CLINICAL GUIDELINES

About time: diagnostic guidelines that help clinicians

R Foy, P Warner

Qual Saf Health Care 2003;**12**:205–209

Clinical guidelines often make recommendations on the use of diagnostic tests. Compared with sensitivity and specificity, the use of pre- and post-test probabilities allows a more explicit and rational selection and interpretation of diagnostic tests. Ideally, clinical guidelines relating to diagnosis should routinely incorporate this information to enhance individualised decision making. We report our experience of incorporating pre- and post-test probabilities into a guideline on the investigation of women with postmenopausal bleeding developed by the Scottish Intercollegiate Guidelines Network. Issues relating to their application are highlighted, including the limitations of available evidence on diagnostic tests and prevalence of disease, acceptability to guideline users, and the uncertain impact on actual clinical decision making. Despite these potential difficulties, the incorporation of data on pre- and post-test probabilities into the development and presentation of guideline recommendations may offer an important opportunity to make clinical decision making more transparent for both clinicians and patients.

See end of article for authors' affiliations

Correspondence to:
R Foy, Centre for Health Services Research, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4AA, UK; R.C.Foy@ncl.ac.uk

Accepted for publication 23 March 2003

Clinical guidelines are "systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances".[1] Guidelines often make recommendations on the use of diagnostic tests. Parameters essential to the evaluation of diagnostic test performance, such as sensitivity and specificity, are neither intuitive nor readily transferable to clinical situations.[2] Much has been written highlighting the superiority of likelihood ratios in providing information more relevant to clinical decisions concerning individual patients,[3–7] yet these parameters are seldom presented in clinical guidelines. The Scottish Intercollegiate Guideline Network (SIGN) guidelines are developed by multidisciplinary groups according to rigorous criteria, including explicit methods of appraising evidence and grading recommendations.[8] None of 17 SIGN guidelines published between 1996 and 2001 which dealt substantially with diagnostic processes mentioned the use of likelihood ratios in the development of recommendations.

We were members of a SIGN guideline development group that addressed the selection and interpretation of diagnostic procedures for the assessment of women with postmenopausal bleeding. During the course of the group's deliberations we were struck by the discomfort expressed by other group members and the target audience (mainly encompassing gynaecologists, radiologists and general practitioners) when "newer" approaches to understanding diagnostic tests were introduced. This paper draws on our deliberations to demonstrate how the development of diagnostic recommendations within clinical guidelines can be improved and highlights potential difficulties with their application.

LIMITATIONS OF SENSITIVITY AND SPECIFICITY IN THE CLINICAL ENCOUNTER

Postmenopausal bleeding represents one of the most common reasons for referral to gynaecological services, largely because of the need to detect or exclude endometrial carcinoma. The most rigorously evaluated investigation is transvaginal ultrasonography[9 10] which measures endometrial thickness. The diagnostic rationale is that, for an individual woman, the greater the measured endometrial thickness, the higher the probability that cancer is present. Transvaginal ultrasonography is attractive as an initial investigation as it is non-invasive and well tolerated. Sufficient reassurance from a negative result can help avoid unnecessary and more invasive procedures seeking a tissue diagnosis. The Guideline Development Group had to decide whether or not to recommend the use of this as a first line test and, if so, what diagnostic threshold to recommend.

Although transvaginal ultrasonography obtains an actual measurement of endometrial thickness, test results are typically reported simply as "positive" or "negative" depending on whether the thickness is above or below a specified threshold. In such circumstances, performance of the diagnostic test across the study group is usually summarised in terms of sensitivity and specificity. Sensitivity is the probability of testing positive if the disease is truly present. Specificity is the probability of testing negative if the disease is truly absent.

