

EVIDENCE-BASED REPRODUCTIVE HEALTH

Evidence-based reproductive health care: using evidence about diagnostic tests

Mike Crilly, MRCGP, MFPHM, *Senior Lecturer in Clinical Epidemiology and Public Health Medicine, Department of Public Health, Aberdeen University Medical School, Aberdeen, UK*; **Robbie Foy**, MRCGP, MFPHM, *Senior Lecturer in Primary Care, Centre for Health Services Research, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne, UK*; **Pamela Warner**, BSc, PgDipStats, *Senior Lecturer in Medical Statistics, Centre for Public Health and Primary Care Research, University of Edinburgh, Edinburgh, UK*

Correspondence: Dr M Crilly, Department of Public Health, Aberdeen University Medical School, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK. E-mail: mike.crilly@abdn.ac.uk

(Accepted 31 July 2003)

Journal of Family Planning and Reproductive Health Care 2003; **29**(4): 245–248

Clinical scenario

At lunch a clinical colleague asks for your opinion concerning the care that her 55-year-old mother-in-law has recently received from a general practitioner (GP) for a recent single episode of postmenopausal bleeding (PMB). The GP who told her that the pelvic examination was normal and arranged for her to have a transvaginal ultrasound scan at the local hospital to exclude 'anything serious'. The GP reassured her that the ultrasound scan was 'normal' and that no further action was required. She has never taken any regular medication such as hormone replacement therapy (HRT) and her last normal menstrual period was some 10–15 years previously.

Your colleague is somewhat concerned. She was always taught that PMB was an absolute indication for dilatation and curettage (D&C) to exclude endometrial cancer. She doesn't want to undermine the GP (or alarm her mother-in-law unnecessarily), but on the other hand she's worried that an endometrial cancer may have been missed.

Evidence-based reproductive health care

In the first article of this series we described evidence-based medicine (EBM) as a five-step process of: (1) asking answerable questions, (2) acquiring the best available evidence, (3) appraising the evidence, (4) applying evidence appropriately to patients and (5) assessing how consistently we perform the previous four steps.^{1,2} The previous article in this series looked at the EBM approach to therapy,³ but the same approach can be applied to diagnosis.^{2,4,5}

Asking a diagnosis question

PMB is a common presenting symptom of endometrial cancer and usually indicates a need for further investigation. D&C has traditionally been used to obtain a histological diagnosis, but in recent years less invasive diagnostic investigations have been developed such as ultrasound measurement of endometrial thickness. The clinical predicament that your colleague's mother-in-law raises can be summarised into the same four-part 'PICO' format (Population Intervention Comparison Outcome) that is the basis of applied clinical research. Whilst the diagnostic 'intervention' remains the same, the actual comparison to be made is between one test result as opposed to another ('positive' or 'negative' ultrasound result). The outcome of interest is the histopathological confirmation of the presence (or absence) of endometrial cancer using an established 'gold-standard' (such as D&C).

- P: In a 55-year-old woman (not on HRT) with a single episode of PMB.
- I & C: Can a negative (or positive) test result from transvaginal ultrasound?
- O: Rule-out (or rule-in) the presence of endometrial cancer.

For questions of diagnosis the best clinical study design is 'cross-sectional'. That is a study where all patients first undergo the diagnostic test under evaluation followed soon afterwards by a diagnostic 'gold-standard'. Whilst this approach is commonly used for diagnostic tests, the same approach can also be applied to any of the key clinical features that physicians routinely elicit from the clinical history and physical examination. For example, the ability of the absence of adnexal tenderness on pelvic examination to rule out pelvic inflammatory disease.⁶

Acquiring the best available evidence

PubMed (MEDLINE) Clinical Queries is currently the best place to look for evidence about issues of diagnosis (www.ncbi.nlm.nih.gov/entrez/query/static/clinical.html). In PubMed Clinical Queries (see Figure 1) the built-in search 'filters' are set up to retrieve the most relevant studies relating to diagnosis. A free text search ('endometrial cancer AND ultrasound') limited to 'systematic reviews' retrieved a total of 11 articles (as of 1 August 2003). Three were systematic reviews, and all were recently published and available electronically.^{7–9}

