

How to appraise an article on diagnosis

Hagen Rampes, James P. Warner and Robert Blizard

In our evidence-based journal club we appraised an article investigating the use of the CAGE questions in screening psychiatric populations for alcohol misuse. We calculated a likelihood ratio of four for a score of two or more positive CAGE questions, suggesting the CAGE is a moderately useful screening instrument.

The importance of evidence-based medicine in psychiatry is gradually becoming apparent (Geddes, 1996). We continue our series of articles, based on the experience of a journal club, introducing the principles that are utilised in critical appraisal. The present article deals with a paper on a diagnostic test. This article is intended to give a brief, focused appraisal, not a comprehensive critique of the paper selected for appraisal. It would be useful, but not essential, to read the paper appraised in conjunction with this article.

Use of the CAGE questionnaire in detecting alcohol dependence

Vignette

A 36-year-old unemployed male presented with a two-week history of worsening of auditory hallucinations. He had a past psychiatric history of paranoid schizophrenia. He scored two out of four questions on the CAGE questionnaire (Mayfield *et al.*, 1974). The CAGE questions are a brief screening tool widely used to detect problem drinking, consisting of: Have you ever felt you should cut down on your drinking?; Have people annoyed you by criticising your drinking?; Have you ever felt bad or guilty about your drinking?; Have you ever had a drink first thing in the morning to steady your nerves or get rid of a hangover (eye-opener)?

In the ensuing discussion of this case, the significance of the score of the CAGE questionnaire was discussed. It was decided to investigate this further.

Question

What is the evidence to support the use of the CAGE questionnaire in detecting alcohol dependence or misuse in psychiatric in-patients?

It is helpful to formulate a question in three parts. The first part concerns the population (i.e. the type of patient that you are interested in), here we are interested in psychiatric in-patients. We could narrow the field further, for example to young men with schizophrenia, although this would reduce our chances of finding a relevant study. The second part concerns the intervention or manoeuvre. Here the manoeuvre is the use of the CAGE questionnaire. The third part of the question is the outcome. We are interested in a diagnosis of alcohol dependence and or misuse.

Literature search

One of us remembered seeing a discussion of the merits of the CAGE questionnaire in *Bandolier* (Moore *et al.*, 1997). In the discussion in *Bandolier*, a paper on the use of CAGE was appraised. However, the population was from an out-patient medical practice in an urban teaching hospital from Virginia, USA (Buchsbaum *et al.*, 1991). Clearly it would be preferable to find a paper concerning the patient in our vignette (i.e. a psychiatric in-patient). A Medline search (WinSpirs 2.0) using the term 'CAGE' for a textword search was carried out from 1966 to 1987. This identified 4662 articles. Second, a textword search for articles containing the word 'hospital' was carried out, revealing 657 016 articles. Third, a textword search for articles containing the words starting with 'psychiat' was conducted by truncating the word with an asterisk (Psychiat*) (revealing 142 424 articles). Combining records one, two and three by using the boolean term 'and' refines our search and identifies a manageable number of relevant articles. The abstracts of these 19 articles were scanned and the one that appeared most relevant was extracted.

The article chosen was entitled 'Comparison of questionnaire and laboratory tests in the detection of excessive drinking and alcoholism' (Bernadt *et al.*, 1982).

Getting the article

The article was easily retrieved from the Chelsea and Westminster Hospital Library and obtained immediately.