Box 1 presents an illustrative example of a woman referred to a gynaecology outpatient department for assessment of postmenopausal bleeding. She and her general practitioner were concerned about the possibility of endometrial cancer. Transvaginal ultrasound was performed using a threshold of >5 mm to define abnormal endometrial thickening. Based upon findings from a recent meta-analysis, transvaginal ultrasonography using this threshold has a sensitivity of 91% and specificity of 58% for the detection of endometrial carcinoma.[10] On hearing her test result was negative, the woman questioned her gynaecologist:

Box 1 Use of pre-test and post-test probabilities to judge the usefulness of a diagnostic test for an individual patient

A woman aged 55 was referred to a gynaecology outpatient department for assessment of postmenopausal bleeding. She and her general practitioner were concerned about the possibility of endometrial cancer. She had no history of using hormone replacement therapy. Transvaginal ultrasound was performed using a threshold of 5 mm or less to define a negative result. The sensitivity of the test is 91% and the specificity is 58%.

Using pre- and post-test probabilities, the pre-test probability that this woman has endometrial cancer is estimated as 10%. Following a positive test, her probability of cancer is 19%, and following a negative test, her probability of cancer is 1.7%.

Woman: This means I can't have cancer then?

Gynaecologist: It's unlikely. The test picks up 91 out of every 100 cancers.

Woman: So I have a 9 in 100 chance of actually having cancer then.

Gynaecologist: Mmm . . . not exactly. It depends . . .

Following a negative test, both the woman and her gynaecologist may wish to have a more individualised estimate of risk, an idea of how likely it is that she has cancer. Apart from highlighting some of the problems clinicians encounter with interpreting sensitivity and specificity,[11] this also shows that, regardless of the test result, if only these parameters are known it is not always clear whether further investigation is justified.

ENHANCING THE INTERPRETATION OF DIAGNOSTIC TEST RESULTS

A more informed decision involves the use of Bayes' theorem. The pre-test probability of disease is combined with test performance to estimate the post-test probability of cancer. For each of the two possibilities—a positive or negative transvaginal ultrasonography result—a formula involving the sensitivity and specificity of the test can be used to calculate the corresponding likelihood ratio. The post-test probability of cancer is then derived from the pre-test probability and the relevant likelihood ratio using either a formula or nomogram. In the examples that follow the actual likelihood ratios are not presented, largely because clinicians are likely to be more conversant with probabilities.[12]

The pre-test probability (also known as prior probability or prevalence) quantifies what proportion of patients has the condition of interest—in this case the proportion of women presenting with postmenopausal bleeding who have endometrial cancer. The positive post-test probability estimates the probability that a patient has disease given a positive test result. Conversely, the negative post-test probability estimates the probability that a patient has disease given a negative test result.

This is also illustrated in box 1. Assuming a pre-test probability of 10% and using a test threshold of 5 mm, the woman's post-test probability of cancer following a positive test is estimated at 19%. Her post-test probability of cancer given a *negative* test result is 1.7% (approximately 1 in 60). Therefore, if the woman asks about the possibility that she has cancer given a negative result, her gynaecologist can reply: "About 1 in 60". It is uncertain whether she and her gynaecologist would be sufficiently reassured by a 1.7% probability to render further investigation unnecessary. However, using the post-test probability provides a quantifiable estimate of the probability of disease being present for the patient in question, rather than largely relying upon clinical intuition.

CLARIFYING THE RATIONALE FOR THE SELECTION OF DIAGNOSTIC TESTS BY CLINICIANS

Some clinicians prefer lower test thresholds to determine a negative result. Using a threshold of 3 mm is more sensitive (98%) but less specific (53%) than 5 mm. What would using the 3 mm threshold mean for the woman in the clinical scenario in box 1? Based upon a pre-test probability of 10%, her post-test probability of cancer following a negative test result is 0.4% (instead of the 1.7% when a less sensitive threshold was used). Both the woman and her doctor may be satisfied that this probability of cancer (1 in 250) is sufficiently low that further investigation is unnecessary unless symptoms recur.

If ultrasonography can be requested using a specified thickness threshold, each threshold for defining a positive or negative result represents, in effect, a different diagnostic test. The use of post-test probabilities can inform individual decisions about which threshold value for transvaginal ultrasonography is most appropriate. Despite a negative test result using the 5 mm threshold, the woman in this case might prefer further investigation and hence initial transvaginal ultrasonography may not alter subsequent clinical management. The use of post-test probabilities indicates that a more sensitive 3 mm threshold test might be necessary to ensure that a negative result provided adequate reassurance. Calculating post-test probabilities can therefore prompt an explicit consideration of whether an investigation (at a specified threshold) is worth performing.

