Journal of Crohn's and Colitis, 2016, S542–S547
doi:10.1093/ecco-jcc/jjv171
Article

Article

Central Endoscopy Reading in Inflammatory Bowel Diseases

Julián Panés,[a] Brian G. Feagan,[b] Fez Hussain,[c] Barrett G. Levesque,[b,d] Simon P. Travis[e]

[a]Department of Gastroenterology, Hospital Clínic de Barcelona, IDIBAPS, CIBERehd, Barcelona, Spain [b]Robarts Clinical Trials, Robarts Research Institute, University of Western Ontario, London, Ontario, Canada [c]Quintiles Immunology and Internal Medicine, Medical Strategy & Science, Rockville, MD, USA [d]Division of Gastroenterology, University of California San Diego, La Jolla, CA, USA [e]Translational Gastroenterology Unit, Oxford University Hospitals, Oxford, UK

Corresponding author: Julián Panés, MD, Department of Gastroenterology, Hospital Clínic de Barcelona, Villarroel 170, 08036 Barcelona, Spain. Tel: +34 932275418; Fax: +34 932279387; Email: jpanes@clinic.cat

Conference presentation: Presented at the 10th Congress of the European Crohn's and Colitis Organisation, February 18–21, 2015, Barcelona, Spain

Abstract

Endoscopic assessment of the presence and severity of endoscopic lesions has become an essential part of clinical trials in ulcerative colitis and Crohn's disease, for both patient eligibility and outcome measures. Variability in lesion interpretation between and within observers and the potential bias of local investigators in patient assessment have long been recognized. This variability can be reduced, although not completely removed, by independent evaluation of the examinations by experienced off-site (central) readers, properly trained in regard to lesion definition and identification, that should be removed from direct patient contact and blinded to any other clinical or study data. Adding endoscopic demonstration of active disease to eligibility criteria has the potential to reduce placebo response rates, whereas in outcome assessment it has the potential to provide a more precise estimation of the treatment effect, increasing the efficiency of the study. Central endoscopy reading is still at the beginning of its development, and the paradigms of central reading need refinement in terms of the number of readers, the process by which a final score is assigned, the selection and sequence of central readers, and the endoscopic indices of choice.

1. Background

Clinical trials rely on measures of disease activity for determining patient eligibility and drug efficacy. In Crohn's disease (CD) the CD Activity Index (CDAI) became the gold standard for clinical trials.[1] Although the CDAI may be adequate for measuring disease activity in some circumstances and proved useful for identification of effective therapies in CD, the index has some important limitations, including the fact that approximately 40% of the index is derived from three subjective criteria (diarrhoea, abdominal pain and sense of well-being), which account for 80% of the responsiveness to change.[2] A study looking at the relationship between CDAI and objective measures of disease activity, including biomarkers [C-reactive protein (CRP)] or endoscopic disease activity measured by the CD Endoscopic Index of Severity (CDEIS), found no correlation.[3] Recently, a sub-analysis of the SONIC study showed that at baseline 18% of patients with moderate to severe CD as measured by the CDAI (>220) had no endoscopic evidence of active CD. More importantly, after 26 weeks of treatment 47% of patients in clinical remission (CDAI <150) had still significant endoscopic lesions, whereas 35% of patients not in remission (CDAI >150) had a healed mucosa.[4] Taken together, these findings indicate that, for drug development programmes in CD, clinical assessment needs to be combined with an objective measure

of disease activity, such as endoscopy. These are the current recommendations of the regulatory bodies.

In ulcerative colitis (UC) coupling endoscopy to clinical assessment has been more common, as studies in UC have used composite clinical and endoscopic activity indices such as the Mayo Clinic score or the Sutherland Index, or have assessed endoscopic lesions using an *ad hoc* index such as the (modified) Baron score in conjunction with a clinical activity index.[5,6] This may have contributed to the perception that UC trials produce cleaner, less noisy results than CD and may be a preferred choice for initial testing of drugs in inflammatory bowel disease (IBD). Similar to CD, regulatory bodies now require demonstration of clinical efficacy and ability to heal mucosal lesions to approve new drugs for treatment of UC.