Brief outline of the article

The aim of the article was to compare the use of different questionnaires and laboratory tests in detecting 'excessive drinking' and 'alcoholism'. Three hundred and eighty-five (198 male) psychiatric in-patients (aged 16–65) at the Maudsley and Bethlem Hospitals were recruited in a 10-month period in 1980. A research nurse administered a structured interview that included the CAGE test and two other screening tests; the Brief Michigan Alcohol Screening Test (Selzer, 1971) and the Reich interview (Reich *et al*, 1975). Within 48 hours of admission, blood was taken for mean corpuscular volume, gamma glutamyl transpeptidase test, aspartate transaminase, alkaline phosphatase, and other tests thought to be influenced by alcohol consumption. A score of two or more on the CAGE was taken as indicating 'alcoholism'. Consumption of more than 16 'drinks' per day over the year before admission was taken as drinking a hazardous amount or 'excessive drinking'. Case notes were perused after discharge to determine whether a primary or secondary diagnosis of 'alcoholism' was recorded. Out of 385 patients, 371 were included in the data analysis (185 male). Forty-two were categorised as 'excessive drinkers' and 49 had primary or secondary 'alcoholism'.

Critical appraisal of the article on diagnosis

This followed the recommendations of the Evidence-Based Medicine Working Group (Jaeschke *et al*, 1993, 1994). There are three main components to the appraisal: Are the results valid? What are the results? Will the results help me in patient care?

Are the results of the study valid?

Was there an independent, blind comparison with a reference standard? The reference standards used were clinician's diagnosis, Murray's Screening interview (Murray, 1977) and Research Diagnostic Criteria for alcoholism (Spitzer & Endicott, 1978). Bernadt *et al* do not state whether these were conducted independently or blind to the result of the screening test.

Did the patient sample include an appropriate spectrum of patients to whom the diagnostic test will be applied in clinical practice? The patients were in a group of 385 consecutive admissions to

the Bethlem and Maudsley Hospitals in a 10-month period. Although the Maudsley is a tertiary referral centre, taking some specialist patients, we feel that it is unlikely this will influence the diagnosis of alcoholism to any large degree.

Did the results of the test being evaluated influence the decision to perform the reference standard? No. It appears that all admissions had the screening test and the reference standards.

Were the methods for performing the test described in sufficient detail to permit replication? Yes. The authors report the method in considerable detail.

What were the results?

The authors only present data for patients who scored two or more on the CAGE questionnaire. For this cut-off, we are given a sensitivity of 0.91 and specificity of 0.77 for the detection of alcoholism. From this we can reconstitute the results table (see Table 1). Therefore, using this criterion, 91% of 'alcoholics' and 77% of 'non-alcoholics' will be correctly classified.

Are likelihood ratios for the test results presented or data necessary for their calculation provided? Likelihood ratio is the ratio of the likelihood of the disease being present given a positive test result, to the likelihood of no disease being present given a positive test result. In other words, it provides an index of increased likelihood of a disease being present given a positive test result. The likelihood ratio for a positive result is calculated simply by: sensitivity/(1 – specificity)

Likelihood ratios may also be calculated for a negative test result. In this case, the likelihood ratio for a negative test provides an index of how likely a disease is to be absent given a negative result (see Sackett, 1997).

Using the sensitivity and specificity provided, the likelihood ratio for a score of two or more on the CAGE is approximately four. The usefulness

Table 1. Reconstituted results table for CAGE performance in detecting alcoholism (Adapted from Bernadt *et al*, 1982)

	Alcoholism present	Alcoholism absent	
CAGE score 2+	45	74	119
CAGE score 1 or 0	4	248	252
Total	49	322	371

Sensitivity=0.91; specificity=0.77; likelihood ratio for a positive test=0.91/0.23 \approx 4.

Table 2. Likelihood ratio values and their relevance to diagnosis

Likelihood ratio	Diagnostic impact
10	Very positive
3	Moderately positive
1	Neutral
0.3	Moderately negative
0.1	Very negative

The higher the likelihood ratio the more likely a positive test will accurately indicate the presence of a disease. A test with a likelihood ratio of one will not help clinicians decide whether a disease is present or absent. Likelihood ratios for negative tests are calculated by $(1 - \text{sensitivity})/\text{specificity}$. A likelihood ratio of less than one will reduce the post-test odds. The lower the likelihood ratio for a negative test, the more likely that negative test excludes the disease (after Sackett *et al.* 1997).

of likelihood ratios, given in Table 2, shows this is in the region of a moderately helpful result. In other words, a positive result of a test with a likelihood ratio of four moderately increases our certainty of the disease being present.

