When reviewers disagree

I B Pless

The manuscript decision process at *Injury Prevention*

Editors must satisfy two constituencies—authors and readers. (Sadly, they do not always overlap!) Readers care about the scientific quality of papers we publish and the manner in which they are written. Authors care about being accepted with the least possible hassle. To help us satisfy both, we rely on the advice of reviewers. Journals differ on how they use reviewers but our policy has remained quite consistent since the start. We ask three reviewers, one of whom is usually a member of the editorial board, to assess each paper along four dimensions: Significance, Appropriateness, Science, Writing. Each of these is rated on a three-point scale: high, medium, or low, along with a composite recommendation—accept as is (exceptionally rare), provisionally accept, provisionally reject, or reject (relatively common). Concerns are conveyed to the authors and summarized for the editor. Then the handling editor makes a final decision or sends the paper to a committee comprising the editor, deputy editors, statistical editor, and the book review editor. The committee step mostly happens when

reviewers' composite recommendations sharply disagree. I say "sharply" because when I review their comments there is often some inconsistency between the final recommendation and the strength of the comments.

An underlying question authors (and some readers) may ask is how well do reviewers actually agree? Based on a small study I conducted, the answer is—more often than you might expect. Before summarizing the results, I need to place the figures in context. Ideally, when examining agreement we should use a statistic like kappa that takes chance agreement into account. But kappa does not lend itself easily to the situation I have described with only three reviewers. Instead, I used a measure of simple agreement such that if one reviewer rated originality as "high" and another as "medium" and the third as "low" this would represent 0% agreement, whereas if all three chose high it would be 100%. To be quite clear, because it occurs so often, if two of the three chose the same rating, agreement would be 66%. But is this really much better than chance? How often would all

three reviewers agree if each flipped coins? The answer is not clear. One expert suggests 3.6% and another puts the number at about 10%. Statistically inclined readers may wish to send letters to the editor about what the correct estimate of chance agreement is for two of three reviewers.

So, what did we find? I randomly chose 20 recent papers and tabulated the responses. The average percent agreement, range, and standard deviation (SD) for each of the categories rated is shown in table 1.

My conclusion from these simple figures is that reviewers agree surprisingly well on the appropriateness of a paper for *Injury Prevention*. They agree a bit less well with respect to ratings of writing, science, and significance. The sharp differences represented by the zeros in the range rarely occur, and we almost never see the complete unanimity represented by the 100%. The standard deviations are all fairly narrow particularly with regard to the writing and overall ratings.

The take home message for authors is that our reviewers seem to agree more often than not on how well your paper meets the criteria we examine when deciding whether it should be published. The process is fair and reasonably objective, not whimsical. For readers, the message is that you can be assured that what you see has been scrutinized carefully and met what we believe to be an acceptable standard. Despite this system, we may occasionally get it wrong by publishing papers that are flawed or rejecting those that are far better than our reviewers judged. In this context we encourage readers to express their criticisms either by way of a letter or email, a follow up paper, or direct contact with the editor. All criticism will help us to continue to improve our Journal.

Injury Prevention 2006;12:211.

doi: 10.1136/ip.2006.090806

Correspondence to: Professor I B Pless, Editor; barry.pless@mcgill.ca

Table 1 Agreement of three or more reviewers on four key elements rated on a three-point scale

	Significance	Appropriate	Science	Writing	Overall
Mean	56%	69%	60%	62%	56%
Range	0–100	0–100	0–100	0–100	0–75
SD	30.7	28.5	33.5	24.0	24.3