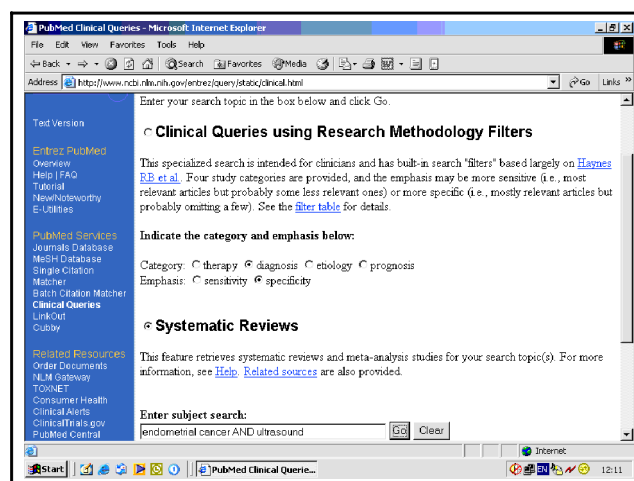


Figure 1 Home page from PubMed (MEDLINE) Clinical Queries (www.ncbi.nlm.nih.gov/entrez/query/static/clinical.html)

Searching for evidence can be like waiting for a bus. Sometimes lots of good-quality information turns up right away whilst on other occasions almost nothing can be found despite intensive searching. In this instance we are spoiled for choice with three extensive systematic reviews. Browsing the PubMed abstracts suggests that the review by Gupta and colleagues may be the best one to go for since they searched two databases (MEDLINE and EMBASE) and appraised the quality of the studies they found.⁷ The other two reviews only searched MEDLINE (and consequently may have missed some important studies).^{8,9} Neither abstract mentions appraising the quality of the studies found and one of the reviews also excludes non-English studies.⁹

Systematic reviews (where investigators trawl the literature to find all the available evidence) are traditionally associated with questions of therapy and randomised clinical trials (RCTs). It may seem surprising that several systematic reviews are available for a diagnostic topic, but such reviews of diagnostic approaches are becoming increasingly common. They can also include a pooling together of the results from several studies ('meta-analysis'). The systematic review by Gupta and colleagues⁷ identified 56 studies evaluating ultrasound in the diagnosis of endometrial cancer in women with PMB, although only five studies were of the highest quality (four of these five relating to an endometrial thickness ≤ 5 mm).⁷

Appraising evidence about diagnosis

An Athens-based hospital study¹⁰ was the largest (250 women with PMB) of the four studies with the highest quality rating in the Gupta review.⁷ The four key elements for assessing the design of a diagnostic test are shown in Box 1).^{2,4,5} Each is considered below in relation to the 'Athens study'.¹⁰

Box 1: Validity of diagnostic information

1. Independent blind comparison with diagnostic 'gold-standard'.
2. Evaluated in an appropriate clinical spectrum of patients.
3. 'Gold-standard' applied regardless of the diagnostic test result.
4. Validated in a second independent group of patients.

Validity of the 'Athens study' design¹⁰

1. The diagnostic 'gold-standard' was the histological confirmation of endometrial carcinoma by a pathologist blind to the ultrasound result (although it doesn't appear that the surgeons performing the D&C were blinded).
2. Ultrasound scanning was evaluated in women with PMB (not on HRT) attending an Athens hospital for D&C. No information is provided about the women's age or duration of amenorrhoea, nor any details about how they were recruited (so that we don't know if they are a highly selected group).
3. The reference 'gold standard' (D&C) was applied to all 250 women regardless of the ultrasound result.
4. Although ultrasonography was not evaluated in a second independent group of women in the 'Athens study' the systematic review included a further 55 studies.⁷

Applying information about a diagnostic test

The 'Athens study' evaluated the performance of transvaginal ultrasound in detecting endometrial cancer in 250 women admitted to hospital for a D&C for PMB. Endometrial thickness was measured using transvaginal ultrasound (of both endometrial layers, i.e. 'double layer') performed by a single ultrasonographer the day before D&C under general anaesthetic.