The problem with endoscopy, however, is that, just as for patient-reported outcomes, interpretation is subjective. Variability in the assessment of endoscopic activity for UC may be an issue, even among experienced investigators.[7] This is not unique to endoscopic assessment. Variability in lesion interpretation between (and within) observers and the potential bias of local investigators in patient assessment have long been recognized with other outcome measures such as radiographic imaging and histopathology. This variability can be reduced, but not removed, by independent evaluation of the images by experienced readers, properly trained in regard to lesion definition and identification.[8] The term 'central reading' is often used to mean that the interpretation of imaging findings is not, or not only, done by the local site reader (endoscopist at the clinical trial centre) but instead is supervised, amended or adjudicated by at least one off-site reader. The 'central reader' is expected to have less bias, or more expertise, than the site reader, although neither can be assumed and questions of training, re-validation, adjudication and totality of assessment (all or a random selection of videos) arise.[9] Blinding of the central reader, removed from direct patient contact, is now considered indispensable and required by regulatory agencies.

2. The need to use centrally read endoscopy in IBD

In contrast to study designs in other areas of medicine, such as rheumatology or neurology, in which objective eligibility and efficacy endpoints include imaging assessed by a central research team, initial lesion assessment in IBD trials has traditionally relied on the interpretation of endoscopic findings by local physicians. Whilst one may imagine that local endoscopists are best equipped to provide the most accurate score, there is increasing evidence to suggest otherwise. This was highlighted in a recent analysis of a study on mesalamine in UC, which, in contrast to previous studies of the same drug, showed no statistical difference between treatment and placebo.[10] A possible explanation for this unexpected result was established when the endoscopic videos were blindly re-assessed by a central reader. No less than 31% subjects did not have the minimum disease activity on endoscopy required for inclusion in the study. *Post hoc* analysis showed that had these subjects been excluded by central readers assessing subjects for eligibility, the study would have demonstrated statistically significant efficacy, with results similar to those of other trials of mesalamine.

There is also a tendency of local investigators to over-score baseline examinations in clinical trials of CD. In a study assessing the efficacy of certolizumab pegol for improving mucosal lesions in patients with active CD, the mean differences between local and centrally read CDEIS scores in baseline examinations was 4.1. The range of CDEIS is 0–44, but the mean CDEIS at baseline in this study was 14.5 ± 5.3. Local scores were systematically higher, whereas the mean differences in CDEIS scores between site and central readers at week 10 and week 54 (n = 33) were considerably smaller (1.0 for each time point), although these differences were not statistically significant.[11]

Recruitment of subjects without a minimum endoscopic severity into trials might be seen as the enthusiasm of investigators to offer their patients access to treatment with an investigational drug, but there is inherent variability in the assessment of endoscopic lesions. In a study performed outside the context of a clinical trial, using a library of videoendoscopies in patients with UC, demonstrated that there was 76% agreement for 'severe' activity, but only 27% agreement for a normal appearance and 37% for moderate severity.[7] Similarly, for CD, both the CDEIS and simplified endoscopic score for Crohn's Disease (SES-CD) have shown intraclass correlation coefficients for inter-reader reliability of 0.71 and 0.83. This reliability drops to 0.62 for assessment of global lesion severity, the most common sources of disagreement being interpretation of superficial ulceration, the definition of disease at ileocolic anastomosis, the assessment of anorectal lesions and grading the severity of a stenosis.[8]

Therefore, despite the availability and use of standardized scoring systems in UC and CD and appropriate training, multicentre studies still carry a risk that inter-observer variability will significantly affect data interpretation and sample size requirements. Central assessment of endoscopic scores helps ensure that each patient enrolled and monitored is assessed using precisely the same criteria, irrespective of local expertise. The implementation of central reading has already been shown to improve efficiency in drug development programmes. In earlier years, high placebo response rates presented a major hurdle to establishing drug efficacy. The first published trial (of etrolizumab for UC) in which central reading was conducted in real time to confirm eligibility resulted in remission rates of 0% in the placebo group at week 10.[12]

3. Paradigms of central reading

If the value of centrally read endoscopy in clinical trials is accepted, many questions arise about its effective execution. These questions include, but are not limited to:

- The number of central readers required
- The process by which the final score is determined
- Whether site reader assessment is considered or only central reads are used
- Whether all time points require the use of the same central reading algorithm
- Selection/sequence of central readers (e.g. the same reader reads both baseline and endpoint vs random selection from a central reader pool)
- The endoscopic index of choice.

