from month 1 after implementation and using the nurse administrator's ESI score as the second assessment, I asked residents to compare high-risk and lower-risk triage scores between the triage nurse and the nurse administrator. The resulting 2 × 2 table is completed as shown in Fig. 1, and calculation of chance agreement proceeds as follows:

kappa = [(observed agreement – expected agreement)/(1 – expected agreement)]

High-risk assessments by nurse administrator: 11/25 = 0.44

High-risk assessments by triage nurse: 10/25 = 0.40

Lower-risk assessments by nurse administrator: 14/25 = 0.56

Lower-risk assessments by triage nurse: 15/25 = 0.60

Observed agreement = (9 + 13)/25 = 0.88

Expected agreement = (chance of high-risk assessment) + (chance of lower-risk assessment)

Chance of high-risk assessment = 0.44 × 0.40 = 0.176

Chance of lower-risk assessment = 0.56 × 0.60 = 0.336

Expected agreement by chance alone = 0.176 + 0.336 = 0.512

kappa = (0.88 – 0.512)/(1 – 0.512) = 0.368/0.488 = 0.75

Table 1 in both the teachers'[1] and learners'[4] versions of this article references Maclure and Willett[5] as a source of the qualitative classification of kappa. My own review of that paper did not reveal any attempt to qualitatively assess kappa, but at least 3 other sources have done so.[6–8] In my experience the most widely used classification for kappa is the last of these,[8] which proposed the guidelines for interpreting kappa values as outlined in Table 1 in this letter.

Christopher R. Carpenter
Assistant Professor
Division of Emergency Medicine
Washington University School of Medicine
St. Louis, Mo.

References

1. McGinn T, Wyer PC, Newman TB, Keitz S, Leipzig R, Guyatt G, for Evidence-Based Medicine Teaching Tips Working Group. Tips for teachers of evidence-based medicine: 3. Understanding and calculating kappa. *CMAJ* 2004;171(11):Online-1 to Online-9. Available: www.cmaj.ca/cgi/data/171/11/1369/DC1/1 (accessed 2005 Jan 10).
2. Dawson B, Trapp RG. *Basic and clinical biostatistics.* 3rd ed. New York: McGraw-Hill; 2001. p. 66-7, 115-7.
3. Wuerz RC, Milne LW, Eitel DR, Travers D, Gilboy N. Reliability and validity of a new five-level triage instrument. *Acad Emerg Med* 2000;7:236-42.
4. McGinn T, Wyer PC, Newman TB, Keitz S, Leipzig R, Guyatt G, for Evidence-Based Medicine Teaching Tips Working Group. Tips for learners of evidence-based medicine: 3. Measures of observer variability (kappa statistic). *CMAJ* 2004;171(11):1369-73.
5. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol* 1987;126:161-9.
6. Altman DG. *Practical statistics for medical students.* London: Chapman and Hall; 1991.
7. Fleiss JL. *Statistical methods for rates and proportions.* 2nd ed. New York: John Wiley and Sons; 1981. p. 218.
8. Byrt T. How good is that agreement? [letter]. *Epidemiology* 1996;7:561.

DOI:10.1503/cmaj.1041742

In their excellent overview of a common statistical measure of agreement, Thomas McGinn and colleagues[1] suggest in Table 1 that values for the kappa statistic range from 0 to 1. However, negative values of kappa are also possible.[2] Although unusual in practice, a negative kappa statistic results when agreement occurs less often than predicted by chance alone. This may indicate genuine disagreement, or it may reflect a problem in the application of a diagnostic test. Readers and researchers who encounter a negative kappa statistic should be aware of its implications, rather than blaming mathematical or typographic errors or computer "gnomes."

David N. Juurlink
Sunnybrook and Women's College Health Sciences Centre
Toronto, Ont.
Allan S. Detsky
Mount Sinai Hospital
Toronto, Ont.

References

1. McGinn T, Wyer PC, Newman TB, Keitz S, Leipzig R, Guyatt G, for Evidence-Based Medicine Teaching Tips Working Group. Tips for learners of evidence-based medicine: 3. Measures of observer variability (kappa statistic). *CMAJ* 2004;171(11):1369-73.
2. Kramer MS, Feinstein AR. Clinical biostatistics. LIV. The biostatistics of concordance. *Clin Pharmacol Ther* 1981;29:111-23.

DOI:10.1503/cmaj.1041744

As a teacher of basic skills in evidence-based medicine (EBM), I have appreciated the *CMAJ* articles that have been appearing in the EBM "tips" series. In particular, I was happy to see the discussion of the kappa statistic,[1] specifically the calculation of chance agreement (e.g., Table 3 in the article).

Unfortunately, discussions of kappa tend to focus on dichotomous variables, such as positive or negative results on mammography or the presence or absence of Murphy's sign. In cases of con-

		Triage nurse: High risk	Triage nurse: Lower risk	Total
Nurse administrator	High risk	9	2	11
	Lower risk	1	13	14
	Total	10	15	25

Fig. 1: Agreement table for triage nurse and nurse administrator at the author's hospital, using the emergency severity index[3] for nursing triage.

Table 1: Qualitative classification of kappa values*

Kappa value	Degree of agreement
≤ 0	None
0.01–0.20	Poor
0.21–0.40	Slight
0.41–0.60	Fair
0.61–0.80	Good
0.81–0.92	Very good
0.93–1.00	Excellent

*Adapted, with permission of the publisher, from Byrt T. How good is that agreement? [letter]. *Epidemiology* 1996;7:561.

tinuous variables, such as jugular venous pressure, how is agreement compared and kappa calculated?

