

Comments on the paper “c-erbB-2 expression and nuclear pleomorphism in canine mammary tumors”

K. Metze and R.L. Adam

Departamento de Anatomia Patológica, Faculdade de Ciências Médicas,
Universidade Estadual de Campinas, Campinas, SP, Brasil

Quantification in histopathology - some pitfalls

Correspondence

K. Metze
Departamento de Anatomia
Patológica
FCM, UNICAMP
13081-970 Campinas, SP, Brasil
Fax: +55-19-3289-3897
E-mail: kmetze@fcm.unicamp.br

K. Metze is a senior researcher of
CNPq.

Publication supported by FAPESP.
.....

Key words

- Canine mammary tumors
 - Nuclear pleomorphism
 - Statistical analysis
 - Scientific methods
 - c-erbB-2 expression
 - MIB-1 expression
-

Received November 26, 2004
Accepted January 3, 2005
.....

In their interesting study on canine mammary tumors published in the *Brazilian Journal of Medical and Biological Research*, Dutra et al. (1) evaluated the influence of c-erbB-2 expression and nuclear polymorphism on survival using a semiquantitative scoring system. Survival was calculated using the Spearman rank correlation test. They obtained a regression coefficient of $r = 0.441$ for c-erbB-2 expression and of $r = -0.295$ for nuclear polymorphism, with $P > 0.05$ in both cases. Therefore, the authors concluded that neither c-erbB-2 nor nuclear polymorphism could be considered to be prognostic factors.

We would like to discuss some of the methodological aspects of the study.

1. In survival studies some individuals will still be alive at the end. Others may have been lost to follow-up some time earlier and thus survival times will be unknown for a subset of the study group. There is a consensus that these so-called “censored” cases must be included in the analysis. Most survival analyses use Kaplan-Meier plots, log rank tests or the Cox proportional hazard regression (2). A Spearman rank order correlation test, as carried out by Dutra et al. (1), can only be used if all the animals under observation die (2).

2. Both variables examined, c-erbB-2 expression and nuclear pleomorphism, were evaluated using a semiquantitative scoring

system. Categorizing variables that are measured on a continuum is a frequently used procedure in medicine, especially in surgical pathology. By looking, for instance, at recently published issues of this Journal, we can see considerable variation in the scoring systems applied in histopathology, not only between different groups but also within the same laboratory (1,3-6).

It is well known that categorization may be useful for clinical decision-making, treatment recommendations, determining study eligibility, or for illustration purposes (7), but we should be well aware that this procedure is dangerous for several reasons.

When the probability of an event associated with a quantitative prognostic factor varies monotonously, the selection of a cut-off point is always arbitrary. This may create scientifically unrealistic models because this procedure discards potentially important information since all values between the cut-off levels have an equal effect on the model. Therefore, categorization considerably lowers the statistical test-power and can be seen as an introduction of measurement errors (8-14).

This can be easily illustrated with a simple experiment. In Table 1 we simulated a possible situation similar to that of the study of Dutra et al. (1): let us assume that 10 of 17 dogs with neoplasia died. The percentages of c-erbB-2-positive cells in the tumors had

Table 1. Simulation of survival data in order to show the effect of categorization.

Final event expired = 1	Follow-up (weeks)	Percentage of positive cells	HercepTest score
1	45	8	0
1	19	3	0
1	75	9	0
0	35	6	0
1	74	9	0
0	55	9	0
1	25	4	0
0	65	4	0
1	50	8	0
0	70	3	0
1	40	9	0
0	18	10	2
1	45	10	2
0	50	10	2
1	18	45	3
1	25	35	3
0	30	13	3

been counted as percentage values and then were transformed by the HercepTest scoring system, assuming that none of the cells showed faint staining. When analyzing the scores by the Kaplan-Meier survival plot followed by the log rank test we cannot show any significant influence of c-erbB-2 on overall survival ($P = 0.131$; SPSS 10.0 software). Similarly, c-erbB-2 scores analyzed by Cox regression were not significant either when simply entering the categories or when using the quotient calculated between observed and expected events in the log rank test of the Kaplan-Meier model, as suggested recently, thus considering the natural order between the categories (15). But when we include the continuous percentage values in the Cox proportional hazard model, we discover a significant influence of c-erbB-2 expression on overall survival ($P = 0.017$, $B = 0.0834$; SPSS 10.0).

Thus, our example shows that even in studies with small sample size it is possible to detect prognostic factors with statistical significance when using continuous data. Choosing the cut-off points of a continuously measured variable is in most cases

arbitrary or opportunistic (16). The so-called minimum P-value approach involves multiple testing which inflates the type I error rates and is therefore not recommended (8). Variations in the choice of cut-off points between different studies provoke a bias called the 'Will Rogers phenomenon', which describes an apparent improvement of prognosis in subgroups due to classification or categorization changes, but which is not accompanied by an overall improvement of survival (11). Therefore, the results of studies using different cut-off points may not be comparable at all. In this context the use of a previously defined classification system, in our case the HercepTest scoring system, is certainly of advantage. But this classification system was created in order to identify patients who would probably benefit from a specific therapy. Yet this is not the topic of the present study on prognostic variables and therefore possible relevant cut-off points may be completely different for this purpose. In summary, the accuracy and reliability of a study may be compromised by categorization of continuous data or qualitative interpretations. Categorization of a variable which can be continuously measured should be done only exceptionally. For instance, when precise measurement is difficult, the distribution of the variable is highly skewed or its relation with another variable is non-linear (8,10,17).