Listed in Table 1 are seven options for the number of central readers, consideration of the site reader and the process by which a final score can be determined. Given the vast differences in complexity, cost and turnaround times of the different algorithms, the identification of the least complex and effective option is important. Paradigms 1, 2, 4, 6 and 7 carry the efficient (but expensive) approach of avoiding site reading and consigning the assessment of eligibility/efficacy solely to the central reader. If over-enrolment due to site reading is avoided and the findings of Feagan et al.[10] are confirmed, then a paradigm based entirely on central reading may turn out to be optimal.

Paradigms 3 and 5 involve the assessments of both the site and central readers, in order to attempt to achieve a higher degree of consensus between the two parties. This might get closer to the 'truth' of

Table 1. Paradigms of central and site reading in IBD trials.

Paradigm	Central readers (CR)	Site reader (SR)	Adjudication
1	1	None	–
2	1 except for suboptimal videos	None	CR1 and CR2 agree = final score CR1 and CR2 disagree = force consensus
3	Up to 2	1	CR1 and SR agree = final score CR1 and SR disagree = CR2 assess CR1 and CR2 agree = final score CR1 and CR2 disagree = force consensus
4	2	None	Use paradigm 2
5	Up to 3	1	CR1 and SR agree = final score CR1 and SR disagree = CR2 assess CR1 and CR2 agree = final score CR1 and CR2 disagree = CR3 assess Final score = majority of CR1, 2 and 3
6	Up to 3	None	CR1 and CR2 agree = final score CR1 and CR2 disagree = CR3 assess Final score = majority of CR1, 2 and 3
7	2 + adjudicating CR	None	CR1 and CR2 agree = final score CR1 and CR2 disagree = CR3 adjudicates

disease activity, since the quality of local endoscopic imaging might be superior to that of smaller, transmitted images read by a central reader and has both motivational and training value. However, involving both site and central reader demands resolution of any differences of interpretation. This implies an additional read by a second central reader. This second central reader would score the images with no knowledge of the site or first central reader's score and no knowledge of the subject's clinical information, although validated endoscopic indices for UC are not influenced by clinical information.[13] The advantage of this paradigm is that it provides an opportunity to resolve a disagreement between site and central reader assessment, while the disadvantage is that the paradigm involves another step in the read process, with an additional time allowance and cost. US regulatory guidance expects that many, if not most, imaging endpoints will be obtained with no reader knowledge of individual-level clinical data (Food and Drug Administration Draft guidance, March 2015). If multiple readers are utilized, then the adjudication process needs to be agreed and pre-specified. A degree of discordance can be expected when two readers assess an examination. Different adjudication designs are listed in Table 1. One approach is to employ no adjudication, using only the score of the second central reader, as in paradigm 7. Disagreements between two central readers may be resolved by forcing consensus, which can be done via a review meeting or by a group of additional readers. Alternatively, disagreement between two central readers may lead to assessment of the examination by a third central reader, with the final score being the majority score of central readers 1, 2 and 3. Should there be no majority, then consensus should be forced as described above. Additional work is needed to determine how paradigms of central reading can be optimized and made most efficient.

4. Regulatory environment

There is no specific guidance from the Food and Drug Administration or the European Medicines Agency on standards for use of endoscopy as an eligibility or efficacy measure in clinical trials in IBD. Both agencies have, however, produced generic guidance documents on standards for clinical trial imaging endpoints, as noted above.

Among the recommendations established by the Food and Drug Administration, some noteworthy statements include the following:

- A centralized image interpretation process is needed when image interpretation represents important components of trial eligibility or safety or efficacy endpoints, and these measurements are vulnerable to considerable variability among clinical sites
- Image interpretation should be blinded to knowledge of treatment and individual level clinical data
- Establish read process: number of readers, independence/consensus of reads, adjudication process, process of re-read
- Describe reader qualifications and training.

The European Medicines Agency recommends central reading when the on-site evaluation may be biased, but should not be presented as sole proof of efficacy, even though this mirrors routine clinical practice. The Food and Drug Administration also recommends that off-site evaluation should be performed by independent readers, that these readers should be blinded and unaware of the clinical context, and that central reading should be performed by two or more readers. However, these are expert, opinion-based recommendations, since the superiority of any particular reading paradigm has yet to be established.

5. Live discussion

JP: I would like to know your views on whether offsite endoscopy reading should be integrated in all IBD studies. Do we always need central reading or we can still develop trials based on clinical symptoms and biomarkers?

