source of major problems when comparisons are made between regions or countries.

We support the use of scoring systems in general terms and certainly do not wish to discourage the use of the injury severity score (or subsequently modified systems) for trauma audit. We suggest, however, that before any revision of the abbreviated injury score and the injury severity score observer variation should carefully be studied and measures adopted to minimise this problem for future users. Otherwise questions must remain over the precision of internal hospital audit, interhospital, and especially interregional or international, comparisons.

In conclusion, our study discloses some major problems with the methodology of the combined trauma and injury severity scoring system, but this does not undermine the need for national collection of trauma data by the United Kingdom major trauma outcome study. Rather, attention to the shortcomings we have identified should allow us—by placing the results in context—to utilise the results more appropriately and obtain the maximum value for these important national and international comparison.

We thank the participating members and in particular Mr J Sloan and Dr W Hulse for clinical input and help.

Participants in the study were: N Zoltie, M Clancy, J Hanson, and D Cartlidge (St James's Hospital, Leeds); M Williams, I Barlow, and J P Sloan (Leeds General Infirmary); A McGowan, A Gourdie, A Hawtin, and A K Marsden (Pinderfields Hospital, Wakefield); M Gibson and M Thirlway (York District Hospital); P Grout (Hull Royal Infirmary); and W Hulse (Harrogate District Hospital).

Appendix: statistical methods

Measures of assessing observer variation are the subject of much statistical debate. Fuller descriptions of many of the alternatives are provided elsewhere.[6,7] Most analyses in trials with multiple observers treat the resultant data as multiple two way comparisons, which was the method used for this study. For 15 observers this would result in 105 paired observations. If all the observers agreed, then the numbers of actual agreements would equal the total possible agreements —that is, the probability of agreement would be 1·0 (100%). For any number less than this the probability of agreement would equal the actual number of agreements divided by the total possible agreements, expressed as a percentage.

The most widely used coefficient of agreement in comparable studies is the κ statistic of Cohen.[11] There are, however, problems with the κ statistic. Cohen himself, for example, suggested several versions. As customarily used, κ measures the difference between observed agreement and the agreement that would be expected by chance in the same setting. A κ value of 0·4 generally represents reasonable agreement and 0·7 good agreement.

Our study discloses the limitations of this form of analysis. Firstly, the values recorded represented numerical assessments on non-numerical information. Secondly, the expected degree of agreement among the 15 observers varied greatly with the type of observation, ranging from the two way choice (major or minor trauma) to the actual score, for which the expected agreement was virtually nil. (In this setting the percentage of agreements was virtually equal to the κ statistic.) Finally, the κ statistic failed to take account of the clinical relevance of the data. Thus the data concerning agreement on major or minor trauma were impressive statistically but concealed an important factor—namely, that for six patients there was disagreement in up to half of comparisons on whether the patient had major or minor trauma and therefore on whether he or she should have been entered into a major trauma outcome study at all.

Brennan and Silman[7] have argued that for complex studies of observer variation more emphasis may have to be placed on raw data. We therefore present these in table A.

TABLE A—*Actual injury severity scores allocated to each patient by each observer*

Observer No	Case No*															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	9	14	29	17	34	17	38	13	29	4	29	25	4	16	25	45
2	9	13	29	17	22	14	45	10	29	4	25	34	9	25	8	50
3	9	13	29	4	27	22	38	10	29	4	16		9	50	18	54
4	9	20	10	16	27	12	45		29	4	75	34		50	4	75
5	9	13	38	25	22	27	38	9	29	4	29	26	9	34	19	51
6	13	22	29	20	22	19	29	13	34	4	25	18	9	50	9	34
7	9	14	42	26	34	14	20	10	20	4	25	27	26	50	26	54
8	9	19	17	10	17	12	30	1	20	4	17	34	1	26	10	50
9	9	13	75	25	27	17	33	9	20	4	29	41	9	25	13	41
10	9	13	38	17	41	18	33	9	20	4	25	26	9	25	5	50
11	13	13	26	13	17	12	29	13	29	4	75	41	9	50	18	42
12	9	27	75	16	22	22	34	10	29	9	29	19	9	50	10	50
13	9	13	27	16	22	17	54	13	29	4	75	35	4	50	14	50
14	9	8	20	16	17	22	34	10	29	4	16	18	9	16	10	50
15	9	13	22	16	27	17	54	10	29	4	25	34		50	9	75

*Graphs derived from table have been reordered, so that case numbers in table do not refer to those in derived graphs.

1 Commission on the Provision of Surgical Services. *Report of the working party on the management of patients with major injuries.* London: Royal College of Surgeons of England, 1988.
2 Phair IC, Barnes MR, Barton DJ, Allen MJ. Deaths following trauma: an audit of performance. *Ann R Coll Surg Engl* 1991;73:53-7.
3 Crawford R. Trauma audit: experience in north-east Scotland. *Br J Surg* 1991;78:1362-6.
4 Boyd CR, Tolson MA, Copes WS. Evaluating trauma care: the TRISS method. *J Trauma* 1987;27:370-7.
5 Yates DW, Woodford M, Hollis S. Preliminary analysis of the care of injured patients in 33 British hospitals: first report of the United Kingdom major trauma outcome study. *BMJ* 1992;305:737-40.
6 De Dombal FT, Softley A. IOIBD report No 1: observer variation in calculating indices of severity and activity in Crohn's disease. *Gut* 1987;28: 474-81.
7 Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ* 1992;304:1491-4.
8 Weiner N. Nonlinear prediction and dynamics. In: *Proceedings of 3rd Berkeley symposium on mathematical statistics and probability.* Vol 3. Berkeley: University of California, 1956:247.
9 Saiger GL. Observations on the probability of error in medical diagnosis. *Am J Intern Med* 1982;56:860-4.
10 Scheff TJ. Decision rules, types of error, and their consequences in medical diagnosis. *Behav Sci* 1963;8:97-107.
11 Cohen J. A coefficient of agreement for nominal scales. *Education and Psychological Measurement* 1960;20:37-46.

(Accepted 17 August 1993)

Correction

Short and long term prognosis of acute myocardial infarction since introduction of thrombolysis

An authors' error occurred in this paper by Robert Stevenson and colleagues (7 August, pp 349-52). In the results section of the abstract and the third paragraph of the subjects and methods section it is unclear how many patients were followed up after discharge from hospital. A total of 608 patients were studied, 89 died in hospital and 12 were lost to follow up after discharge. All 608 patients were followed up until hospital discharge or death in hospital and 507 were followed up after discharge from hospital.