Table 1 Contingency table (2 x 2 table) of the ability of transvaginal ultrasound scanning to correctly identify endometrial cancer in 250 women with postmenopausal bleeding

| TEST: Endometrial thickness (on ultrasound scan) | DISEASE: Endometrial cancer (at D&C) | |
|---|---|-------------|
| | Present | Absent |
| Positive (>5 mm) | 24 (a) | 75 (b) |
| Negative (≤ 5 mm) | 0.1 (c) | 151 (d) |
| | 24 (a + c) | 226 (b + d) |

Prevalence = $a + c / (a + b + c + d) = 10\%$

LR+ = $[a / (a + c)] \div [b / (b + d)] = 3.0$

LR- = $[c / (a + c)] \div [d / (b + d)] = 0.01$

D&C, dilatation and curettage; LR, likelihood ratio.

The 2 x 2 table

Some 151 women had a 'negative' scan (endometrial thickness ≤ 5 mm) whilst 99 women had a 'positive' scan (>5 mm). A histopathologist confirmed that 24 women had endometrial cancer, a prevalence of 10%. The results from the 'Athens study' can be placed into a simple 2 x 2 table as shown in Table 1.

The two shaded boxes in Table 1 indicate where the ultrasound scan gave the correct result. A perfect test would place all women with PMB into one of these two shaded boxes, but this is not usually the case in real clinical practice. Most diagnostic tests tend to be better at placing patients into one box rather than another. Consequently, selecting a diagnostic test is often a trade-off between 'ruling-in' or 'ruling-out' the target disorder.

Box 'a' is where a positive ultrasound scan indicated that women with PMB had cancer when they did in fact have cancer. Box 'd' is where a negative scan indicated that women didn't have cancer when they didn't in fact have cancer. Overall the test got it right for 70% ($24 + 151/250$) of the women with PMB. Box 'b' is where a positive ultrasound scan incorrectly indicated that a woman with PMB had cancer when in fact she didn't have cancer. Box 'c' is where the scan wrongly indicated that a woman didn't have cancer when in fact she did have cancer. Whilst the ultrasound test gave a 'positive' result for 75 women who didn't actually have endometrial carcinoma ('false-positives'; box 'b'), the ultrasound test didn't produce any 'false-negative' results (box 'c'). Overall the test got it wrong for 30% ($75 + 0/250$) of the women with PMB (to make the calculations easier zero has been replaced with 0.1 in box 'c').

Likelihood ratios

The usefulness of a diagnostic test depends upon likelihood ratios (LRs).^{2,4,5} The concept of LRs is much more helpful than traditional concepts such as 'sensitivity' and 'specificity'. In fact that is the only reference we're going to make to sensitivity and specificity. LRs have the advantage that they can be directly applied to individual patients. They indicate by how much a given test result increases (or decreases) a patient's chances of actually having the targeted disorder. LRs can be calculated for both positive (LR+) and negative (LR-) test results.

The LR for a positive test result (LR+) indicates by how much a patient with a positive test further increases their risk of having the target disorder. The LR ratio for a negative test result (LR-) indicates by how much a patient with a negative test further decreases their risk of having the target disorder. A LR of 1.0 means the test result provides no useful information about the presence or absence of disease. A LR >1.0 increases the probability of disease (the larger the LR the larger the increase) whereas a LR <1.0 reduces the probability of disease (the smaller

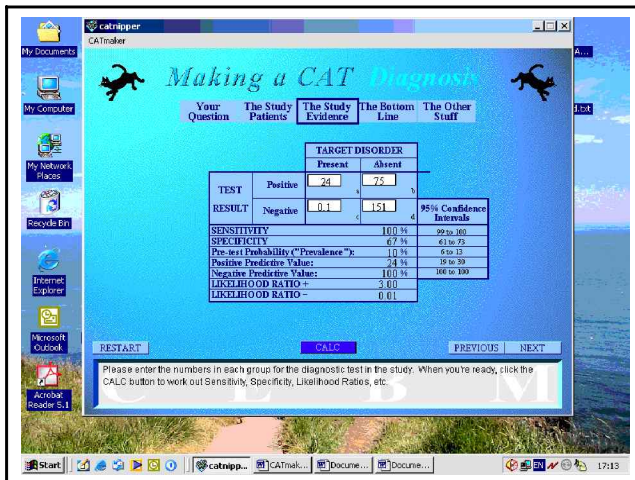


Figure 2 CATmakers 2 x 2 table (www.cebm.net)

the LR the bigger the decrease). An ideal diagnostic test will have a high LR for a positive test result (LR+) and a low LR for a negative test result (LR-).