ST: I don't think that we need offsite endoscopy in all IBD studies. I do think that you need offsite endoscopy reading in all phase III studies. I think the argument for phase II is rather more refined because you have to balance cost against benefit. The message is that you should not trust the opinion of local investigators without some way of independent assessment, and central reading is a convenient way of confirming activity at trial entry. But I think that for some phase II studies which recruit a small defined population in a few centres, you can get away without central reading as long as you have a trained endoscopist with a track record in IBD scoring systems. In phase III studies including larger cohorts and higher number of centres with a diversity of investigators, I do think that you need some central quality control.

BL: I'd agree with the need for offsite centralised for all IBD clinical studies. I think that's the feedback that we are getting from regulatory agencies. The data that we showed earlier that placebo rates particularly can be decreased with that measure are convincing. I think additional data that you can see at this ECCO congress will show that central reading also for outcome can particularly affect the signal. And I think by decreasing the variants of the signal we are getting a more clear picture of efficacy. I think in addition we need to identify patients who have active disease and central reading helps confirming the presence of active disease. Then we can go into biomarker engagement, for example in phase II trials, and add another level of resolution.

FH: I don't think that there is any doubt that central reading makes endoscopy more robust in clinical trials. I think this is strong theoretical grounds and you know some early data which have been presented very nicely. So I think if you are doing a clinical trial then central reading will give you a more robust result with increased separation of interventions and will definitely decrease bias.

ST: We need to recognize the panel may be somewhat biased in that we have two people with a professional interest in central reading, and it would be interesting to hear the view of some industry leads as to whether it is worth paying.

FH: I think that is a fair point. I was going to qualify that the question may be do you need to do endoscopy? I don't think you always necessarily need to do endoscopy. I think it is a balance between cost, patient burden and value you are going to get from endoscopy. So in fact, in phase III trials we know we have to do endoscopy for CD and UC trials, but if you go to proof of concept trials we could distinguish between UC and CD. In UC performing a sigmoidoscopy is really easy, I mean the patient burden isn't too high, we are going to get some valuable information from this examination. CD may be different; if you are doing a small phase II study, particularly in a moderate to severe population who can take quite a long time to respond, then doing endoscopy may not justify the patient burden and the increase in complexity.

Dr Jean-Fred Colombel (Mount Sinai Hospital, New York): I'd like to make several points. The first point is that we don't know if the central reader is doing better than the site reader, since there is absolutely no data about that. Second point is for CD for the calculation of the scores SES-CD and CDEIS. You mentioned the study by Hébuterne et al.;[11] discordance between site and central reads in this study came from the experience of the endoscopists. The central readers were experienced senior readers, while the site readers were less experienced. So I think a very important point is about learning how to score; remember for instance the EXTEND study,[14] where some patients had an SES-CD or CDEIS not acceptable, because they were above the highest level that can be observed. It is mainly a matter of education.

ST: So you are talking about process and I think that is absolutely crucial. Julian touched on the process and FDA recommendations. Training is fundamental and it's not something we have routinely done in recruiting PIs to clinical trials.

JP: Jean-Fred, let me challenge that because you mean that the site readers learn how to score during the trial because there were differences at baseline and there were no differences at the time of the primary endpoint?

Dr Jean-Fred Colombel (Mount Sinai Hospital, New York): The point is that it is very clear that we don't have data about the experience of readers in reading CDEIS or SES-CD and there is a learning curve for this kind of score. I think it is particularly true for CD, more than for UC.

FH: I think if using site readers is part of the algorithm, it is really important to get consistency between the offsite centralised reading and site readers on how they are reading the endoscopy scores. Training is absolutely essential, even to the point of having test videos and requiring site readers to pass those test videos. Interestingly, if you look at studies where they have done correlations between the experience of endoscopists and results, there has not always been the correlation that you would expect. In other words an endoscopist with shorter experience performs almost as well as a very experienced endoscopist. I think the key point is actually to get people trained on something that is not intuitive. The major components of the SES-CD are not aspects that we all evaluate in clinical practice, so it is not surprising that central and site readers don't get the same scores, unless they have that training.

BL: I think I need to defend the bias point. I think in terms of central reading, I may be biased towards doing it, but I am probably more biased to including patients who have active rather than inactive disease in trials And definitely, we are still seeing 30% screen failure rates with central reading due to endoscopic disease not meeting inclusion criteria. The main difference between the central and the site reader may be the interest for helping the patient to obtain a new treatment. This may lead to bias in scoring, and this is why we should rely on the central rather than on the site reader. There may be differences between driving the car and being a passenger, but I still think there is an advantage when you are a passenger to be the one reading the map rather than the driver being distracted.

