

hormonal peptides, can be considered mucosal in nature. These observations may explain the primitive origin of the MALT lymphoma in the thymus. The morphological features of the submandibular gland—that is, expansion of the lymphoid infiltrate and the presence of blast forms of CCL cells, are diagnostic of a lymphomatous process and genetic analysis demonstrated a κ chain rearrangement. The morphological and genotypic analyses are consistent with a diagnosis of B cell low grade malignant lymphoma of the salivary gland² exhibiting features of a MALT lymphoma. As the salivary gland normally lacks lymphoid tissue, the MALT lymphoma may have originated from tissue involved by MESA. However, there was no evidence of underlying MESA in our patient. The infiltrate of the skin was composed of a few clear CCL cells and many monotypic IgA/ κ plasma cells. Plasma cell differentiation was associated with amyloid deposits. The presence of CCL cells and the monoclonality of the plasma cells suggest that the cutaneous infiltrate originated via the same lymphomatous process.

It is of note that the lesions in the salivary gland and the thymus were histologically similar and also had identical clonally rearranged bands. The morphological features of MALT lymphoma in the salivary gland and the thymus have been reported previously.³ It is possible that multiple mucosal sites can be involved by MALT lymphoma or that the lymphoma can spread from the salivary gland to the thymus. This pattern of spread probably depends on the specific circulation pathways and homing characteristics of the lymphocytes present. We can put forward two hypotheses: (1) that the B lymphocytes of the thymus and salivary gland are closely related; and (2) that B lymphocytes can move between these two organs. Similarly, the lymphomatous localisation in the skin could have resulted from peripheral spread of the tumours and could be related, at the least

in part, to lymphocyte–endothelial cell recognition. Skin directly exposed to the external environment may be a site for so-called acquired B cell, skin associated lymphoid tissue. Cutaneous B cell lymphomas have numerous aspects in common with MALT lymphoma, not least being originated from lymphoid tissue.¹⁰

In conclusion, our case is unique because the patient has evidence of MALT lymphoma in the salivary gland and in the thymus showing the same genotypic pattern, with further localisation in the skin. This confirms the possibility of multiple mucosal sites involved by lymphoma or the spread of a MALT lymphoma from one mucosal site to another, including the skin.

- 1 Isaacson PG, Spencer J. Malignant lymphoma of mucosa-associated lymphoid tissue. *Histopathology* 1987;11:445–62.
- 2 Hyjek E, Smith WJ, Isaacson PG. Primary B cell lymphoma of salivary gland and its relationship to myoepithelial sialadenitis (MESA). *Hum Pathol* 1988;19:766–76.
- 3 Isaacson PG, Chan JKC, Tang C, Addis BJ. Low-grade B-cell lymphoma of mucosa-associated lymphoid tissue arising in the thymus. A thymic lymphoma mimicking myoepithelial sialadenitis. *Am J Surg Pathol* 1990;14:342–51.
- 4 Li G, Hansmann ML, Zwingers T, Lennert K. Primary lymphomas of the lung: morphological, immunohistochemical and clinical features. *Histopathology* 1990;16:519–33.
- 5 Norton AJ, Isaacson PG. Detailed phenotypic analysis of B-cell lymphoma using a panel of antibodies reactive in routinely fixed wax embedded tissue. *Am J Pathol* 1987;128:225–40.
- 6 Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning. A laboratory manual*. 2nd edn. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 1989.
- 7 Hieter PA, Max EE, Seidman JG, Marzel JV Jr, Leder P. Cloned human and mouse kappa immunoglobulin constant and J region genes conserve homology in functional segments. *Cell* 1980;22:197–207.
- 8 Hieter PA, Hollis GF, Horsmeyer SJ, Waldmann TA, Leder P. Clustered arrangement of immunoglobulin lambda constant region genes in man. *Nature* 1981;294:536–40.
- 9 Takagi N, Nakamura S, Yamamoto K, Kunishima K, Takagi I, Suyama M, et al. Malignant lymphoma of mucosa-associated lymphoid tissue arising in the thymus of a patient with Sjogren's syndrome. *Cancer* 1992;69:1347–55.
- 10 Gianotti B, Santucci M. Skin-associated lymphoid tissue (SALT)-related B-cell lymphoma (primary cutaneous B-cell lymphoma). *Arch Dermatol* 1993;129:353–5.

Kappa statistics as indicators of quality assurance in histopathology and cytopathology

S S Cross

Abstract

Kappa statistics are widely-used to assess performance in quality assurance schemes. Low values, however, are difficult to interpret, especially when confidence intervals have not been calculated. A model of a dichotomous decision in pathology (benignancy or malignancy in fine needle aspirates of the breast) was used to calculate kappa statistics (with confidence limits) for increasing false

positive rates. It was found that the level at which the upper 95% confidence interval for the kappa statistic fell below 1 was an insensitive method of detecting unsatisfactory performance as at that level the false positive rate was unacceptably high (>1%) for all populations of specimens less than 800 in number. Either large populations of samples are required in quality assurance schemes which use kappa statistics (which may well be impractical) or

Department of Pathology, University of Sheffield Medical School, Beech Hill Road, Sheffield S10 2UL

email: s.s.cross@sheffield.ac.uk

Accepted for publication 21 March 1996

other methods of assessing performance, possibly with weighted outcomes, are required.