POTENTIAL PITFALLS AND POSSIBLE SOLUTIONS

The incorporation of pre-test and post-test probabilities into clinical guidelines may allow more rational selection and interpretation of diagnostic tests. However, guideline developers and users need to be aware of potential difficulties associated with their application.

Quality of evidence

Unfortunately, the evidence base for a 3 mm threshold is sparse and less reliable than that for the 5 mm threshold.[10] Poorer quality studies may overestimate the accuracy of diagnostic tests.[13] The Guideline Development Group faced the dilemma of whether to recommend transvaginal ultrasonography at a 5 mm threshold which would probably be too insensitive as a diagnostic test in most women, or to recommend a 3 mm threshold based on evidence possibly overestimating its accuracy. It was decided to make approximate adjustments to the likelihood ratio to account for bias (described in Appendix). After consideration of the issues there was a consensus decision that the 3 mm threshold should be recommended. For this the best and worst cases for negative post-test probabilities ranged from 0.6% to 0.8%. While acknowledging that this was an extrapolation, it was considered to be a reasonable pragmatic decision.

Uncertainty around estimates of test performance

Confidence intervals (CIs) demonstrate the degree of statistical uncertainty around point estimates of diagnostic accuracy. The robustness of recommendations can be checked if confidence intervals are used to provide a "worst case scenario" for test performance and subsequent post-test probabilities.[10]

In the case of the woman with a pre-test probability of 10% tested using the 3 mm threshold, her post-test probability of cancer given a negative result is estimated as 0.8% (using the most cautious likelihood ratio negative of 0.07). However, if uncertainty is allowed for (95% confidence), the probability of

Box 2 Definitions and formulae

Pre-test probability = prevalence (the proportion of the population with the disorder)
Pre-test odds = prevalence/(1 – prevalence)
Likelihood ratio for a positive test = sensitivity/(1 – specificity)
Likelihood ratio for a negative test = (1 – sensitivity)/specificity
Post-test odds = pre-test odds × likelihood ratio
Post-test probability = post-test odds/(post-test odds + 1)

cancer may be as high as 2.1%. Some women and clinicians may feel inadequately reassured even by a negative transvaginal ultrasonography result if aware of the level of prevailing uncertainty.

Estimates of post-test probability should reflect statistical uncertainty around estimates of both pre-test probability and test performance. Although guideline users may feel uncomfortable at having to confront the uncertainty around test results, this does prompt a more explicit consideration of the potential limitations of diagnostic approaches.

Estimating pre-test probabilities

Information on pre-test probability is central to the likelihood ratio approach to interpretation of diagnostic test results. However, there is a paucity of reliable prevalence data. The Guideline Development Group estimated the overall pre-test probability of endometrial cancer in women referred with postmenopausal bleeding to be 10%.[14–16] Estimating pre-test probabilities for more specific subgroups is problematic. The Group used extrapolated data to estimate a pre-test probability of 1% of endometrial cancer for women on sequential hormone replacement therapy presenting with unscheduled bleeding.[17 18] Such estimates may be contentious, particularly if they have a major impact on estimates of post-test probability and subsequent criteria for test selection and interpretation. If there is substantial doubt as to the pre-test probability that applies, then the post-test probability can be calculated for a range of pre-test values, particularly for the plausible upper limit for pre-test probability. If new data on relevant pre-test probabilities become available after publication, the revised post-test probability can then be calculated more accurately. In both cases this depends on guidelines explaining the formula used (box 2). More generally, pre-test probabilities can be raised or lowered according to levels of clinical suspicion, thus providing more pertinent post-test probabilities.[2]