Will the results help me in caring for my patients?

Will the reproducibility of the test result and its interpretation be satisfactory in my setting? Yes, we feel it is. The CAGE questionnaire is simple to administer and interpret.

Are the results applicable to my patient? If a patient in a similar setting to that of the study scores two or more on the CAGE, then the likelihood ratio of a diagnosis of alcoholism is $0.91/0.23 \approx 4$. This result would be about four times as likely to be seen in someone with alcoholism, as opposed to someone without alcoholism. By using the likelihood ratio in conjunction with the probability of a disease being present before the test was carried out, the probability of the disease being present with a positive result can be derived. This may be done by multiplying the pre-test odds by the likelihood ratio to get the post-test odds.

Odds are related to probability and are calculated by dividing the probability by $1 - \text{probability}$. For example, the probability of a pregnant woman having a boy is 50%. The odds are $0.5/(1 - 0.5)$ or 1 (evens). In our example the pre-test probability of a patient having alcoholism was around 15% (a figure that accords to our clinical practices). The odds are $0.15/0.85 = 1/6$. If we then multiply $1/6$ by the likelihood ratio (4), we get $4/6$ (or 0.666). This is the post-test odds. Converting this back to the post-test probability

$0.666/(1+0.666)=0.4$. So if our patient scores two or more on the CAGE, the probability he has alcoholism has risen from 15% to 40%.

Will the results change my management? Quite possibly. The CAGE is not a diagnostic tool, but a positive result should alert clinicians to a greater possibility of alcohol misuse. Furthermore, searching questions could then be asked in order to confirm or refute the diagnosis. If alcohol misuse is present this could have a significant impact on symptomatology and outcome. Comorbidity is a major problem that has important implications for treatment and risk assessment.

Will patients be better off as a result of the test? This is more difficult to answer. The prevailing clinical model: test → diagnosis → treatment → improved outcome, represents an ideal which, in practice, may not be achievable. A positive test should result in a further evaluation of the patient. Our patient's current symptoms may be due to alcohol alone, in which case detoxification may lead to a resolution of his symptoms. Alternatively, he may have schizophrenia and problem drinking. In this event, tackling his drinking may help his psychotic symptoms, or may lead to better adherence to his treatment regimen. In this scenario, whether the appropriate treatment for his alcohol misuse leads to improved outcome would merit a further evidence-based medicine exercise.

Comment

Alcohol misuse is common in psychiatric practice. The CAGE is a well established screening tool for alcohol misuse, and the article reviewed here investigates its utility in a psychiatric setting. The results of the article suggest a score of two or more on the CAGE should alert clinicians to the possibility of alcohol misuse. The article was quite old and terminology in this domain has changed. Many people no longer find the term 'alcoholism' acceptable, but Bernadt *et al* and clinicians today will have similar interpretations of the term.

The likelihood ratio of four for a score of two or more on the CAGE in the study population suggests that it is a reasonable test. However, the likelihood ratio for the Brief Michigan Alcohol Screening Test, using a cut-off of six, was eight; double that of the CAGE. Other tests frequently used to diagnose alcohol misuse perform less well; the likelihood ratio of an abnormal mean corpuscular volume is two, and a raised gamma glutamyl transpeptidase test 2.5. Therefore, the CAGE appears to be a much better discriminator than either of these biochemical tests.

It is unfortunate that the authors used an *a priori* cut-off of two in the CAGE. Previous studies have found much higher likelihood ratios for a score of two, three or four. Bush *et al* (1987), in a study of 518 medical patients found a likelihood ratio of 19 for a score of two, 170 for three and infinity for a score of four. Incidentally, they found similarly disappointing utility of mean corpuscular volume and gamma glutamyl transpeptidase test. Sackett *et al* (1991), in their analysis of the interpretation of the use of the CAGE, found a cut-off of two or more provided a likelihood ratio of seven. The current article suggests a lower value. This may be because the 'gold standard' diagnosis was different, or that a psychiatric population responds differently to the questions. For example, psychiatric patients' responses may be contaminated by disturbance of affect. In particular, two CAGE items expressly refer to guilt and annoyance. The suggestion that the CAGE performs less well in a psychiatric population raises some questions about its applicability in this group.

We found this a particularly rewarding exercise. The use of likelihood ratios helps to put 'meat on the bones' of screening and diagnostic tests. Although the maths may appear complex, it is relatively easily mastered, and readers are referred to Sackett *et al* (1997) for a clear explanation and example.

References

- BUCHSBAUM, D. G., BUCHANAN, R. G., CENTOR, R. M., *et al* (1991) Screening for alcohol abuse using CAGE scores and likelihood ratios. *Annals of Internal Medicine*, **115**, 774-777.
- BERNADT, M. W., MUMFORD, J., TAYLOR, C., *et al* (1982) Comparison of questionnaire and laboratory tests in the detection of excessive drinking and alcoholism. *Lancet*, *i*, 325-328.
- BUSH, B., SHAW, S., CLEARY, P., *et al* (1987) Screening for alcohol abuse using the CAGE questionnaire. *American Journal of Medicine*, **82**, 231-235.
- GEDDES, J. (1996) On the need for evidence-based psychiatry. *Evidence-Based Medicine*, **1**, 199-200.
- JAESCHKE, R., GUYATT, G. H. & SACKETT, D. L. (1993) Users' guides to the medical literature. III. How to use an article about a diagnostic test A. Are the results of the study valid? *Journal of the American Medical Association*, **271**, 389-391.
- , — & — (1994) Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *Journal of the American Medical Association*, **271**, 703-707.
- MAYFIELD, D., MCLEOD, G. & HALL, P. (1974) The CAGE questionnaire: validation of a new alcoholism screening instrument. *American Journal of Psychiatry*, **131**, 1121-1123.
- MOORE, A., MCQUAY, H. & MUIR GRAY, J. A. (1997) Clinical scoring for alcohol abuse. *Bandolier*, **4**, 5-6.
- MURRAY, R. P. (1977) Screening and early detection instruments for disabilities related to alcohol consumption. In *Alcohol-Related Disabilities* (eds G. Edwards *et al*). Geneva: World Health Organization.
- REICH, T., ROBINS, L. N., WOODRUFF, R. A., *et al* (1975) Computer assisted derivation of a screening interview for alcoholism. *Archives of General Psychiatry*, **32**, 847-852.
- SACKETT, D. L., HAYNES, R. B., GUYATT, G. H., *et al* (1991) The interpretation of diagnostic data. In *Clinical Epidemiology: A Basic Science for Clinical Medicine* (2nd edn). Boston, MA: Little Brown and Company.
- , RICHARDSON, W. S., ROSENBERG, W., *et al* (1997) *Evidence-Based Medicine: How to Practice and Teach EBM*. New York: Churchill Livingstone.
- SELZER, M. L. (1971) The Michigan Alcoholism Screening Test. *American Journal of Psychiatry*, **127**, 1653-1658.
- SPITZER, R. L. & ENDICOTT, J. (1978) Medical and mental disorder: proposed definition and criteria. In *Critical Issues in Psychiatric Diagnosis* (eds R. L. Spitzer & D. F. Klein). New York: Raven Press.
- Hagen Rapses, Senior Registrar, South Kensington and Chelsea Mental Health Centre, Chelsea and Westminster Hospital, London and *James P. Warner, Lecturer, Robert Blizard, Medical Statistician, University Department of Psychiatry, Royal Free Hospital School of Medicine, Rowland Hill Street, London NW3 2QG

*Correspondence