be used routinely in all patients considered at risk for eating disorders.

We thank Martin Bland, professor of medical statistics, for guidance in planning the research methods.

References

- 1 King MB. Eating disorders in a general practice population: prevalence, characteristics and follow-up at 12 to 18 months. *Psychol Med Monogr* Suppl 1989;16:191-194.
- 2 Henderson M, Freeman CPL. A self-rating scale for bulimia: the "BITE." Br J Psychiatry 1987;150:18-24.
- 3 Garner DM, Olmstead MA, Polivy J. Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *Int J Eating Disorders* 1983;2:15-34.
- 4 Ewing JA. Detecting alcoholism: the CAGE questionnaire. *JAMA* 1984;252:1905-1907.
- 5 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. 4th ed. Washington (DC): American Psychiatric Press; 1990.
- 6 Greenhalgh T. Papers that report diagnostic or screening tests. *BMJ* 1997;315:540-543.

COMMENTARY A promising instrument, but more research is needed

One strength of the SCOFF questionnaire is its simplicity and ease of use. The instrument can be administered and scored in a matter of minutes. Further, it seems to need no specialized training or qualifications to use. Preliminary diagnostic results from the SCOFF questionnaire are impressive. The SCOFF questionnaire was able to correctly identify 100% of participants with eating disorders (both anorexia nervosa and bulimia nervosa) and correctly rule out 87.5% of controls who did not have eating disorders. The sample size, although not impressive, was sufficient to provide meaningful conclusions in the present sample.

It is unclear from the article how the diagnoses of eating disorders and the absence of eating disorders in controls were established. The use of a structured diagnostic interview such as the Structured Clinical Interview from the *Diagnostic and Statistical Manual of Mental Disorders* (SCID) to confirm or rule out eating disorder diagnoses would clearly strengthen the study.

It is also unclear that Americans will correctly interpret "make you really sick" in the first question ("make yourself sick" may be clearer), and the use of "one stone" will be meaningless (about 6.3 kg, or 14 lb). Also, some will find the term "SCOFF" unpleasant at best.

The sample is limited in several respects. First, all cases presumably met full diagnostic criteria for anorexia nervosa or bulimia nervosa. It is, therefore, unclear how participants who were subthreshold for these conditions would score on the SCOFF questionnaire. Further, the ratio of patients with eating disorders to those without eating disorders in the present study-roughly 1.2:1-is not representative of that found in the general population. Although this should not influence estimates of sensitivity and specificity (because each is calculated within cases and controls, respectively), it would artificially inflate estimates of positive predictive power-that is, what percentage of those testing positively actually have an eating disorder. Another concern with the present sample is that all of the patients had previously been diagnosed as having an eating disorder. This may have artificially inflated the sensitivity of the instrument because these participants may have been more willing to answer yes to items on the SCOFF

questionnaire than those with eating disorders who had not been previously diagnosed.

No data are presented on the reliability of the SCOFF questionnaire. Because the proposed scoring assumes equivalence of items (that all items have equal weight), it would be important to examine the internal consistency of the items (for example, the Cronbach alpha). Test-retest reliability coefficients would also be informative. Further, if the SCOFF questionnaire is intended to be administered in interview form (as it was in the present study), interrater reliability may also be relevant. It would have been informative if data had been presented on responses to individual SCOFF items by diagnostic group. Also, the authors should consider presenting a receiver operating characteristic curve showing sensitivity and specificity across a range of cutoff scores.

One issue that is likely to be raised with respect to the SCOFF questionnaire-but that is by no means specific to the instrument-is the relatively low base rate of eating disorders (anorexia nervosa, 0.5%, and bulimia nervosa, 1%-2% in the age groups at risk) in the general population. As a result, the SCOFF questionnaire is likely to produce a relatively large number of false-positive cases (respondents who test positively on the instrument but do not have a diagnosis of an eating disorder). Whether this false-positive rate is a matter of concern can only be evaluated by the relative costs, financial and otherwise, of classification errors in the specific situation in which the SCOFF questionnaire is used. Costs to be considered (including those to patients) should include those associated with administration, with following up a patient who is falsely positive, and with missing a patient who has the disease. Different situations may have different costs associated with these errors. These costs may dictate whether the instrument should be used at all, or whether alternative cutoff points should be used to better balance the relative costs of misclassification.

This study represents the first logical step in the psychometric development of the SCOFF questionnaire. Although the preliminary results are impressive, further development is clearly needed. Ross D Crosby Department of Biomedical Statistics James E Mitchell Neuropsychiatric Research Institute 700 First Ave S, Box 1415 Fargo, ND 58107 Correspondence to:

Dr Mitchell Jmitchell@nrifargo.com

West J Med 2000;172:165