Reviewer's report

Title: Outcome Measurement in Clinical Trials for Ulcerative Colitis: Toward Standardisation

Version: 3 Date: 2 March 2007

Reviewer: Karel Geboes

Reviewer's report:

General
The authors present a nice overview of the different outcome measurements available for ulcerative colitis and the problems observed when using these in clinical trials. While this is a very well written review, which accurately describes the complexity of the topic, I have still a few minor comments which could be addressed.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Discretionary Revisions (which the author can choose to ignore)

- In the introduction the authors indicate that they will propose potential solutions for the limitations of the current system. In fact, they propose only to address inter-observer variation in sigmoidoscopy.
- The authors indicate that there is no need to combine clinical, sigmoidoscopic, histopathology and other scores. I entirely agree for this point. The scores which are needed or which are optimal depend in fact upon the primary or secondary endpoints of a study. As stated by the authors "scoring systems are driven by the need to select appropriate patients and monitor response..." Usually primary endpoints are clinical response or remission, because this is in the interest of the drug. Long term remission is not usually an endpoint, although histological remission with disappearance of disease activity could be an important issue given the possible relationship between inflammation and carcinogenesis. It would therefore be interesting to find a few lines, in which the authors discuss the more common aim of the clinical trials which is clinical response or remission and lack of adverse events. This is not always the same as what can be measured by some of the indices or scoring systems which were developed sometimes on clinical grounds.
- The different scoring systems do not need an overlap because first of all, they measure different variables, and secondly, the changes induced in the bowel wall, during the course of the disease may be responsible for functional disturbances subsequently not due to disease activity. Endoscopy and histology for instance do not necessarily need to give similar results.

What next?: Accept after discretionary revisions

Level of interest: An article of outstanding merit and interest in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.