they will, because “fourth hurdle” issues alone will demand that change. Industry will learn that evidence-based medicine sells drugs, and they will find that it makes for good pharmacoeconomic arguments, too.

References:

OBESEITY AND LIFE EXPECTANCY

Being overweight is properly regarded as being a bad thing as pointed out in a report from the National Audit Office featured in Bandolier 85. What we need, perhaps, are better ways of telling people just what effect it has both on the quality and quantity of life. A Dutch analysis of the famous Framingham study quantifies the loss of life expectancy from being overweight [1].

Study

Information from 3,457 people who were 30-49 years old in about 1950 with initial height, weight and smoking status information on entry formed the cohort. Those chosen were neither underweight nor had cardiovascular disease at entry, and had at least four years of follow up. They were classified into three groups, with BMI between 18.5 and 24.9 kg/sq metre, 25 to 29.9 kg/sq metre, and 30 kg/sq metre or greater.

The main outcome was life expectancy at 40 years of age over a follow up of 40 years.

Results

There were 728 deaths in the period of four to 28 years of follow up, and 919 deaths in the period between 29 and 40 years of follow up. The probability of death increased with each step increase in BMI. Adjustment for physical activity at baseline and education made no difference to results. Smoking status at baseline, but not sex, modified effects of obesity, and results were given for smoking and non smoking women and men.

People with a BMI of 30 kg/sq metre or more at baseline, both women and men, smokers and non smokers, lost an average of seven years of life (Figure 1). For those with a BMI between 25 and 30 kg/sq metre fewer years were lost, and significant loss of years of life was restricted to female non smokers.

After 20 years, most survivors (67%) were in the same BMI category as they were initially. Of the remainder, 27% had increased by at least one BMI category, while only 6% had fallen by at least one BMI category.

Comment

This study makes explicit the decreased life expectancy associated with obesity. Those with BMI of 30 kg/sq metre or more can expect to have seven fewer years of life than their slimmer brothers and sisters. The effect is independent of sex and smoking, but the reason cannot be gleaned from this study, but it undoubtedly stems from a number of factors, including diet and exercise.

We know (Bandolier 78) that a healthy lifestyle greatly reduces the risk of heart attack and stroke, and we know they benefit in other areas too. If being overweight is a marker of an unhealthy lifestyle, the fact that it takes years off a life should be no big surprise.

References:

Figure 1: Years of life lost from age 40, compared with BMI less than 25

<table>
<thead>
<tr>
<th></th>
<th>BMI 25-30</th>
<th>BMI &gt;30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female nonsmoker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female smoker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male nonsmoker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male smoker</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Years of life lost from age 40, relative to BMI less than 25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female nonsmoker</td>
</tr>
<tr>
<td>Female smoker</td>
</tr>
<tr>
<td>Male nonsmoker</td>
</tr>
<tr>
<td>Male smoker</td>
</tr>
</tbody>
</table>

Bandolier 112 6 www.ebandolier.com
HYSTEROscopy for Endometrial Cancer and Hyperplasia

Diagnostic test evidence is different from that of therapeutics. It is more difficult because the quality of studies is often very low, so that they suffer from bias (Bandolier 26 & 70), and high quality studies are rare. It is also difficult because the way in which studies and meta-analyses are reported are hard to understand, and are neither intuitive nor, for most people, immediately useful (Bandolier 28). A systematic review of hysteroscopy [1] offers an opportunity to examine some of these issues again.

Review

The review focused on observational studies in which hysteroscopy was compared with results of the reference standard of endometrial histology. Verification of hysteroscopy diagnosis was followed either at the same time or after a short delay. The outcome was the accuracy of endometrial cancer and hyperplasia diagnosis.

Searching involved MEDLINE and EMBASE to the end of 2001, as well as the Cochrane Library, and without language restriction. Studies retrieved were assessed for methodological quality using a five-item hierarchy (Table 1). Studies in levels 1-3 were considered to be high quality, and 4 and 5 low quality. Technical failure in hysteroscopy so that no diagnosis was made were categorised as failed procedures. Information from the studies was collected on setting and pre or postmenopausal status.

