Assessment of evidence quality in IBD guidance: The use and misuse of GRADE

Short Title: Evidence quality in IBD guidance

Authors: Professor Morris Gordon[1,2], Professor Gordon Guyatt[3]

1. School of Medicine, University of Central Lancashire, UK
2. Families Division, Blackpool victoria Hospital, UK
3. Department of Clinical Epidemiology and Biostatistics, McMaster University, Canada

No grant support

Abbreviations:

AGA – American Gastroenterological Association

ACG – American College of Gastroenterology

CD – Crohn's Disease

ESPGHAN – European society for Pediatric Gastroenterology, Hepatology and Nutrition

ECCO – European Crohn's and Colitis Organisation

GRADE – Grading of Recommendations Assessment, Development and Evaluation

IBD – Inflammatory Bowel Disease

NICE – National institute for Health and Care Excellence

UC – Ulcerative Colitis

Corresponding author:

Professor Morris Gordon

Room HA340,

Harrington building

University of Central Lancashire,

Preston,

Lancashire, PR1 2HE, UK

Email: mgordon@uclan.ac.uk

Phone: ++44 1772 895945

Disclosures: none relevant to declare.

Professor Gordon has previously received none restricted travel grants to attend internal scientific meetings from various companies, but received no other funding or performed any other roles for them. Since July 2019, Professor Gordon has made a personal undertaking to receive no future pharma funding for any activity, to comply with the enhanced requirements of Cochrane COI policies. Both Professor Gordon and Professor Guyatt work in the academic fields of synthesis and gain personal academic advantages from publication of such works.

Author Contributions:

Professor Gordon conceived this paper, led the writing and has approved the final version. Professor Guyatt has reviewed and rewritten the paper and approved the final version.

A core principle of evidence based medicine is that optimal clinical care must be grounded in systematic summaries of the best available evidence. Over 40 years, the methodology of systematic reviews has evolved sophisticated approaches to framing study questions, conducting comprehensive searches, evaluating risk of bias, and most recently and perhaps most crucially making judgements of the quality (also referred to as confidence or certainty) of the evidence. More recently – largely in the last 20 years – methodologists have developed and applied scientific standards to the process of moving from evidence to recommendations; that is, the process of creating clinical practice guidelines.

In this article, we will discuss this evolution in the particular context of inflammatory bowel disease. Although clinical leaders and stakeholder organisations within gastroenterology have embraced scientific standards for systematic reviews and guidelines, problematic practices that risk undermining the evidence based credentials of guidelines and undoing the progress made over the last decades still remain.

Evidence based medicine in Gastroenterology

The term 'Evidence Based Medicine' that first appeared in the literature in 1991[1], refers to a paradigm shift in approach to health professionals' use of evidence to guide their clinical practice, an approach grounded in systematic reviews of the best available evidence. Calls have been issued in this very journal for the field of gastroenterology to keep up with the trend as 'current systematic review articles seem to be published outside the field'[2]. It wasn't until 2002 that the first systematic review was published in the journal[3], although in 2018 there were 12 such articles, emphasising the core role they now play in the wider scholarly discourse.

Rigorous systematic reviews involve a number of steps, including formulating explicit eligibility criteria, postulating possible subgroup effects, conducting a comprehensive search, judging the risk of bias of individual eligible studies, and presenting coherent evidence summaries. As with all important methodologies within scholarly endeavours, execution at this stage can be capricious, with recent work suggesting that systematic reviews completed outside the Cochrane collaboration looking at identical contexts and questions will differ by up to 65% in the evidence they include[4]. If, however, these vital and generally well agreed steps are performed transparently, meticulously and without bias, this will set the appropriate stage for the next even more challenging step, which is deciding on the quality of the resulting body of evidence.

At the time systematic reviews captured the evidence summary science, multiple systems of judging quality of evidence existed, none of the remotely rigorous or satisfactory. Methodologists were aware of considerations such as inconsistency in research results, concerns about publication bias (suppression of negative trial results), and whether the magnitude of intervention benefits and harms is large enough to be important to patients. They did not, however, have an approach that would optimally integrate such issues.

GRADE contributing to Evidence Based Medicine and guidelines

The informal organization that became the GRADE working group began to meet in 2000[5] published its first guidance in the BMJ in 2004, and has continued to publish further detailed guidance (including a six-part series in the BMJ in 2008, and a thus far 22 part series in the Journal of Clinical

Epidemiology). There are now more than 110 organisations from 19 countries that have endorsed or use this approach to judging the quality of evidence bearing on clinical questions, and developing corresponding clinical practice guidelines, including key organisations within gastroenterology. There is no scholarly discourse that highlights any clear competing approach or identifies fundamental flaws in this approach that is now seen as the universal gold standard to assess quality of evidence and strength of recommendations.