Shifting pre-test probabilities

Pre-test probability may also alter over time within either setting as one or more of background risk factors for disease, consultation patterns, or referral thresholds change. For example, pre-test probabilities for patients consulting general practitioners may differ from those for patients referred to a hospital clinic. General practitioners may "filter" out lower risk patients from referrals to secondary care. Introducing direct access to hospital investigations may reduce referral thresholds so that patients with lower pre-test probabilities of disease are investigated. Subsequently, the post-test probabilities change, as can test performance.[19] There is therefore a need for up to date epidemiological studies on the prevalence of disease associated with indications for investigation in various settings such as primary and secondary care. In particular, more studies of diagnostic approaches are required in primary care as this is usually where the most critical decision—whether to refer for further investigation—is made.[20]

Acceptability to guideline users

Although pre-test and post-test probabilities offer a more transparent basis for clinical decision making, clinicians may be deterred from using them because of actual and perceived complexities in their application. When the draft guideline on postmenopausal bleeding was pre-tested on a range of clinicians, concern was expressed that using pre-test and post-test probabilities might be too complex (SIGN National Meeting, Edinburgh, 12 May 2000). This view is difficult to reconcile with the fact that probabilities are being used commonly to support and individualise treatment decisions.[21] In coronary heart disease prevention, the absolute benefits of treatment vary according to pre-treatment risk.[22] Risk assessment charts, now commonplace in coronary heart disease guidelines, enable clinicians to assess an individual patient's absolute risk based on a number of identifiable risk factors.

An advantage of pre-test and post-test probabilities is that they avoid the "one size fits all" approach, the basis for widespread scepticism of guidelines among clinicians.[23] In developing its recommendations, the Guideline Development Group suggested that a less than 1% probability of having cancer given a negative result would be sufficiently reassuring to justify the avoidance of further more invasive investigations. Individual women and other clinicians may hold different views about what constitutes a "safe" probability. Ideally, the selection of diagnostic tests should be driven by the acceptability of corresponding post-test probabilities. These data should be presented to allow individualised decisions to be made. We recommend extending the use of pre-test and post-test probabilities in diagnosis, for example, to criteria for referral and investigation of malignancies other than endometrial cancer.[24]