In our example of ultrasound and endometrial cancer, the LR is the ratio of the likelihood of a test result (either positive or negative) in women with endometrial cancer, divided by the likelihood of the same test result in those without cancer. Table 1 shows the algebra for calculating LRs for positive and negative test results. But don't be too concerned if your algebra is rusty as several available packages (such as CATmaker at www.cebm.net) are available to work out the LRs (see Figure 2)

The '10-fold difference rule'

How big (for a LR+) or small (for a LR-) does a LR have to be in order to be diagnostically useful? A helpful rule of thumb is the '10-fold' one. Tests that increase (or decrease) the likelihood of a disease by a factor of 10 or more will be excellent at ruling-in (or ruling-out) the presence of disease. A LR+ of 10.0 or more will rule-in disease, whereas a LR- of 0.10 or less will rule-out disease. LRs that make somewhere between a 5- and 10-fold difference to the likelihood of the targeted disorder only have an intermediate ability to rule disease in and out.

Determining the risk of cancer

It would be nice if we could simply multiply the risk of our patient having cancer before we do the test ('pretest probability') by the appropriate 'LR' and obtain the risk of our patient having cancer after a 'positive' or a 'negative'

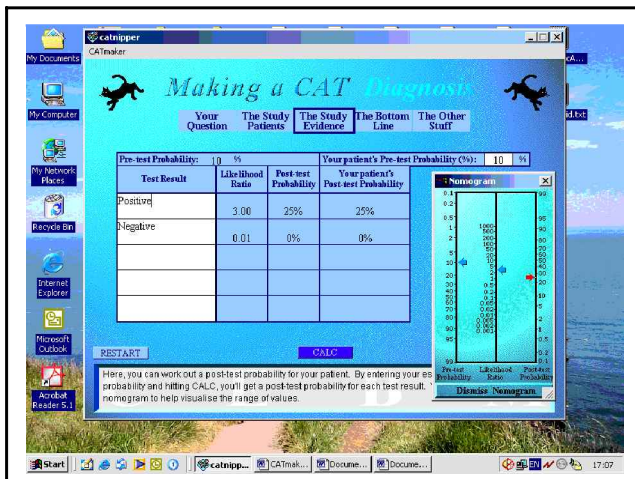


Figure 3 CATmakers posttest probability calculator and nomogram (www.cebm.net)

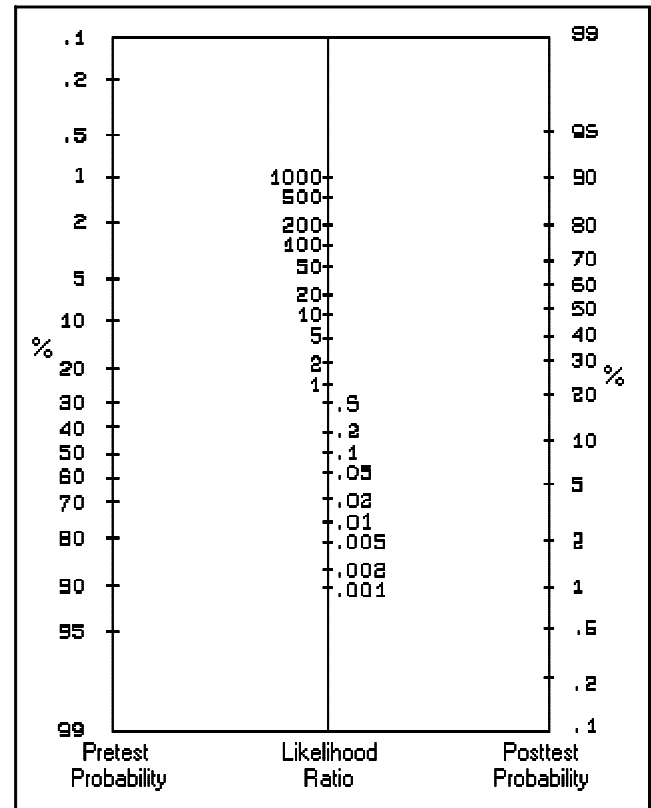


Figure 4 Likelihood ratio nomogram. Figure adapted from Fagan¹¹

test result ('post-test probability'). Unfortunately LRs are derived from something called 'Bayes theorem' which is based on 'odds' rather than 'probabilities'. To avoid getting tied up in the technicalities of using 'odds' we can adopt a 'high-tech' or a 'low-tech' approach.