A recently published example by one of the authors of this commentary highlights the GRADE approach to quality of evidence in systematic reviews and how it supports, but in no way mandates, specific recommendations for practice. In an updated Cochrane review on the use of 5-ASA agents for maintenance of post-surgical remission in Crohn's disease[6], the authors identified a statistically significant reduction in relapses when using 5-ASA agents compared with placebo. During a follow-up period of 12 to 72 months, 36% (131/361) of 5-ASA participants relapsed compared to 43% (160/369) of placebo participants (RR 0.83, 95% CI 0.72 to 0.96).

When considering a GRADE approach to rating the quality of the evidence specific to relapse prevention, as this was analysis of randomised controlled trials, the quality of the evidence begins as high in GRADE's four category rating (high, moderate, low, and very low quality evidence). Five categories of limitations can, however, results in rating down evidence quality. Risks of bias (item 1) for the trials in this comparison were low across key items; there was no evidence of publication bias (item 2); the populations and outcomes were appropriate suggesting no concerns of indirectness (item 3); and inconsistency (item 4) was low, no evidence of statistical heterogeneity. Thus, none of these four elements required rating down of quality. There were, however, relatively low numbers of events (291) and the smallest plausible effect (the upper boundary of the confidence interval, a 4% relative risk reduction), might not be important to patients. The authors therefore rated down the quality of the evidence by one level for imprecision, leaving us moderately confident of an important impact of 5-ASA in reducing recurrence following post-surgical remission in patients with Crohn's disease

This relatively robust evidence will be considered in a variety of international guidelines in the future. Despite the quality or certainty of the evidence from the review, guideline committees will not necessarily issue a strong clinical recommendation for 5-ASA. Following GRADE methodology, guideline developers will need to consider harms and burdens associated with the intervention, costs, feasibility, acceptability and equity – as well as alternative treatment options and the magnitude of effect. This last element may be particularly relevant, as the absolute risk reduction in recurrences over a period of 1-5 years is 7%, meaning that clinicians must treat 14 patients to prevent one clinical relapse over this time period[6]. This may lead those in a guideline process to consider that despite the moderate certainty of the evidence, the relative small magnitude of effect (weak recommendation).

This is the ultimate achievement of GRADE for clinicians and patients, with quality of evidence not necessarily implying strong recommendations or vice versa, but a requirement to nonetheless assess evidence in an objective manner in all cases. This balanced approach to considering not just the quality but wider context of evidence is at the heart of evidence based medicine and one that is now well reflected across growing guidance in IBD.

Current IBD guidance and quality of evidence according to GRADE

Evidence based treatment guidelines for IBD have become the increasing standard in the last decade. It is not surprising that whilst these began to be released around the time the GRADE process was being endorsed by key stakeholders, early guideline committees lacking familiarity in this approach did not use the tool. Instead, these used an approach to consider evidence quality which mostly focussed on individual study level assessments, but no method to consider the evidence base as a whole when making recommendations.

Table 1 presents the most recent publications on IBD from the major guideline producing societies across USA and Europe. It can be seen that whilst a number of these recent guidelines have fully embraced the GRADE process, there are some clear areas of divergence. One reason for this is understandable in that a society may be updating their processes to use GRADE. An example of this was the recently published ECCO adult Crohn's disease medical management guidelines[7], which is the first from the society to use GRADE.

A more alarming situation is the attribution of use of GRADE within a guideline process, when what authors are using is a misinterpreted or misappropriated version of GRADE or indeed committees simply not using GRADE at all. Three examples (UK National Institute for health and care excellence NICE CD guideline update[8], ESPGHAN UC[9] guidelines and ECCO surgical CD guidelines[10]) illustrate the problem.

The UK NICE guidelines[8] have a standard operating procedure aligned with GRADE[11]. Whilst there is no question within these procedures or their underpinning GRADE guidelines that authors rate the quality of an entire body of evidence for a given outcome, the authors present GRADE analysis of individual studies. This is truly a non sequitar: two GRADE domains, inconsistency and publication bias apply only to bodies of evidence; for two others, risk of bias and directness, individual studies can differ and GRADE provides guidance for making across study ratings; and for precision the evaluation is made on pooled estimates rather than individual studies. Both these guidelines link to external methodologies (NICE guideline policies[11] and Cochrane Handbook[12]) that accurately summarize GRADE guidance, yet neither follow the summaries presented there.

In the case of the recent ESPGHAN guidelines for Paediatric UC[9], it is unclear. There is mention of GRADING evidence, but whilst readers may infer alignment within this rhetoric, the methods involve scoring of individual studies much more aligned with earlier systems. Finally, the recent ECCO surgical CD guidelines[10] explicitly address and choose to not use GRADE. They state that 'due to the peculiarities of the surgical literature, appraisal of the systematically researched literature was conducted according to the Oxford methodology.' We believe this is again a non-sequitur. Whilst a given body of evidence may be relatively weaker than other areas in a field, using an out of date and out of consensus method for judging evidence does not solve this issue. This is akin to using a fractured lens to examine broken glass. It merely leaves readers even more unclear as to the nature of the evidence. As the example above showed, higher certainty evidence does not necessitate strong recommendations any more than lower quality evidence limits committees making stronger recommendations. Therefore, using a lower fidelity tool to reflect lower quality evidence offers no advantages to the guideline committee and further complicates the role of guideline developers and readers.