Impact on planning services

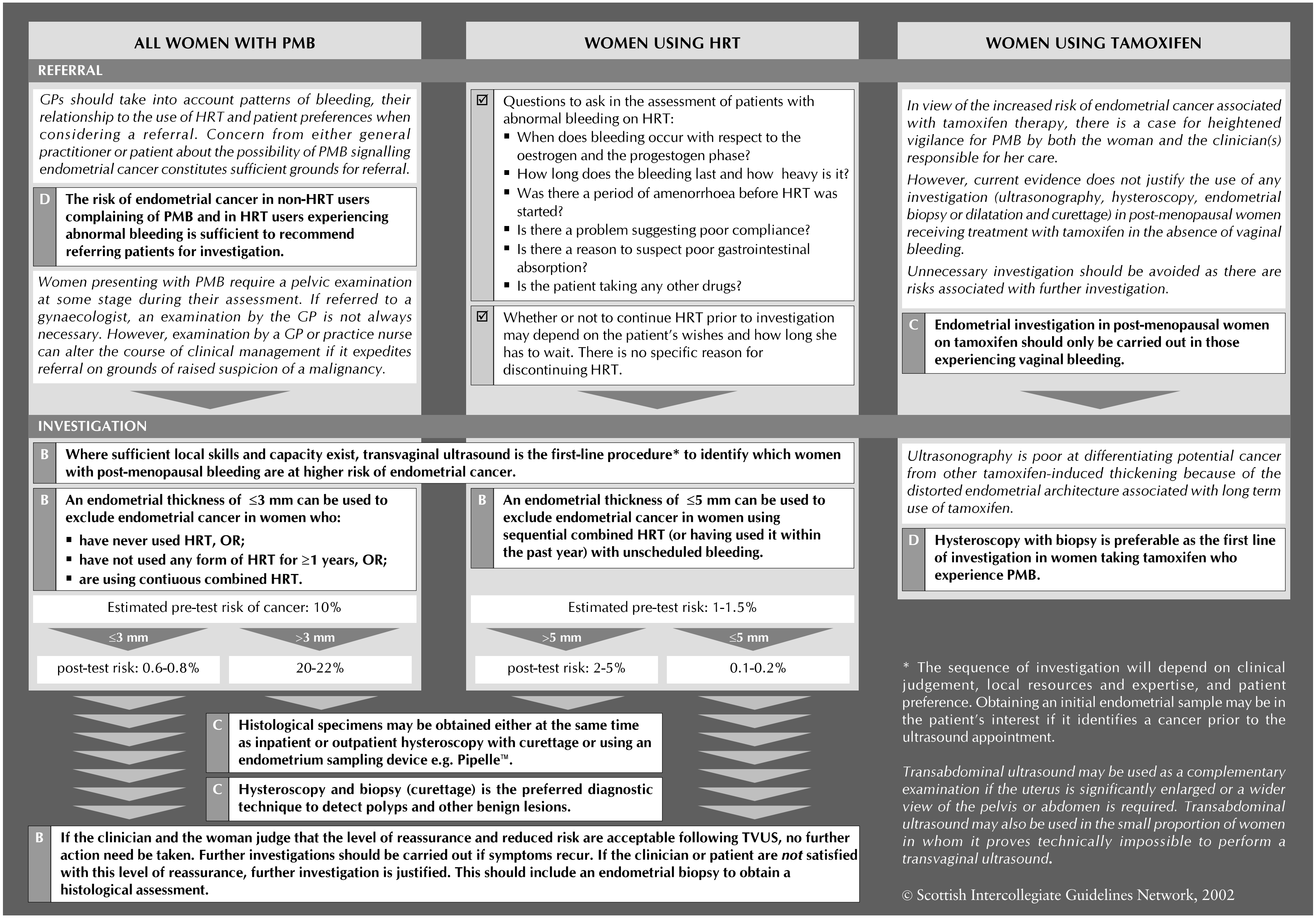
Sensitivity and specificity are still useful in the overall planning of healthcare programmes. Lower transvaginal ultrasonography thresholds tend to be more sensitive and hence miss fewer cancers. Such thresholds tend to be less specific and will result in more false positives and hence more patients being unnecessarily investigated. Guideline developers therefore need to balance the needs of individual patients against population needs in formulating recommendations. Health economic techniques can help make the costs and benefits involved in such trade-offs more explicit.

Impact on clinical decision making

There is a substantial body of research on how clinicians make decisions[25] which we cannot address in full here. Non-specific diagnostic guideline recommendations are more likely to result in inappropriate or potentially harmful decisions.[26] Presenting clinicians with information on likelihood ratios can improve their interpretation of diagnostic tests compared with information on sensitivity and sensitivity.[11] It is not known whether knowledge of pre- and post-test probabilities improves clinical decision making further. Teaching clinicians to make better judgements about disease probability may not alter treatment decisions.[27] However, it is unrealistic to expect that—by itself—enabling clinicians to estimate post-test probabilities of disease more accurately will lead to more rational decision making. Evidence on changing professional practice, for example, indicates that the simple distribution of clinical guidelines seldom changes clinical practice.[28] Therefore, as with any clinical guideline, active strategies are required to support the implementation of recommendations within diagnostic guidelines.

IMPLICATIONS FOR GUIDELINE DEVELOPMENT PROGRAMMES

The SIGN Guideline Development Group agreed to use pre- and post-test probabilities in the development and presentation of its recommendations. A summary of the relevant recommendations from the Quick Reference Guide is shown in fig 1. In England and Wales the National Institute for Clinical Excellence (NICE) has embarked on a guideline development programme and will need to consider how to develop



ALL WOMEN WITH PMB

WOMEN USING HRT

WOMEN USING TAMOXIFEN

REFERRAL

GPs should take into account patterns of bleeding, their relationship to the use of HRT and patient preferences when considering a referral. Concern from either general practitioner or patient about the possibility of PMB signalling endometrial cancer constitutes sufficient grounds for referral.