Results

There were 65 primary studies involving 26,346 women. Of these studies, only 12 (18%) were of high quality (Table 1). Only one study was of the ideal quality, that is, an independent, blind comparison with reference standard among an appropriate population of consecutive patients.

In 35 studies failure rates were reported, with an overall failure to make a diagnosis with hysteroscopy of 3.6%. Potentially severe uterine complications were reported in eight cases, but only 19 studies with 9,413 procedures explicitly stated an intention to report them. A worst case risk of a serious complication would then be 1 in 1177 cases.

Endometrial cancer

For endometrial cancer there were 56 unique studies with 61 sets of data, only 11 of which were deemed of high quality. The post-test probability of endometrial cancer with positive hysteroscopy and with a prevalence of 3.9% is shown in Figure 1. There was considerable variability according to high versus low quality, by different settings, and by menopausal status.

In high quality studies the likelihood ratio for a positive test was reported as 35, giving a post-test probability of endometrial cancer of 59%. The likelihood ratio for a negative test was reported as 0.2, giving a post-test probability of endometrial cancer of 0.8%.

Endometrial disease

Endometrial disease was defined as endometrial cancer, hyperplasia, or both. For endometrial disease there were 41 unique studies with 71 sets of data, only 12 of which were deemed of high quality. The post-test probability of endometrial disease with positive hysteroscopy and with a prevalence of 10.6% is shown in Figure 2. There was considerable variability according to high versus low quality, by different settings, and by menopausal status.

In high quality studies the likelihood ratio for a positive test was reported as 5.5, giving a post-test probability of endometrial disease of 39%. The likelihood ratio for a negative test was reported as 0.3, giving a post-test probability of endometrial disease of 3.5%.

Comment

The definition of high quality studies in this review was probably justified. Certainly those in grades 4 and 5 are studies known to be associated with bias. There may still be a question mark over whether non-blind comparisons are subject to bias, and if they were, then 10 out of the 12 studies rated high quality would also be subject to possible bias. That would leave only two studies, not much grist for the mill of meta-analysis there.

Why then bother to perform an analysis using studies known to have potential faults? Perhaps because there’s nothing else. Good studies of diagnostic tests are exceptional, as this review proves again.

Table 1: Definitions of quality criteria for study and number of studies with particular quality scores. Scores 1-3 were rated high quality

<table>
<thead>
<tr>
<th>Quality</th>
<th>Definition</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>An independent, blind comparison with reference standard among an appropriate population of consecutive patients</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>An independent, blind comparison with reference standard among an appropriate population of nonconsecutive patients or confined to a narrow population of study patients</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>An independent, nonblind comparison with reference standard among an appropriate population of consecutive patients</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>An independent, nonblind comparison with reference standard among an appropriate population of nonconsecutive patients or confined to a narrow population of study patients</td>
<td>42</td>
</tr>
<tr>
<td>5</td>
<td>An independent, blind comparison among an appropriate population of patients, but reference standard not applied to all patients</td>
<td>11</td>
</tr>
</tbody>
</table>
So even if we include studies of poor design, what was the result? One way of looking at it for cohorts of 1,000 theoretical women is shown in Figure 3.

- Women with a positive test for endometrial cancer have an 8 out of 10 chance of actually having endometrial cancer, and those with a negative test have only a 1 in 200 chance of having endometrial cancer.

- Women with a positive test for endometrial disease have a 7 out of 10 chance of actually having endometrial cancer, and those with a negative test have only a 1 in 40 chance of having endometrial cancer.

But these results may just be the figments of the imagination of biased trials. To know how good hysteroscopy actually is, we have to wait for new studies. We also need reassurance that the gold standard, of histopathology, really is gold and not some shiny base metal. Agreement in histopathology is not always high, and the quality of histopathology will have a major impact on the results.

Once again, the ethics of involving patients in studies that are unlikely to produce a useful result, in this case the best part of 24,000 women, should be considered. Once again, the deficient design quality of diagnostic test studies has been highlighted.

Reference: