

## Diagnosis: shifting the ROC curve

### INTRODUCTION

In 1972 Howie asked whether diagnosis was the Achilles heel of general practice.<sup>1</sup> Accurate diagnosis was then, and remains, the fundamental concept of all medical practice. Are we any good at it now?

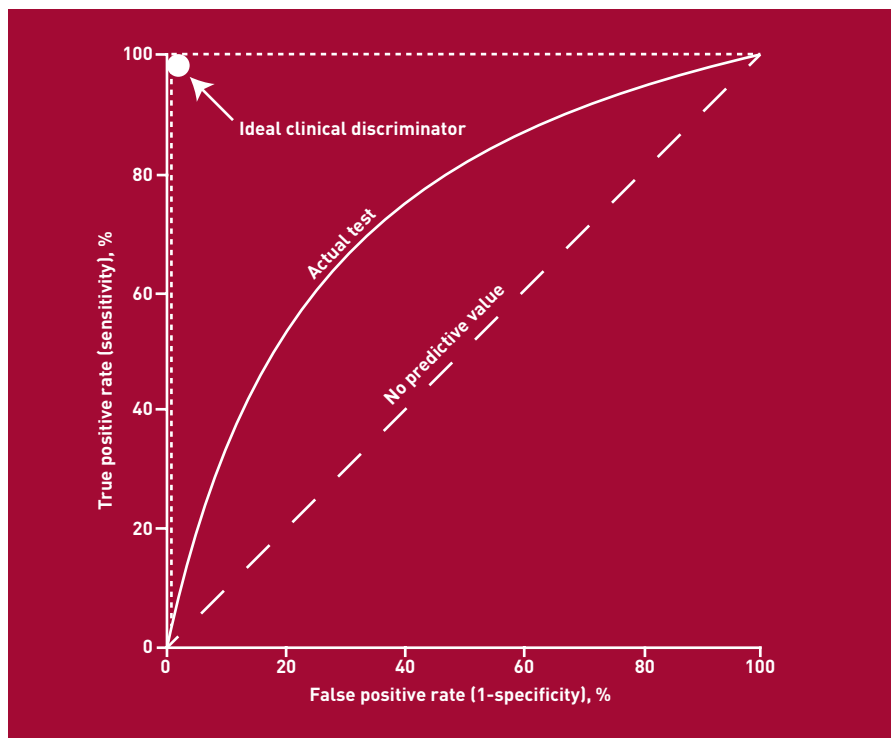
### HOW ARE WE DOING?

This issue of the *BJGP* contains several papers on diagnosis and decision-making which shed light on this important area of practice. The results reported by Meechan *et al* give food for thought.<sup>2</sup> The researchers analysed national referrals for suspected cancer under the '2-Week Wait' rule (2WW) during 2009. Out of 865 494 referrals, 11% had cancer. Of all cancers being newly treated in that year, 43% were referred under the 2WW. So far — not bad.

However, they also demonstrate wide disparity in practice, with clear evidence that some practices are good at diagnosing cancer while some are poor. This quality difference does not correlate with deprivation, region, practice size, or referral rates (which themselves vary threefold between practices<sup>3</sup>).

High-quality practices have high 'conversion rates': of those referred under the 2WW rule maybe 14% are diagnosed with cancer. However, the same practices also have high 'detection rates': of all cancers diagnosed in the practice, maybe 40% to 50% have been suspected and referred under the 2WW rule. In poorly performing practices the figures may be 4% and 17% respectively.

The scatter diagram in Figure 1 of the Meechan *et al* paper is a powerful argument against complacency. If all practices were roughly the same the scatter diagram would look like a large spherical globular cluster with the spread explained by random variation. But that is not what the data show. In the practices in the top right of the main grouping, 20% to 30% of all patients referred under 2WW were diagnosed with cancer. Those same practices were detecting 60% to 70% of all cancers via 2WW referrals. In practices in the bottom left of the Meechan *et al* paper, maybe only 5% of those referred under the 2WW had cancer, yet only 10% of all cancers were detected via 2WW rule referrals. If this was random variation it would not explain the correlation between conversion rates and detection rates.