D **The risk of endometrial cancer in non-HRT users complaining of PMB and in HRT users experiencing abnormal bleeding is sufficient to recommend referring patients for investigation.**

Women presenting with PMB require a pelvic examination at some stage during their assessment. If referred to a gynaecologist, an examination by the GP is not always necessary. However, examination by a GP or practice nurse can alter the course of clinical management if it expedites referral on grounds of raised suspicion of a malignancy.

☑ Questions to ask in the assessment of patients with abnormal bleeding on HRT:

- When does bleeding occur with respect to the oestrogen and the progestogen phase?
- How long does the bleeding last and how heavy is it?
- Was there a period of amenorrhoea before HRT was started?
- Is there a problem suggesting poor compliance?
- Is there a reason to suspect poor gastrointestinal absorption?
- Is the patient taking any other drugs?

☑ Whether or not to continue HRT prior to investigation may depend on the patient's wishes and how long she has to wait. There is no specific reason for discontinuing HRT.

In view of the increased risk of endometrial cancer associated with tamoxifen therapy, there is a case for heightened vigilance for PMB by both the woman and the clinician(s) responsible for her care.

However, current evidence does not justify the use of any investigation (ultrasonography, hysteroscopy, endometrial biopsy or dilatation and curettage) in post-menopausal women receiving treatment with tamoxifen in the absence of vaginal bleeding.

Unnecessary investigation should be avoided as there are risks associated with further investigation.

C **Endometrial investigation in post-menopausal women on tamoxifen should only be carried out in those experiencing vaginal bleeding.**

INVESTIGATION

B **Where sufficient local skills and capacity exist, transvaginal ultrasound is the first-line procedure* to identify which women with post-menopausal bleeding are at higher risk of endometrial cancer.**

B **An endometrial thickness of ≤3 mm can be used to exclude endometrial cancer in women who:**

- **have never used HRT, OR;**
- **have not used any form of HRT for ≥1 years, OR;**
- **are using contiuous combined HRT.**

Estimated pre-test risk of cancer: 10%

≤3 mm	>3 mm
post-test risk: 0.6-0.8%	20-22%

B **An endometrial thickness of ≤5 mm can be used to exclude endometrial cancer in women using sequential combined HRT (or having used it within the past year) with unscheduled bleeding.**

Estimated pre-test risk: 1-1.5%

>5 mm	≤5 mm
post-test risk: 2-5%	0.1-0.2%

Ultrasonography is poor at differentiating potential cancer from other tamoxifen-induced thickening because of the distorted endometrial architecture associated with long term use of tamoxifen.

D **Hysteroscopy with biopsy is preferable as the first line of investigation in women taking tamoxifen who experience PMB.**

C **Histological specimens may be obtained either at the same time as inpatient or outpatient hysteroscopy with curettage or using an endometrium sampling device e.g. Pipelle™.**

C **Hysteroscopy and biopsy (curettage) is the preferred diagnostic technique to detect polyps and other benign lesions.**

B **If the clinician and the woman judge that the level of reassurance and reduced risk are acceptable following TVUS, no further action need be taken. Further investigations should be carried out if symptoms recur. If the clinician or patient are *not* satisfied with this level of reassurance, further investigation is justified. This should include an endometrial biopsy to obtain a histological assessment.**

* The sequence of investigation will depend on clinical judgement, local resources and expertise, and patient preference. Obtaining an initial endometrial sample may be in the patient's interest if it identifies a cancer prior to the ultrasound appointment.

Transabdominal ultrasound may be used as a complementary examination if the uterus is significantly enlarged or a wider view of the pelvis or abdomen is required. Transabdominal ultrasound may also be used in the small proportion of women in whom it proves technically impossible to perform a transvaginal ultrasound.