G. Michael Allan
Department of Family Medicine
University of Alberta
Edmonton, Alta.

Reference

1. McGinn T, Wyer PC, Newman TB, Keitz S, Leipzig R, Guyatt G, for Evidence-Based Medicine Teaching Tips Working Group. Tips for learners of evidence-based medicine: 3. Measures of observer variability (kappa statistic). *CMAJ* 2004;171(11):1369-73.

DOI:10.1503/cmaj.1050010

[Two of the authors respond:]

Christopher Carpenter suggests another approach to teaching the kappa statistic: giving real data to the students and having them do the requisite calculations. Our approach takes them through the principles of the calculations, step by step. Calculating the kappa score from real data would be an excellent subsequent step for the most enthusiastic students. We encourage readers to consult the teachers' version of our article on the kappa score,[1] which has interactive components and data that may be helpful in understanding the concept of kappa.

Regarding the interpretation of kappa scores, Table 1 in both the teachers'[1] and learners'[2] versions of our article was based on a text by Sackett and colleagues[3] and not, as Carpenter correctly points out, the article by Maclure and associates.[4] Carpenter is also correct in noting that several different versions of this table are available in the literature. All have 3 basic categories: poor agreement, fair to good agreement, and very good to excellent agreement. In our view, further differentiating within these groups adds little to the practical clinical discussion.

Although we have never seen it in real life, David Juurlink and Allan Detsky correctly state that kappa can theoretically be less than 0 when agreement is poorer than chance. This is most likely to occur when both observers call almost every observation positive or almost every observation negative. In these circumstances, chance agreement would be close to zero and at times could be negative; determining chance-independent agreement (the phi statistic) may represent a better approach.[5]

Michael Allan asks about chance-corrected agreement when outcomes are categorical or continuous. One useful approach to this problem is the "weighted kappa," which gives maximal credit for full agreement, partial credit for partial agreement and no credit when disagreement is extreme. For example, in the case of ventilation-perfusion scans for the evaluation of pulmonary embolus, if both people reading a scan interpret the test result as normal (or both say there is intermediate or high probability of embolus), they get full credit for their agreement (weight of 1.0). If one reads the result as normal and the other as high probability, they get no credit (weight of 0). If one assessor classifies the scan as low probability and the other calls the same scan high probability or normal, they get partial credit (a weight of 0.75, for instance). Readers can find a more in-depth explanation and the details of how to calculate weighted kappa from the Web site of MedCalc Software (Mariakerke, Belgium; www.MedCalc.org).

Thomas McGinn
Department of Medicine
Division of General Internal Medicine
Mount Sinai Medical Center
New York, NY
Gordon Guyatt
Departments of Medicine and of Clinical Epidemiology and Biostatistics
McMaster University
Hamilton, Ont.

References

1. McGinn T, Wyer PC, Newman TB, Keitz S, Leipzig R, Guyatt G, for Evidence-Based Medicine Teaching Tips Working Group. Tips for teachers of evidence-based medicine: 3. Understanding and calculating kappa. *CMAJ* 2004;171(11):Online-1 to Online-9. Available: www.cmaj.ca/cgi/data/171/11/1369/DC1/1 (accessed 2005 May 16).
2. McGinn T, Wyer PC, Newman TB, Keitz S, Leipzig R, Guyatt G, for Evidence-Based Medicine Teaching Tips Working Group. Tips for learners of evidence-based medicine: 3. Measures of observer variability (kappa statistic). *CMAJ* 2004;171(11):1369-73.
3. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine*. 2nd ed. Boston: Little, Brown and Co; 1991. p. 30.
4. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol* 1987;126:161-9.
5. Meade MO, Cook RJ, Guyatt GH, Groll R, Kachura JR, Bédard M, et al. Interobserver variation in interpreting chest radiographs for the diagnosis of acute respiratory distress syndrome. *Am J Respir Crit Care Med* 2000;161:85-90.

Competing interests: None declared.

DOI:10.1503/cmaj.1050048

More on India's HIV-1 epidemic

Paul Arora and colleagues,[1] in their overview of the HIV epidemic in India, appropriately identify known risk groups and possible contributors to the problem, focusing on heterosexual sex, specifically involving commercial sex workers. However, there may also be a need to revisit the relative contributions of other major groups.

One factor, the population of men who have sex with men (MSM), is rarely discussed because homosexual acts are illegal in India. Recent work presented at the International AIDS Conference in Thailand indicates that MSM may contribute substantially to the epidemic in both the homosexual and heterosexual communities. A recent survey of more than 3000 men in Andhra Pradesh found that over half of all anal sex acts between MSM were unprotected; in addition, almost half of the MSM were married, and more than half had had sex with a woman within the previous 3 months (and most of these encounters were unprotected).[2,3] MSM in India represent a hard-to-service group, as the stigma of homosexuality and the responsibilities of marriage make disclosure difficult.

The effective interventions proposed by Arora and colleagues[1] are essential to slow the rate of infections. Accomplishing this goal will require countrywide recognition of the HIV epidemic, sex education (both within and outside the classroom) and access to free voluntary counselling and testing, all of which must reach commercial sex