**Figure 1.** The receiver operating characteristics (ROC) curve shows how good a clinical feature or test is in discriminating between disease and non disease. A good test achieves high levels of true positives with low levels of false positives. The better the test, the greater the area under the ROC curve.

### THE BALANCE BETWEEN DOING TOO LITTLE AND DOING TOO MUCH

Diagnosis is difficult and GPs overall probably do not do badly. But can we improve? One possibility would be to refer and investigate patients at lower absolute risk of disease, although there are two problems with this. First, it would be expensive. Second, we risk medicalising huge numbers of people at low risk of disease. If most practices are currently referring under the 2WW at absolute levels of risk of 5% or 8% the financial implications of referring at 1% levels of risk are likely to be huge. The 'symptom iceberg' is a pyramid and not a cube.<sup>4</sup> A shift in threshold from 5% risk to 1% risk is likely to generate

far more than a fivefold increase in referral and investigation. The gap between waits for routine and urgent referral, wide already, will become even wider.

Vedsted and Olesen analysed data from the EURO CARE-4 study and showed reduced cancer survival in those countries with GP 'gatekeepers'.<sup>5</sup> The implication is that GPs are barriers to early diagnosis. Others have criticised the completeness of the EURO CARE-4 dataset<sup>6</sup> (it covers 1% of the German population and 80% of the UK population). But surely we are more than 'gatekeepers'. Equally important is protecting patients against 'the cascade effects of medical technology'.<sup>7</sup> There is increasing evidence of the harmful effects of

*"Diagnosis is difficult and GPs overall probably do not do badly. But can we improve? One possibility would be to refer and investigate patients at lower absolute risk of disease, although there are two problems with this ..."*

*"The rather startling implication of the evidence presented by Meechan et al is that, for once, there is not a straight trade off between sensitivity and specificity..."*

over investigation and over diagnosis.<sup>8</sup> If I am investigated at a 1% absolute level of risk I will need 69 (0.9969 = 0.5) colonoscopies, cystoscopies, prostate biopsies, head magnetic resonance imaging (MRI with a 7% risk of incidental abnormalities), computed-tomography calcium scores, and more to have a 50% chance of being diagnosed with something significant and possibly benefiting. (The figure for 5% risk generates the need for a more acceptable 13 referrals).

A decade ago Haslam described how GPs not only see huge numbers of patients in the NHS, but also absorb huge amounts of risk.<sup>9</sup> In so doing they allow a publicly-funded health service to function. He made the analogy with the heat sink in a computer: it does not get the headlines but is nonetheless essential. We need to be wary of dismantling that 'risk sink' with over inclusive referral and treatment guidelines. Sceptics seem to be underrepresented on expert committees. As generalists, we need to be cautious about both advice from tertiary specialists on the lecture circuit, and of patient pressure groups with anecdotal experience of disease. In general practice we tread a fine line between exercising too little and too much caution.

### IMPROVING ACCURACY

So how can we improve diagnostic precision and yet avoid medicalising the world? One answer would be to increase the area under the ROC (receiver operating characteristic) curve (Figure 1).

The rather startling implication of the evidence presented by Meechan *et al* is that, for once, there is not a straight trade off between sensitivity and specificity. Both could be improved if all practices performed as well as the high performing practices. Most of us would want to know where our diagnoses fit on that scatter diagram (Meechan *et al*). Are we bottom left or top right? It surely represents a more meaningful measure of quality of practice than many of the current Quality and Outcomes Framework targets. The results of Meechan *et al* suggest that we could do better with the tools that we already have — with improved education and practice.

A second possibility would be to improve the evidence base of diagnosis and for the various diagnostic methods we use. Over investigation potentially leads to colossal financial costs. It is arguably one of the largest areas for potential cost savings and better use of resources. Yet it is surprising that so few resources are directed to improving use of diagnostic facilities. Much referral and management guidance gets inexorably more over-inclusive with time.