The 'high-tech' way is easy. We simply get a machine (either a PC or handheld device) to run one of the readily available packages to do the maths (such as CATmaker, see Figure 3). Alternatively we can use a printed 'nomogram' (as shown in Figure 4) just like an old-fashioned slide-rule. The 'low-tech' way is also easy. Line up two points, join the dots by drawing a straight line across to the vertical line on the right, and read off the 'post-test probability'. If you want to combine both approaches you can play with the 'electronic nomogram' in CATmaker (see Figure 3).

Applying the 'Athens study' data

Assuming that the pretest probability ('prevalence') of endometrial cancer is 10% in women with PMB (as in the 'Athens study'), then the probability of having endometrial cancer rises to 25% after a positive test (see nomogram and table in Figure 3), but falls to 0.1% after a negative test result (see the table in Figure 3). You might want to double-check that we got this right by using the 'printed nomogram' shown in Figure 4.

From a diagnostic perspective it is the negative test result (endometrium thickness ≤ 5 mm) that is clinically the more informative. If we reassure 1000 women with a negative scan that they don't have endometrial cancer, then 999 will have been correctly advised, although we would miss the diagnosis of cancer in one additional woman. Alternatively, if we tell 100 women with a positive scan that they have endometrial cancer then we will be correct for 25 women (but cause anxiety to 75/100 who don't have cancer). A negative test result is better at 'ruling-out' endometrial cancer than a positive test result is at 'ruling-in' endometrial cancer.

Table 2 Likelihood ratios^a for different thicknesses of the endometrium on ultrasound scanning for the prediction of endometrial cancer in women presenting with postmenopausal bleeding

| Endometrial thickness on ultrasound (mm) | Positive LR _s (LR ₊) | Negative LR _s (LR ₋) | Studies combined (n) |
|--|---|---|----------------------|
| ≤6 | 2.5 | 0.20 | 2 |
| ≤5 | 2.2 | 0.15 | 21 |
| ≤4 | 2.0 | 0.08 | 9 |
| ≤3 | 2.1 | 0.04 | 2 |

^aFrom systematic review by Gupta and colleagues (2002).⁷
LR, likelihood ratio.

Different levels of endometrial thickness

Ultrasound measurement of endometrial thickness is made in millimetres, but the results are typically reported as either 'positive' or 'negative', depending on whether the thickness is above or below a certain threshold ('cut-off point'). Table 2 shows the 'pooled' LR_s for several different cut-off levels of endometrial thickness on from the systematic review by Gupta.⁷

Positive likelihood ratios (LR₊)

We can see from Table 2 that that the cut-off point selected for endometrial thickness on ultrasound makes little difference to the ability of a 'positive test' result (LR₊) to 'rule-in' cancer. The LR₊ changes little as the thickness increases, presumably because other conditions (such as endometrial hyperplasia) also show up as endometrial thickening on ultrasound scanning.

Negative likelihood ratios (LR₋)

The ability of a 'negative test' result to 'rule-out' cancer improves as we select ever-thinner levels of endometrial thickness on ultrasound. As the cut-point for endometrial thickness is reduced, the LR₋ moves further and further away from unity (i.e. a LR of 1.0). At an endometrial thickness of ≤3 mm the LR₋ is 0.04. If we apply this to a women with a pretest probability of cancer of say 14% (the average from the Gupta review)⁷ a 'negative' test result decreases her risk of cancer to 0.7% (7 in 1000). Hence we would only (falsely) reassure 7/1000 women with a negative scan that they didn't have endometrial cancer (when in fact they did), compared to correctly reassuring 993 women with PMB that they didn't have cancer (which they don't) – and potentially avoiding D&C to confirm this histologically.