© Scottish Intercollegiate Guidelines Network, 2002

Figure 1 Summary of recommendations in the Quick Reference Guide from the SIGN guideline on the investigation of postmenopausal bleeding (PMB).[30]

Key messages

- The use of pre-test and post-test probabilities allows a more explicit and rational selection and interpretation of diagnostic tests.
- Clinical guidelines relating to diagnosis should routinely incorporate this information to enhance decision making.
- More epidemiological research is needed on the probability of disease being present across a range of settings to better inform pre-test probabilities.

and present recommendations about diagnostic tests or processes. For potential users of guidelines, available generic checklists do not offer specific advice on assessing guidelines concerned with diagnostic processes.[29]

We recommend that diagnostic clinical guidelines routinely present post-test probabilities, with some indication of the uncertainty around these estimates, since these are the statistics most relevant to patient management. We further recommend that the pre-test probabilities and likelihood ratios from which the post-test probabilities have been derived are also presented, for four main reasons:

(1) the joint impact on decision making of test performance and the patient's pre-existing risk of disease is made explicit;

(2) upstream to guideline development, further research priorities are highlighted. This includes reporting of pre- and post-test probabilities for different subgroups of patients in future diagnostic studies;

(3) revised post-test probabilities can be calculated if pre-test probabilities change or better performing tests become available; and

(4) clinicians need to be exposed to pre- and post-test probabilities through guidelines more frequently if they are to gain confidence in their application.

ACKNOWLEDGEMENTS

We are grateful to other members of the SIGN Guideline Development Group for contributing to the development of the views expressed in this paper, to Jayne Parry, David Weller and Phil Hannaford for comments on earlier drafts, and to Patrick Chien and Janesh Gupta for providing data.

APPENDIX: ADJUSTMENTS MADE TO SENSITIVITY AND SPECIFICITY FOR 3 MM THRESHOLD

The systematic review of transvaginal ultrasonography identified two primary studies evaluating a 3 mm threshold, neither of which was judged to be of high quality.[10] The review identified 21 primary studies evaluating a 5 mm threshold, four of which were judged to be of high quality.[10] Pooled sensitivity was only slightly higher for the 21 studies (91% *v* 89%) and pooled specificity was lower (58% *v* 68%). The *differences* in sensitivity and specificity between all 21 studies and the four high quality studies of the 5 mm threshold were applied to data reported for the 3 mm studies.

The resulting estimates of sensitivity and specificity were used to calculate the positive and negative likelihood ratios. This allowed an exploration of the extent to which the lower quality data for the 3 mm threshold might be misleading. The adjustments were applied in three ways: exactly as observed at 5 mm (sensitivity decreased by 2%, specificity increased by 10%); half the change (–1% and +5%); and the worst combination of these (–2% and +5%). The adjusted likelihood ratio negatives ranged from 0.05 to 0.07 compared with 0.04 for the unadjusted likelihood ratio.

Authors' affiliations

R Foy, Department of Reproductive and Developmental Sciences, University of Edinburgh, Edinburgh EH3 9ER, UK

P Warner, Public Health Sciences, Department of Community Health Sciences, University of Edinburgh, Edinburgh EH8 9AG, UK