Research that identifies more discriminatory clinical features or tests would improve the precision of referral. In this issue of the *BJGP*, Shephard *et al* analyse clinical features that may improve identification of bladder cancer.<sup>10</sup> But realistically, the new clinical features introduced perform rather poorly when used individually. Visible haematuria remains the key. Overall, it is likely that, after generations of clinical observation, few new clinical features will be discovered that, individually, change the game.

However, clinical features that are poor discriminators in isolation may become more significant when incorporated into a clinical prediction rule. A tool that identified the patient as being at 0.2% risk of pulmonary embolism rather than 20% would be very useful. Although such risk scores are cumbersome to use, they may be preferable to investigating legions of people at low risk. However, the evidence that such tools are both used and are successful is mostly lacking, particularly in primary care.<sup>11</sup>

Lastly, there is good evidence that if GPs are given access to good diagnostic investigations (such as MRI) they use them carefully and appropriately.<sup>12</sup>

My own belief is that the 2WW guidance is 12 years old and should be restructured into something less proscriptive, less cumbersome, more educational, and more flexible. The binary decision for GPs between a 2-week wait and a 4-month wait is unacceptable. But we cannot avoid the powerful evidence that, at least in the field of cancer diagnosis, there are wide and unexplained variations in practice. Before we plunge more resources into investigating people at low risk we need to

### ADDRESS FOR CORRESPONDENCE

Kevin Barraclough

Painswick Surgery, Hoyland House, Painswick, Stroud, GL6 6RD, UK.

E-mail: k.barraclough@btinternet.com

find out why there is such variation and how we can shift that ROC curve northwest.

Kevin Barraclough,

GP, Painswick Surgery, Painswick, Stroud.

### Provenance

Commissioned; not externally peer-reviewed.

### Acknowledgements

William Hamilton argued constructively with the author's views.

### Competing interests

The author has declared no competing interests.

DOI: 10.3399/bjgp12X653796

### REFERENCES

1. Howie JGR. Diagnosis — the Achilles heel? *J R Coll Gen Pract* 1972; **22(118)**: 310–315.
2. Meechan D, Gildea C, Hollingworth L, *et al*. Variation in use of the 2-week referral pathway for suspected cancer: a cross sectional analysis. *Br J Gen Pract* 2012; DOI: 10.3399/bjgp12X654551.
3. National Cancer Intelligence Network. *General practice profiles for cancer*. London: NCIN. [http://www.ncin.org.uk/cancer\\_information\\_tools/profiles/gp\\_profiles.aspx](http://www.ncin.org.uk/cancer_information_tools/profiles/gp_profiles.aspx) [accessed 8 Aug 2012].
4. Crosland A, Jones R. Rectal bleeding: prevalence and consultation behaviour. *BMJ* 1995; **311(7003)**: 486–488.
5. Vedsted P, Olesen F. Are the serious problems in cancer survival partly rooted in gatekeeper principles? *Br J Gen Pract* 2011; DOI: 10.3399/bjgp11X588484.
6. Souhami R. Are UK cancer cure rates worse than in most other European countries? *Br J Gen Pract* 2010; **60(571)**: 81–82.
7. Deyo RA. Cascade effects of medical technology. *Annu Rev Public Health* 2002; **23**: 23–44.
8. Moynihan R, Doust J, Henry D. Preventing overdiagnosis: how to stop harming the healthy. *BMJ* 2012; **344**: e3502.
9. Haslam D. 'Schools and hospitals' for 'education and health'. *BMJ* 2003; **326(7383)**: 234–235.
10. Shephard EA, Stapley S, Neal RD, *et al*. Clinical features of bladder cancer in primary care. *Br J Gen Pract* 2012; DOI: 10.3399/bjgp12X654560.
11. Reilly BM, Evans AT. Translating clinical research into practice: impact of using prediction rules to make decisions. *Ann Intern Med* 2006; **44(3)**: 201–209.
12. Gough-Palmer AL, Burnett C, Gedroyc WM. Open access to MRI for general practitioners: 12 years' experience at one institution—a retrospective analysis. *Br J Radiol* 2009; **82(980)**: 687–690.