Dr Elmer Schabel (BfArm, Bonn, Germany and European Medicines Agency, London, UK): Isn't it also a question of trial design? The Tillotts trial[10] is a typical trial of a medicine that is only a new strength, it is a well-known substance, and the trial is placebo-controlled, in a mild to moderate population of patients with UC. So patients are not really that ill and since it is a placebo-controlled study the investigator tends to get their patients in if they are not so much ill. So the point is if there are trial designs with higher risk of bias for eligibility or efficacy assessment. One has to differentiate between these situations to determine the need for central reading of endoscopies.

BL: I think it brings up the important point that there is the reliability or reproducibility issue for selecting the population entering a study on one hand, and then on the other there is the potential for bias in assessment of efficacy, particularly when this would influence the inclusion of the patient into a maintenance phase. So, I think these are two important points.

FH: That is a really interesting question and that may be answered in the years to come when we learn more about central reading. But something that we have noticed is that when site investigators realized there is a central reading process implemented, they know they have someone watching them, and then the bias decreases, so you get much less bias across. If you look at the great original Tillotts study,[10] this was retrospective, so the local investigators did not know that there was a central reading process that would be conducted afterwards, but actually peoples' behaviour changes when they know there is a central reading process.

Dr Sharon O'Byrne (Genentech-Roche): From the sponsor side there is a little bit of discussion of the importance of central reading in phase II trials, but from our point of view the big decision to invest in a very large phase III trial, with huge costs, is based on a proof of concept phase II study. Human experiment is key to getting drugs that are actually going to make a difference to patients through the gate, and the quality of how that experiment is conducted which comes

down to proof of concept and dose ranging is of paramount importance. I think central reading has a huge role there. We have to be very objective in how we conduct those studies because we are often competing for funds in the portfolio and we won't get it unless the data is somewhat convincing, so I think that is another piece to consider here in favour of implementing central reading in phase II trials.

JP: In fact, when I posed that question initially I just had considered phase II because I absolutely see what you are saying: a key decision is taken based on the results of phase II, and you want really robust data on this part, that informs not only the decision to pursue the development of the drug, but other aspects of the design of a potential phase III trial, in terms of selection of doses, endpoints, time points, etc.

Dr Geert D'Haens (AMC, Amsterdam): Two trials will be presented in the ECCO congress that were basically negative when the site readers' scores were used, but that become positive using the central reading score, so I think this is a very essential point, this is phase II. I wanted to come back to Dr Colombel's comment about the training of the site readers and perhaps it is our responsibility; we haven't trained our IBD fellows in scoring IBD and we have just now, in Amsterdam, implemented this system to force them to do the scores. You know, we were never formally trained to calculate the scores, so it is a critical duty, also to ECCO, to teach our fellows to do that.

And then just a last comment as a central reader. The quality of what you get on your screen determines what you can see and score. So again there is a need for training the site investigators so they provide high-quality evaluable examinations. There is also a duty for the CROs to teach the centres how to record and how things should be done.

My question is, do you really think that the site investigators' scoring should be taken into account when assessing drug efficacy?

BL: I don't think for the primary outcome measure that the site reader should be taken into account. I think it is reasonable to do sensitivity analysis on this. But I think there are two issues: one, the bias likely in inclusion and in outcome; and two, there is a need for a system in place for validation of the outcome measure and I think that system is being built around central reading and I don't see it built around site reading. So I think for the purpose of the trial I would go with central reading.

JP: According to the data we have, for the outcome measure there is a deviation between central and site reader, but it is a 10% difference on both sides, and it is very coincident for the outcome measures. So whereas we understand that there may be a bias in evaluations for eligibility for reasons already commented, the data available may be not strongly supporting the need for central reading for outcome evaluation.

BL: I definitely agree, we need more data. I think it will show that it will make a difference for outcome measures as well, but I think that the data remains to be published. On that note as well, I still think that we do see, even as the trial goes on and new sites are involved, there is a learning process in the individual sites. Not to say that one is better than the other, but I think that one is potentially more reliable and less biased.

FH: What we have seen with a central reading process implemented is that a trained site reader has roughly 70–75 percent agreement for the completely blinded central reader, which again is a very efficient way of doing it. But again you have to train the site readers well, which comes back to what you were saying. This won't work well unless you train someone really robustly on the requirements.

