University of York Department of Health Sciences

Measurement in Health and Disease

Exercise: sensitivity and specificity

1. The following is taken from a paper on the detection of HIV.

In Africa, serological testing for HIV infection is both expensive and difficult to obtain and so a study sought to assess the value of regional lymph node enlargement, by site and by size, as a predictor of HIV disease. The sensitivities and specificities were given as follows:

Site (size (cm)) of lymph node	Sensitivity (Rate (%))	Specificity (Rate (%))
Axillary (≥1)	63/146 (43)	107/113 (95)
Axillary (≥0.5)	110/146 (75)	78/113 (69)
Submandibular (≥1)	31/146 (21)	109/113 (96)
Submandibular (≥0.5)	110/146 (75)	84/113 (74)
Epitrochlear (≥1)	53/146 (36)	102/113 (90)
Epitrochlear (≥0.5)	123/146 (84)	92/113 (81)
Epitrochlear (≥0.5) +axillary (≥1)	63/146 (43)	108/113 (96)
Epitrochlear (≥0.5) +submandibular (≥1)	31/146 (21)	111/113 (98)
Axillary (≥1) +submandibular (≥1)	34/146 (23)	111/113 (98)
Epitrochlear (≥0.5) +axillary (≥1) +submandibular (≥1)	26/146 (18)	112/113 (99)

The positive and negative predictive values were also given:

Site (size (cm)) of lymph node	Positive predictive value (Rate (%))	Negative predictive value (Rate (%))
Axillary (≥1)	63/69 (91)	107/190 (56)
Axillary (≥0.5)	110/145 (76)	78/114 (68)
Submandibular (≥1)	31/35 (89)	109/224 (49)
Submandibular (≥0.5)	110/139 (79)	84/120 (70)
Epitrochlear (≥1)	53/64 (83)	102/195 (52)
Epitrochlear (≥0.5)	123/144 (85)	92/115 (80)
Epitrochlear (≥0.5) +axillary (≥1)	63/68 (93)	108/191 (57)
Epitrochlear (≥0.5) +submandibular (≥1)	31/33 (94)	111/226 (49)
Axillary (≥1) +submandibular (≥1)	34/36 (94)	111/223 (50)
Epitrochlear (≥0.5) +axillary (≥1) +submandibular (≥1)	26/27 (96)	112/242 (46)

(Malin, A., Ternouth, I., and Sarbah, S. Epitrochlear lymph nodes as marker of HIV disease in sub-Saharan Africa. *British Medical Journal* 1994; **309**: 1550-1.)

QUESTIONS

a) What are the sensitivity and specificity of a test? What do they tell us?

b) How could we present the sensitivity and specificity graphically?

c) What are positive and negative predictive values and on what do they depend?

d) Why are the denominators constant in the sensitivity and specificity columns but varying in the positive predictive value column?

e) Why does the sensitivity go up and the specificity go down if we reduce the size of node which we consider positive?

f) Which two diagnostic tests give the strongest relationship to HIV and how do they differ?

(Source: Bland M and Peacock JL. (2000) *Statistical Questions in Evidence-based Medicine* Oxford University Press, Oxford.*)*

2. In a discussion of testing for HIV the child of an HIV positive mother, the following appeared: ‘These tests are not “notorious for false positives”. It is well known that they may produce them, but the reliability increases in high prevalence situations.’

(Talbot. Positively false on HIV. *The Guardian*, London, page 21, 28 September, 1999).

QUESTION

In what sense does the reliability of a test increase as the prevalence increases and how is this relevant to testing such a child?

(Source: Bland M and Peacock JL. (2000) *Statistical Questions in Evidence-based Medicine* Oxford University Press, Oxford.*)*

3. The following is the abstract of a paper.

Background: This study investigated the sensitivity and specificity of a computer-automated telephone system to evaluate cognitive impairment in elderly callers to identify signs of early dementia.

Methods: The Clinical Dementia Rating Scale was used to assess 155 subjects aged 56 to 93 years (n=74, 27, 42, and 12, with a Clinical Dementia Rating Scale score of 0, 0.5, 1, and 2, respectively). These subjects performed a battery of tests administered by an interactive voice response system using standard Touch-Tone telephones. Seventy-four collateral informants also completed an interactive voice response version of the Symptoms of Dementia Screener.

Results: Sixteen cognitively impaired subjects were unable to complete the telephone call, Performances on 6 of 8 tasks were significantly influenced by Clinical Dementia Rating Scale status. The mean (SD) call length was 12 minutes 27 seconds (2 minutes 32 seconds). A subsample (n=116) was analyzed using machine-learning methods, producing a scoring algorithm that combined performances across 4 tasks. Results indicated a potential sensitivity of 82.0% and specificity of 85.5%. The scoring model generalized to a validation subsample (n=39), producing 85.0% sensitivity and 78.9% specificity. The kappa agreement between predicted and actual group membership was 0.64 (P<.001). Of the 16 subjects unable to complete the call, 11 provided sufficient information to permit us to classify them as impaired, Standard scoring of the interactive voice response-administered Symptoms of Dementia Screener (completed by informants) produced a screening sensitivity of 63.5% and 100% specificity. A lower criterion found a 90.4% sensitivity, without lowering specificity.

Conclusions: Computer-automated telephone screening for early dementia using either informant or direct assessment is feasible. Such systems could provide wide-scale, cost-effective screening, education, and referral services to patients and caregivers.

(Mundt JC, Ferber KL, Rizzo M, Greist JH. Computer-automated dementia screening using a touch-tone telephone. *Archives of Internal Medicine* 2001; **161**: 2481-2487.)

QUESTIONS

a) How might we describe the agreement between predicted and actual group membership?

b) Lowering the criterion for dementia increased sensitivity without lowering specificity. What does this tell us about the relationship between the score and dementia classification?