This is the reason why the current SIGN guideline for the investigation of PMB (www.sign.ac.uk/) suggests that an endometrial thickness of '3 mm or less' on ultrasound can be used to exclude endometrial cancer in women with PMB,¹² although we should cautiously point out that only two studies used the ≤3 mm cut-off point (see Table 2).

A minor point: readers may have noticed (but don't worry if you didn't) that the LR_s for ≤5 mm thickness shown in Table 2 are different to the previous LR_s we calculated (from the 'Athens study') in Table 1. This is simply because Table 2 relates to LR_s derived from pooling the 'Athens study' result with those from the other 20 studies.

Resolution of the scenario

A key element in the case of your colleague's 'postmenopausal mother-in-law' is the cut-off level that the

ultrasonographer uses to distinguish between a 'normal' negative scan and an 'abnormal' positive scan. For the mother-in-law's negative scan result, a cut-off point of ≤3 mm would be of greater reassurance than one of say ≤5 mm, although the majority of the international evidence currently available on this topic is for the ≤5 mm level (21 studies) rather than the 3 mm level (two studies).

The review by Gupta⁷ reported a 14% prevalence of endometrial cancer in women presenting with PMB, although a prospective study has reported a lower prevalence of 8%.¹³ The 'Athens study' reported 10%.¹⁰

The pooled results from the four highest quality studies (using the ≤5 mm cut-off)⁷ indicates that if your colleague's mother-in-law's pretest risk of endometrial cancer is a 'high' 14%, then her risk of still having cancer after a 'normal' ultrasound scan is reduced to 2.3%. Combining together only the four 'high-quality' studies may provide a more reliable estimate of the LR_s than combining 21 studies of variable quality. A posttest probability of cancer of 2.3% means that one in 40 women with a negative scan following PMB will still have an underlying endometrial cancer. Further action probably depends upon how much risk the 'post-menopausal mother-in-law' (and her relatives) are ready to tolerate.

Statements on funding and competing interests

Funding. None identified.

Competing interests. Mike Crilly has received payments from the 'Oxford Centre for EBM' for tutoring on their workshops.

References

- Crilly M, Foy R. Evidence-based family planning: finding answers to clinical questions. *J Fam Plann Reprod Health Care* 2003; **29**(2): 48–52.
- Sackett D, Straus S, Richardson W, et al. *Evidence-based medicine. How to practise and teach EBM* (2nd edn). Edinburgh: Churchill-Livingstone, 2000.
- Foy R, Crilly M, Brechin S. Evidence-based reproductive health: testing times for treatments. *J Fam Plann Reprod Health Care* 2003; **29**(3): 165–168.
- Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? The Evidence-Based Medicine Working Group. *JAMA* 1994; **271**: 389–391.
- Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994; **271**: 703–707.
- Peipert JF, Ness RB, Blume J, et al. Clinical predictors of endometritis in women with symptoms and signs of pelvic inflammatory disease. *Am J Obstet Gynecol* 2001; **184**: 856–863.
- Gupta JK, Chien PF, Voit D, et al. Ultrasonographic endometrial thickness for diagnosing endometrial pathology in women with postmenopausal bleeding: a meta-analysis. *Acta Obstet Gynecol Scand* 2002; **81**: 799–816.
- Smith-Bindman R, Kerlikowske K, Feldstein VA, et al. Endovaginal ultrasound to exclude endometrial cancer and other endometrial abnormalities. *JAMA* 1998; **280**: 1510–1517.
- Tabor A, Watt HC, Wald NJ. Endometrial thickness as a test for endometrial cancer in women with postmenopausal vaginal bleeding. *Obstet Gynecol* 2002; **99**: 663–670.
- Grigoriou O, Kalovidouras A, Papadimas C, et al. Transvaginal sonography of the endometrium in women with postmenopausal bleeding. *Maturitas* 1996; **23**: 9–14.
- Fagan T. Nomogram for Bayes theorem. *N Engl J Med* 1975; **293**: 257.
- Foy R, Warner P. About time: diagnostic guidelines that help clinicians. *Qual Saf Health Care* 2003; **12**: 205–209.
- Gredmark T, Kvint S, Havel G, et al. Histopathological findings in women with postmenopausal bleeding. *Br J Obstet Gynaecol* 1995; **102**: 133–136.

Visit the Faculty Website at: www.ffprhc.org.uk