(*J Clin Pathol* 1996;49:597-599)

Keywords: quality assurance schemes, kappa statistics, cytopathology.

Internal and external quality assurance schemes, proficiency testing and related exercises are seen as useful ways of promoting uniform high quality reporting of specimens in histopathology and cytopathology. Assessment of results in these schemes differ from other pathological specialties, such as clinical chemistry and haematology, in that the results are discrete diagnostic categories (for example, severe dyskaryosis in a cervical smear) rather than variable parameters (for example, haemoglobin) and for this reason kappa statistics are often used as indicators of performance.¹⁻³ Kappa statistics measure levels of agreement between two observers and make allowance for the degree of agreement that would occur by chance alone. In histopathology and cytopathology there has been discussion about the most appropriate reference diagnosis to be used in the calculation of kappa statistics (for example, the consensus of participants in the scheme, an expert diagnosis, a diagnosis verified by non-histopathological criteria, etc.) but another important factor to be considered is the significance of any calculated kappa statistic in its context. In almost all quality assurance schemes which use kappa statistics the statistic is given as a single figure for each diagnosis, without confidence limits, and it is difficult to assess the significance of low values.⁴ In the present study a mathematical model of a simple quality assurance scheme is used with calculation of kappa statistics with confidence limits and the significance of kappa statistic values is discussed in relation to other test performance statistics.

Methods

A statistical model was developed which simulated reporting of breast fine needle aspirates with a dichotomous outcome (benign/malignant). Populations from 50 to 1600 specimens were used and for increasing numbers of false positive results, the kappa statistics were calculated with 95% confidence intervals (CI) using the method described by Silcocks.⁵ Details of the model and calculations are given in the appendix. Other performance parameters, including the positive predictive value (PPV) of the test were, also calculated. The number of false positive results sufficient for the calculated upper 95% CI to fall below 1 was found for each population size. Two systems with differing mixes of benign and malignant specimens were used; a 50% benign, 50% malignant mix which could be used in a quality assurance scheme and a 80% benign, 20% malignant mix which represents the prior probabilities in breast fine needle aspirates received in the Department of Pathology at the Royal Hallamshire Hospital.

Table 1 Results for populations of 50% benign, 50% malignant specimens

No. of cases	False positive results*	Kappa	Lower CI	Upper CI	PPV†
50	4	0.78	0.58	0.99	71.4
100	5	0.86	0.74	0.98	80.0
200	5	0.90	0.81	0.99	88.9
400	6	0.95	0.92	0.99	93.0
800	6	0.97	0.95	0.99	96.4
1600	9	0.98	0.97	0.99	97.3

*Number of false positive results required for the upper 95% CI of the kappa statistic to fall below 1.

†PPV of a malignant result.

Table 2 Results for populations of 80% benign, 20% malignant specimens

No. of cases	False positive results*	Kappa	Lower CI	Upper CI	PPV†
50	4	0.84	0.63	0.97	83.3
100	5	0.90	0.81	0.99	90.9
200	5	0.95	0.91	0.99	95.2
400	6	0.97	0.95	0.99	97.0
800	8	0.98	0.97	0.99	98.0
1600	11	0.99	0.98	0.99	98.6

*Number of false positive results required for the upper 95% CI of the kappa statistic to fall below 1.

†PPV of a malignant result.

Results and Discussion

Tables 1 and 2 summarise the results.

If the kappa value is less than one, then we are 100% sure that performance is less than perfect and the upper confidence limit must also be less than one. In practice, however, methods of calculating confidence limits are approximate and high kappa values do have calculated upper limits of one when standard methods are used. Thus, to show by calculation that performance is significantly ($p < 0.05$) below a very high level (close to perfection) the calculated upper confidence limit must be less than one. At all population sizes and mixes in the present study, however, the number of false positive results required to produce an upper confidence limit below one was unacceptably high. The Cytology Subgroup of the National Coordinating Committee for Breast Screening Pathology recommend that the false positive rate should be less than 1%,⁶ but in this study the upper confidence limit of the kappa statistic fell below one, with a false positive rate of less than 1% only in populations with 800 and 1600 cases in a 50:50 population mix (rows 5 and 6 of table 1) and in the population of 1600 cases in the 80:20 mix (row 6 of table 2). Population numbers of 800 and 1600 represent the number of fine needle aspirates of the breast a pathologist might see during one or two years in a busy laboratory rather than the number of specimens that could be included in a quality assurance scheme.