JP: I think somehow we have addressed the first questions. I would like to ask Dr Hussain and Dr Levesque about where are we in the process of standardizing offsite endoscopy reading? What do we know about how many readers do we need, what would be the best process of adjudication? How do we certify the readers and how do we educate the readers? Are there any initiatives in place or are these necessary to implement?

BL: In terms of adjudication and what's the optimal process, I think there are different approaches we can take. I think we took advantage of looking at how other disciplines have done central reading or offsite reading and tried to apply that to endoscopy. I think endoscopy in some way is unique because of differences of image quality and resolution that are available to the central and the site reader. The other part of that is that I think in order to answer it accurately we need data. There is the trial data, we'll look at it with different adjudication algorithms and see how the outcomes are influenced.

FH: I agree with Dr Levesque. I think we are very far from being standardized in terms of the algorithms that we use for central reading. I think that answer will come after central reading is used in more and more trials and relate different processes to outcomes in different ways. Also in terms of standardization, we'll also go back to where we are with the scales. There has been a lot of work in CD with the SCS-CD. With ulcerative colitis there is still a long way to go. Most studies are using the traditional Mayo index. Dr Travis did some great work with the UCEIS. There is still a lot of work to be done with standardization and the descriptive in ulcerative colitis.

ST: I think the performance characteristics of the UCEIS need to be defined, but the simplicity of the UCEIS and the fact that the components are mutually exclusive actually helps the scoring. Responsiveness of the index still needs to be characterized in prospective studies.

Dr Roopal Thakkar (AbbVie): As sponsors, one thing that we continue to deal with regarding the sites is site frustration when they are told by the central reader that the patient is not eligible. I am curious to see what your current thoughts are on this and what constitutes an eligible patient. Adding endoscopy to the more classical clinical criteria for eligibility, the screen failure rate has gone from 25–30% to 40–65% and almost all of it is driven by central reviews. This has an additional negative effect on trial development because some sites decide they don't want to participate in this trial and move on to something else.

BL: We need valid entry criteria and to do that we are looking at trials with treatments of known efficacy and then look to Food and Drug Administration guidance and the need for patient-reported outcome measures together with an endoscopic measure for inclusion. We are in the process of evaluating how different cutoffs for those entry criteria would influence eligibility and outcomes. Then, I think there is also a process of education of the sites not being confrontational in terms of us vs you for better or worse. Here is a reliable method, let's work with you, educate you as we talked about and then use those valid cutoffs.

JP: Thank you very much. I think it was a really interesting critical analysis of the role of central reading. It has become clear that there are a lot of opportunities for improvement and for making trials most efficient with better use of these tools.

Conflict of interest

JP has received unrestricted grants from AbbVie and MSD and consulting or speaker fees from AbbVie, Arena Pharmaceuticals, Boehringer Ingelheim, BMS, Celltrion, Ferring, Galapagos, Genentech-Roche, Hospira, Janssen, MSD, Nutrition Science Partners, Pfizer, Robarts, Takeda and TiGenix.

BF is CEO of Robarts Clinical Trials Inc. and has received grants from AbbVie, Amgen, Astra Zeneca, Bristol-Myers Squibb, Roche/Genentech, JnJ/Janssen, Millennium, Pfizer, Receptos, Tillotts and UCB, and personal

fees from AbbVie, Actogenix, Akros, Albireo, Amgen, Astra Zeneca, Avaxia Biologics, Avir Pharma, Baxter Healthcare Corp, Biogen Idec, Boehringer Ingelheim, Bristol-Myers Squibb, Calypso Biotech, Celgene, Elan/Biogen, EnGene, Ferring Pharma, Roche/Genentech, GiCare Pharma, Gilead, Given Imaging, GSK, Ironwood Pharma, JnJ/Janssen, Kyowa Kakko Kirin Co. Lt., Lexicon, Lilly, Lycera Biotech, Merck, Mesoblast Pharma, Millennium, Nektar, Nestles, Novartis, NovoNordisk, Pfizer, Prometheus Therapeutics & Diagnostics, Protagonist, Receptos, Salix, Shire, Sigmoid Pharma, Synergy Pharma Inc., Takeda, Teva Pharma, TiGenix, Tillotts, UCB, Vertex Pharma, VHsquared Ltd, Warner-Chilcott, Wyeth, Zealand and Zyngenia.