REFERENCES

1 **Field MJ**, Lohr KN, eds. *Clinical practice guidelines: directions for a new program*. Washington, DC: National Academy Press, 1990.
2 **Sackett DL**, Richardson WS. Rosenberg W, *et al. Evidence-based medicine: how to practice and teach EBM*. London: Churchill Livingstone, 1997.
3 **Greenhalgh T**. Papers that report diagnostic or screening tests. *BMJ* 1997;**315**:540–3.
4 **Halkin A**, Reichman J, Schwaber M, *et al.* Likelihood ratios: getting diagnostic testing into perspective. *Q J Med* 1998;**91**:247–58.
5 **Jaeschke R**, Guyatt GH, Sackett DL, for the Evidence-based Medicine Working Group. Users' guide to the medical literature. II. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *JAMA* 1994;**271**:703–7.
6 **Deeks JJ**, Morris JM. Evaluating diagnostic tests. *Bailliere's Clin Obstet Gynaecol* 1996;**10**:613–30.
7 **Chien PFW**, Khan KS. Evaluation of a diagnostic test. II: Assessment of validity. *Br J Obstet Gynaecol* 2001;**108**:568–72.
8 **Scottish Intercollegiate Guidelines Network (SIGN)**. *A guideline developers' handbook*. SIGN Publication 50. Edinburgh: SIGN, 2001.
9 **Smith-Bindman R**, Kerlikowske K, Feldstein VA, *et al.* Endovaginal ultrasound to exclude endometrial cancer and other endometrial abnormalities. *JAMA* 2000;**280**:1510–7.
10 **Gupta JK**, Chien PFW, Voit D, *et al.* Ultrasonography endometrial thickness for diagnosing endometrial pathology in women with postmenopausal bleeding: a meta-analysis. *Acta Obstet Gynaecol Scand* 2003 (in press).
11 **Steurer J**, Fischer JE, Bachmann LM, *et al.* Communicating accuracy of tests to general practitioners: a controlled study. *BMJ* 2002;**324**:824–6.
12 **Johnson MR**, Good CD, Penny WD, *et al.* Playing the odds in clinical decision making: lessons from berry aneuysms undetected by magnetic resonance angiography. *BMJ* 2001;**322**:1347–9.
13 **Lijmer JG**, Willem MB, Heisterkamp S, *et al.* Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;**282**:1061–6.
14 **Lidor A**, Ismajovich B, Confino E, *et al.* Histopathological findings in 226 women with post-menopausal uterine bleeding. *Acta Obstet Gynaecol Scand* 1986;**65**:41–3.
15 **Gredmark T**, Kvint S, Havel G, *et al.* Histopathological findings in women with postmenopausal uterine bleeding. *Br J Obstet Gynaecol* 1995;**102**:133–6.
16 **Symonds I**. Ultrasound, hysteroscopy and endometrial biopsy in the investigation of endometrial cancer. *Bailliere's Clin Obstet Gynaecol* 2001;**15**:381–91.
17 **Gambrell RD**, Massey FM, Castaneda TA, *et al.* Use of the progestogen challenge test to reduce the risk of endometrial cancer. *Obstet Gynecol* 1980;**55**:732–8.
18 **Persson I**, Adami HO, Bergkvist L, *et al.* Risk of endometrial cancer after treatment with oestrogens alone or in conjunction with progestogens: results of a prospective study. *BMJ* 1989;**298**:147–51.
19 **Sackett DL**, Haynes RB. The architecture of diagnostic research. *BMJ* 2002;**324**:539–41.
20 **Summerton N**. Diagnosis and general practice. *Br J Gen Pract* 2000;**50**:995–1000.
21 **Glasziou PP**, Irwig LM. An evidence based approach to individualising treatment. *BMJ* 1995;**311**:1356–9.
22 **Jackson R**. Guidelines on preventing cardiovascular disease in clinical practice. *BMJ* 2000;**320**:659–61.
23 **Cabana MD**, Rand CS, Powe NR, *et al.* Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999;**282**:1458–65.
24 **Department of Health**. *Referral guidelines for suspected cancer.* London: Department of Health, 2001.
25 **Ayton P, Wright G, Rowe G.** *Medical decision-making*. In: Baum A, Newman S, Weinman J, *et al*, eds. *Cambridge handbook of psychology, health and medicine.* Cambridge: Cambridge University Press, 1997: 294–7.
26 **Shekelle PG**, Kravitz RL, Beart J, *et al.* Are nonspecific practice guidelines potentially harmful? A randomized comparison of the effect of nonspecific versus specific guidelines on physician decision making. *Health Serv Res* 2000;**34**:1429–48.
27 **Poses RM**, Cebul RD, Wigton RS. You can lead a horse to water – improving physicians' knowledge of probabilities may not affect their decisions. *Med Decis Making* 1995;**15**:65–75.
28 **Bero LA**, Grilli R, Grimshaw J, *et al.* Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. *BMJ* 1998;**317**:465–8.
29 **Cluzeau FA**, Littlejohns P, Grimshaw JM, *et al.* Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Qual Health Care* 1999;**11**:21–8.
30 **Scottish Intercollegiate Guidelines Network.** *Investigation of post-menopausal bleeding.* Publication 61. Edinburgh: Royal College of Physicians, 2002.