The model used in this study is the simplest that could occur in pathology with a dichotomous outcome that can be verified (by clinical examination, mammography, histology, and follow up). Most other diagnostic situations in pathology would involve more diagnostic outcomes and those outcomes would be less easily verified (for example, diagnosis of inflammatory bowel disease by colorectal biopsy⁷ or assessment of chronic gastritis⁸). A

greater number of diagnostic outcomes would produce wider confidence limits for the kappa statistics if the population size remained the same.⁹ It therefore seems that the kappa statistic is not a sensitive indicator of performance in cytopathological or histopathological quality assurance schemes unless large numbers of samples are used. Other indicators should be considered, possibly using systems which weight the outcomes according to the importance of the result in clinical practice.¹⁰ The present study did not include the function of time and the persistence of low kappa statistics over a number of cycles of a quality assurance scheme would be a more reliable indicator of possible unsatisfactory performance. However, the frequency of cycles in most schemes is such that at least two years would elapse before such an assessment would be possible.

- 1 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- 2 Silcocks PB. Some issues in observer error studies in pathology. *J Pathol* 1992;168:255-6.
- 3 Svanholm H, Starklint H, Gundersen HJ, Fabricius J, Barlebo H, Olsen S. Reproducibility of histomorphologic diagnoses with special reference to the kappa statistic. *APMIS* 1989;97:689-98.
- 4 Donker DK, Hasman A, van Geijn HP. Interpretation of low kappa values. *Int J Biomed Comput* 1993;33:55-64.
- 5 Silcocks PB. Measuring repeatability and validity of histological diagnosis - a brief review with some practical examples. *J Clin Pathol* 1983;36:1269-75.
- 6 Cytology Sub-Group of the National Coordinating Committee for Breast Screening Pathology. Guidelines for Cytology Procedures and Reporting in Breast Cancer Screening. NHSBSP Publication No 22. Sheffield: NHSBSP Publications, 1993.
- 7 Seldenrijk CA, Morson BC, Meuwissen SG, Schipper NW, Lindeman J, Meijer CJ. Histopathological evaluation of colonic mucosal biopsy specimens in chronic inflammatory bowel disease: diagnostic implications. *Gut* 1991;32:1514-20.
- 8 Andrew A, Wyatt JJ, Dixon MF. Observer variation in the assessment of chronic gastritis according to the Sydney system. *Histopathology* 1994;25:317-22.
- 9 Morris JA. Information and observer disagreement in histopathology. *Histopathology* 1994;25:123-8.

10 Furness PN, Lauder I. Response analysis in histopathology EQA schemes. *J Clin Pathol* 1993;46:357-63.

Appendix

Outcome	Diagnosis		Total
	Benign	Malignant	
Benign	<i>a</i>	<i>b</i>	<i>g</i>
Malignant	<i>c</i>	<i>d</i>	<i>h</i>
Total	<i>e</i>	<i>f</i>	<i>i</i>

In the initial state of the model $b = 0$ and $c = 0$, and the rest of the cases are split between a and d in the proportions for that particular model (either 50:50 or 80:20 as described in the text). For each iteration of the model $b = b + 1$ and $a = a - 1$ —that is, there is replacement of one true negative with a false positive result. For each iteration of the model the parameters given below are calculated.

$$\text{Observed probability } P_o = \frac{a + d}{i}$$

$$\text{Expected probability } P_e = \frac{\left[\frac{g}{i} + \frac{h}{i} \right]}{i}$$

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

$$\text{Standard error of } \kappa \text{ } s.e. = \sqrt{\frac{P_o(1 - P_o)}{i(1 - P_e)^2}}$$

$$95\% \text{ CI } \quad CI = \kappa \pm (1.96 * s.e. \kappa)$$

$$\text{PPV of a malignant diagnosis } \quad PPV = \frac{d}{d + b}$$

J Clin Pathol 1996;49:599-602

Department of Pathology,
Royal Victoria Hospital,
Grosvenor Road,
Belfast BT12 6BA
W G McCluggage
H Bharucha

Department of Haematology
D Hull
E Mayne

The Queen's University of Belfast
H Bharucha

Department of Haematology,
St Mary's Hospital Medical School,
Imperial College of Science, Technology
and Medicine, London
S N Wickramasinghe

Correspondence to:
Dr W G McCluggage.

Accepted for publication
21 March 1996

Malignant lymphoma in congenital dyserythropoietic anaemia type III

W G McCluggage, D Hull, E Mayne, H Bharucha, S N Wickramasinghe

Abstract

A 60 year old woman with congenital dyserythropoietic anaemia (CDA) type III developed a malignant T cell lymphoma with cutaneous and widespread nodal involvement. Bone marrow aspirates showed erythroid hyperplasia and dyserythropoiesis with multinucleate erythroblasts and giantoblasts, in keeping with CDA type III. Electron microscopy showed multinucleate erythroblasts with notably irregular nuclear outlines and

intranuclear clefts. The development of malignant lymphoma in this patient, together with a documented high prevalence of monoclonal gammopathy and multiple myeloma and a single case of Hodgkin's disease, may indicate an increased incidence of lymphoproliferative disease in CDA type III.

(*J Clin Pathol* 1996;49:599-602)

Keywords: congenital dyserythropoietic anaemia type III, malignant lymphoma, electron microscopy.