FH is employed by Quintiles.

BL has received consulting fees from Nestle Health Sciences, Prometheus Labs, AbbVie, Takeda and Santarus and research salary support from Robarts Clinical Trials.

ST has received unrestricted grants from AbbVie, BMS, Cosmo Pharmaceuticals, Lilly and Merck, and acted as advisor to or lecturer for AbbVie, Boehringer Ingelheim, Celgene, Cosmo, Ferring, Genentech, Giuliani, GSK, Novo Nordisk, NPS, Pfizer, Proximagen, Quintiles, Takeda, Topivert and VHsquared. He conducted no industry consulting during his time as President of ECCO (2012–2014). He has received research support from IOIBD and the Norman Collison Foundation and acted as a central reader for Bioclinica and Quintiles.

Acknowledgments

All authors have made substantial contributions to the conception of this review, drafting the article (JP) or revising it critically for important intellectual content (BF, FH, BL, ST). The authors did not have any writing assistance. The authors thank Ms Gabriele Mayr for providing a first draft with live capture of the content of the ECCO session.

JP is supported in part by a grant SAF2012-33560 from Ministerio de Economía y Competitividad, Spain.

References

1. Thia K, Faubion WA Jr, Loftus EV Jr, *et al.* Short CDAI: development and validation of a shortened and simplified Crohn's disease activity index. *Inflamm Bowel Dis* 2011;**17**:105–11.
2. Best WR, Becktel JM, Singleton JW, *et al.* Development of a Crohn's disease activity index: National Cooperative Crohn's Disease Study. *Gastroenterology* 1976;**70**:439–44.
3. Modigliani R, Mary JY, Simon JF, *et al.* Clinical, biological, and endoscopic picture of attacks of Crohn's disease. Evolution on prednisolone. Groupe d'Etude Therapeutique des Affections Inflammatoires Digestives. *Gastroenterology* 1990;**98**:811–8.
4. Peyrin-Biroulet L, Reinisch W, Colombel JF, *et al.* Clinical disease activity, C-reactive protein normalisation and mucosal healing in Crohn's disease in the SONIC trial. *Gut* 2014;**63**:88–95.
5. D'Haens G, Sandborn WJ, Feagan BG, *et al.* A review of activity indices and efficacy end points for clinical trials of medical therapy in adults with ulcerative colitis. *Gastroenterology* 2007;**132**:763–86.
6. Levesque BG, Sandborn WJ, Ruel J, *et al.* Converging goals of treatment of inflammatory bowel disease from clinical trials and practice. *Gastroenterology* 2015;**148**:37–51.
7. Travis SP, Schnell D, Krzeski P, *et al.* Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS). *Gut* 2012;**61**:535–42.
8. Khanna R, Zou G, D'Haens G, *et al.* Reliability among central readers in the evaluation of endoscopic findings from patients with Crohn's disease. *Gut* 2015, in press. PMID: 25935574.
9. Gottlieb K, Travis S, Feagan B, *et al.* Central reading of endoscopy endpoints in inflammatory bowel disease trials. *Inflamm Bowel Dis* 2015;**21**:2475–82.
10. Feagan BG, Sandborn WJ, D'Haens G, *et al.* The role of centralized reading of endoscopy in a randomized controlled trial of mesalamine for ulcerative colitis. *Gastroenterology* 2013;**145**:149–57.
11. Hebuterne X, Lemann M, Bouhnik Y, *et al.* Endoscopic improvement of mucosal lesions in patients with moderate to severe ileocolonic Crohn's disease following treatment with certolizumab pegol. *Gut* 2013;**62**:201–8.
12. Vermeire S, O'Byrne S, Keir M, *et al.* Etrolizumab as induction therapy for ulcerative colitis: a randomised, controlled, phase 2 trial. *Lancet* 2014;**384**:309–18.
13. Travis SP, Schnell D, Feagan BG, *et al.* The impact of clinical information on the assessment of endoscopic activity: characteristics of the Ulcerative Colitis Endoscopic Index Of Severity [UCEIS]. *J Crohns Colitis* 2015;**9**:607–16.
14. Rutgeerts P, Van Assche G, Sandborn WJ, *et al.* Adalimumab induces and maintains mucosal healing in patients with Crohn's disease: data from the EXTEND trial. *Gastroenterology* 2012;**142**:1102–11.