Therefore, we suggest that immunohistochemical expression should be quantified as done for the MIB-1 antibody, i.e., by counting the percentage of antibody-stained cells. One may argue that this method will not take into consideration differences in staining intensity between the cells. In this case the immunohistochemical reaction product may be easily quantified with interactive software applied to digitized images. Nuclear pleomorphism can also be easily and precisely quantified using texture analysis of digitized images (18-21). Therefore, we believe that a re-evaluation of the material

presented in the study by Dutra et al. (1) using continuous data and a survival analysis including censored observations might

reveal the prognostic relevance of the variables under study.

References

- Dutra AP, Granja NVM, Schmitt FC & Cassali GD (2004). c-erbB-2 expression and nuclear pleomorphism in canine mammary tumors. *Brazilian Journal of Medical and Biological Research*, 37: 1673-1681.
- Clark TG, Bradburn MJ, Love SB & Altman DG (2003). Survival Analysis Part I: Basic concepts and first analyses. *British Journal of Cancer*, 89: 232-238.
- Engelman DES, Andrade LALA & Vassallo J (2003). Human papillomavirus infection and p53 protein expression in vulvar intraepithelial neoplasia and invasive squamous cell carcinoma. *Brazilian Journal of Medical and Biological Research*, 36: 1159-1165.
- Sredni ST, Zerbini MCN, Latorre MR & Alves VAF (2003). P53 as a prognostic factor in adrenocortical tumors of adults and children. *Brazilian Journal of Medical and Biological Research*, 36: 23-27.
- Batista SS, Pires RS & Britto LRG (2002). Differential expression of AMPA-type glutamate receptor subunits during development of the chick optic tectum. *Brazilian Journal of Medical and Biological Research*, 35: 973-978.
- Brenna SMF, Zeferino LC, Pinto GA, Souza RA, Andrade LALA, Vassallo J, Martinez EZ & Syrjänen KJ (2002). C-Myc protein expression is not an independent prognostic predictor in cervical squamous cell carcinoma. *Brazilian Journal of Medical and Biological Research*, 35: 425-430.
- Mazumdar M & Glassman JR (2000). Categorizing a prognostic variable: Review of methods, code for easy implementation and applications to decision-making about cancer treatments. *Statistics in Medicine*, 19: 113-132.
- Clark TG, Bradburn MJ, Love SB & Altman DG (2003). Survival Analysis Part IV: Further concepts and methods in survival analysis. *British Journal of Cancer*, 89: 781-786.
- Mazumdar M, Smith A & Bacik J (2003). Methods for categorizing a prognostic variable in a multivariable setting. *Statistics in Medicine*, 22: 559-571.
- Streiner DL (2002). Breaking up is hard to do: The heartbreak of dichotomizing. *Canadian Journal of Psychiatry*, 47: 262-266.
- Donadieu J, Auclerc MF, Baruchel A et al. (2000). Prognostic study of continuous variables (white blood cell count, peripheral blast cell count, haemoglobin level, platelet count and age) in childhood acute lymphoblastic leukaemia. Analysis of a population of 1545 children treated by the French Acute Lymphoblastic Leukaemia Group (FRALLE). *British Journal of Cancer*, 83: 1617-1622.
- Taylor JMG & Yu MG (2002). Bias and efficiency loss due to categorizing an explanatory variable. *Journal of Multivariate Analysis*, 83: 248-263.
- Metze K (1998). Methodological aspects of prognostic factor studies: some caveats. *São Paulo Medical Journal*, 116: 1787-1788.
- Metze K (1998). Methodological problems of grading tumour regression: responders compared to non-responders. *Journal of Cancer Research and Clinical Oncology*, 124: 281-282.
- Lorand-Metze I, Pinheiro MP, Ribeiro E, de Paula EV & Metze K (2004). Factors influencing survival in myelodysplastic syndromes in a Brazilian population: comparison of FAB and WHO classifications. *Leukemia Research*, 28: 587-594.
- Figueiras A & Cadarso-Suarez C (2001). Application of nonparametric models for calculating odds ratios and their confidence intervals for continuous exposures. *American Journal of Epidemiology*, 154: 264-275.
- Vassallo J, Metze K, Traina F, de Souza CA & Lorand-Metze I (2003). The prognostic relevance of apoptosis-related proteins in classical Hodgkin's lymphomas. *Leukemia and Lymphoma*, 44: 483-488.
- Metze K, Souza Filho W, Adam RL & Lorand-Metze I (2001). Analysis of the component "tree" as a new tool for analytical cellular pathology. *Analytical Cellular Pathology*, 22: 65 (Abstract).
- Metze K, Adam RL, Silva PV, De Carvalho RB & Leite NJ (2004). Analysis of chromatin texture by Pinkus' approximate entropy. *Cytometry. Part A*, 59A: 63 (Abstract).
- Adam RL, Leite NJ, De Carvalho RB, Silva PV & Metze K (2004). Granulometric residues as a diagnostic tool in cytology. *Cytometry. Part A*, 59A: 63 (Abstract).
- Adam RL, Ribeiro E, Metze K, Leite NJ & Lorand-Metze I (2004). Morphometric and granulometric features of erythroblasts as a diagnostic tool of hematologic diseases. *Cytometry. Part A*, 59A: 46 (Abstract).