Evidence-Based Medicine

EBM NOTEBOOK

Evidence-based pathology

Evidence-based medicine is currently a highly fashionable and visible concept. Does this idea have implications for cellular pathology? When I suggested elsewhere that it did (1), the editors of *Evidence-Based Medicine* invited me to repeat my arguments for their readers.

Apart from the obvious impact of evidence-based medicine on teaching, a stimulating aspect of the concept is that it challenges medical practitioners to examine the scientific basis of their practice. One analogy for cellular pathologists is to examine how we diagnose and prognosticate. This can be encapsulated in 2 questions: 1) How good are we at reproducibly recognising and quantifying (if appropriate) specific morphological features?; 2) How relevant are these features to diagnosis, management, or outcome? Of the 2 questions, the first must be addressed and made as secure as possible initially: If a feature cannot be reliably and reproducibly identified and quantified by a broad group of pathologists, any conclusion that it is clinically important is almost meaningless and, at worst, misleading. [*Editors' note:* It was for this reason that we, the editors, invited Dr. Fleming to repeat his evidence and conclusions here.] Answering the question of how good pathologists are at reproducibly recognising morphological features means determining the inter-observer variation (2 pathologists independently examine the same specimen) and intra-observer variation (the same pathologist examines the same specimen again, without referring to the earlier report) in reporting a morphological feature on a biopsy or surgical specimen. As might be expected, considerable literature already exists on this, with agreement usually expressed as a kappa score (2). A score of 0 corresponds to agreement no better than chance, whereas 1 is perfect agreement. In practice, a kappa score of greater than 0.4 is taken to indicate agreement is becoming reasonable, whereas 0.6 and above is considered good agreement.

Most pathologists (but perhaps few other clinicians) are aware that the issue of observer variation in pathology reporting has been evaluated for some time (3, 4) but probably feel that the evaluations have little to offer the daily practice of cellular pathology. However, an examination of the evidence suggests that pathologists frequently fail to identify specific morphological features reproducibly. Indeed, apart from grouping lesions into broad categories (e.g., benign versus malignant), poor agreement is common (Table). Moreover, most of these studies involved very small numbers of pathologists, often with specialist knowledge and interest in the topic; often on restricted or selected samples; and, of course, often devoting unusual attention to the material examined. This implies that the observer variation in the general community is likely to be significantly worse than that documented in the literature.

Does the existence of observer variation matter to pathologists and the clinicians who rely on them? Most pathologists probably subconsciously (at least) already know something of the above. However, they usually ignore it, partly because they feel nothing can be done anyway (the problem is inherent in any subjective specialty), partly because the specialists do not realise how bad the variation is and partly because they feel that individually they are internally reproducible (in general, intra-observer variation is less than inter-observer) and "their clinicians" know what they are talking about. However, the existence of observer variation does matter. Management decisions are made on the basis of the pathology diagnosis, often on features that are poorly reproducible. For example, therapeutic decisions for treatment of breast cancer can be influenced by grading of tumours (kappa score 0.18 to 0.36). Steroid or interferon therapy in liver disease can be instituted on the basis of

Table

Reference	Organ feature	Agreement	Kappa score
Theodossi et al. (5)	Liver: piecemeal necrosis	47%	0.223
Thomas et al. (6)	Rectal cancer: grading	50% to 69%	0.11 to 0.5
Holman et al. (7)	Lymph node: Hodgkins classification	56%	0.44
Stenkvist et al. (8)	Breast cancer classification	73%	0.46
Creagh et al. (9)	Cervical intra-epithelial neoplasia	ND	0.009 to 0.519
Carter et al. (10)	Grading anal intra-epithelial neoplasm	ND	0.17 to 0.6
Demetris et al. (11)	Liver transplant acute rejection	ND	0.31 to 0.5
Colloby et al. (12)	Depth of invasion in thin malignant melanoma	Breslow, 82% Clark, 64%	0.68 0.23
Sloane et al. (13)	Breast cancer Grade Invasive subtype Atypical hyperplasia	ND	 0.26 0.00 to 0.31 0.17

ND = not done.

piecemeal necrosis (interface hepatitis) (kappa score 0.2). Prognostic information is given to patients on the basis of depth of invasion in malignant melanoma (kappa score 0.23 to 0.68). Finally, as mentioned already, accurate determination of whether a feature really is clinically relevant (question 2 above) can only be achieved and applied if the feature is reproducibly identified and quantified by pathologists.

What is to be done? I've recommended 5 steps that my colleagues and I in pathology might undertake. The first is to recognise that a problem exists, and the second is to adopt an evidence-based approach to our teaching and training. Despite the inevitable subjectivity of cellular pathology, improvements can be made. For example, recent evidence from the United Kingdom breast screening programme has shown that inter-observer agreement can rise to acceptable limits when repeated rounds of assessment are carried out, particularly concentrating on specific topics. Third, and pertinent to the readers of this journal, editors and referees should expect and require kappa scores when morphological features are the subject of an article. Fourth, the burgeoning Audit, Continuing Medical Education, and Quality Assurance Schemes ought to pay particular attention to observer variation, and readers of evidence-based medicine publications should join pathologists in fostering this development. Finally, and most important, by concentrating on those features that are reproducibly recognisable and clinically relevant and by abandoning those that are not, patient care can be made much more evidence-based, to the ultimate benefit of patients.

Kenneth A. Fleming, FRCPath
University of Oxford
Oxford, England, UK

References

1. **Fleming KA.** J Pathol. 1996;179:127-8.
2. **Altman D.** Practical Statistics for Medical Research. London: Chapman and Hall; 1991.
3. **Wartman WB.** Am J Clin Pathol. 1959; 32:468-71.
4. **Cutler SJ, Black MM, Friedell GH, Vidone RA, Goldenberg IS.** Cancer. 1966;19:75-82.
5. **Theodossi A, Skene AM, Portmann B, et al.** Gastroenterology. 1980;79:232-41.
6. **Thomas GD, Dixon MF, Smeeton NC, Williams NS.** J Clin Pathol. 1983;36:385-91.
7. **Holman CD, Matz LR, Finlay-Jones LR, et al.** Histopathology. 1983;7:399-407.
8. **Stenkvist B, Bengtsson E, Eriksson O, et al.** J Clin Pathol. 1983;36:392-8.
9. **Creagh T, Bridger JE, Kupek E, et al.** J Clin Pathol. 1995;48:59-60.
10. **Carter PS, Sheffield JP, Shepherd N, et al.** J Clin Pathol. 1994;47:1032-4.
11. **Demetris AJ, Seaberg EC, Batts KP, et al.** Hepatology. 1995;21:408-16.
12. **Colloby PS, West KP, Fletcher A.** J Pathol. 1991;163:245-50.
13. **Sloane JP, Ellman R, Anderson TJ, et al.** Eur J Cancer. 1994;30A:1414-9.