These fundamental mischaracterizations or misuses represent a serious matter. The purpose of GRADE quality of evidence ratings is to allow systematic reviewers to present to summarize the hugely complex issue of quality of evidence in a single rating for each outcome that the review has addressed and in turn present this evidence to guideline committees to use in a clear and transparent fashion. What evidence users need, and are entitled to, is quality ratings that apply to the entire body of evidence.

The overall impact of such inconsistent approaches are to substitute a careful systematic approach that includes assessment of inconsistency and indirectness and systematic consideration of between study differences with a former approach sometimes referred to as GOBSAT (good old boys sitting around a table). GOBSAT substitutes intuition for careful, systematic guidance-informed assessment, opening the process to judgments that flow from raters' conflicts of interest. When claiming use of GRADE, guidelines undermine the progress in understanding, using the status of the system that has almost universal acceptance internationally while potentially misleading unsuspecting clinicians. Similarly, when substituting or diverging from GRADE without true justification, clinicians may not necessarily be able to interpret the relative impact on the overall process or utility of the resulting guideline.

We are not suggesting deliberate motives to mislead – though authors may, consciously or subconsciously, giving more free play to their particular opinions in such approaches. Rather, this is complex area and such aberrations are understandable; irrespective of the motivation, the adverse impact is a serious matter. The examples given are recently published guidelines relevant to dozens of countries with a catchment of millions of IBD patients. It is vital that the responsible organizations rectify their practices in future guidelines.

Future directions in IBD and gastroenterology guidelines

There is a growing integration of GRADE approaches to IBD and wider Gastroenterology guidance. Many of the recent publications were the first time these societies have integrated GRADE approaches and we would strongly advise they follow the institute of medicine's guidance and employ a GRADE methodologist, preferably as co-chair, participating in the guideline (as indeed some organizations have adopted as standard practice)[14], as well as following the guidance from the GRADE working group on applying or using GRADE[17].

This movement towards homogenous approaches of evidence collection, evaluation, and presentation may offer new opportunities for collaboration that remain unexploited. In key areas, numerous groups associated – and not associated – with specialty societies are simultaneously conducting comprehensive searches (including for unpublished studies), undertaking risk of bias ratings, extracting data, contacting authors, and conducting GRADE quality of evidence assessments. A collaborative approach to these activities, with the various guideline developers using this single shared data set to inform their own technical reviews and guidelines would be far more efficient manner that the current duplication. Such collaboration will ensure the most rigorous reviews, consistency of evidence, and enhanced expert methodological support through sharing of limited resources.

We strongly encourage key opinion leaders and senior professionals within these societies to discuss the possibility of this sort of collaboration, but this must be built on an alignment to GRADE as standard.

Table 1. Current Use of GRADE in guidelines for IBD

Guideline title (reference)	Society and Location	Date	Methodology	Compliant with GRADE
American Gastroenterological Association (AGA) Institute Guideline on Management of Mild-to-Moderate Ulcerative Colitis[13]	AGA, USA	2019	**CDG author reports**	YES
			Guideline states 'Although the quality of evidence was a key factor in determining the strength of the recommendations (Table 3), the panel also considered the balance between benefit and harm of interventions, patients' values and preferences, and resource utilization.'	
			Judgement of methodology presented	
			GRADE methodology was used to prepare the background information for the guideline and the accompanying technical review and evidence profile. Citation of society publication describing application of GRADE to quality assessment of evidence. Evidence summaries with quality and strength of recommendation ratings are presented and consistent with GRADE[14]	
ACG Clinical Guideline: Management of Crohn's Disease in Adults[15]	American College of Gastroenterology, USA	2018	**CDG author reports**	YES
			To evaluate the level of evidence and strength of recommendations, we used the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system.	
			Judgement of methodology presented	
			Statements are presented with both the quality of evidence rating and strength of recommendation, in line with GRADE, however underpinning evidence profiles are not presented.	
Guidelines on the management of inflammatory bowel disease in adults[16]	British Society of Gastroenterology, UK	2019	**CDG author reports**	YES
			GRADE of evidence for each statement 'considered study type, risk of bias, inconsistency, Indirectness, imprecision, publication bias, effect size, plausible confounding variables, and dose response gradient if applicable. The strength of recommendation was assessed based on considerations of desirable and undesirable	

			anticipated effects, the certainty of the evidence of effects, any important uncertainty about or variability in how much people value the outcome, whether the balance of these effects favours the intervention or comparison, the acceptability of intervention to key stakeholders, and feasibility of intervention implementation.' **Judgement of methodology presented** Significant details of methodology, however, despite some appendices being mentioned, evidence profiles are not presented. Detailed discussion of efficacy and safety data in main text.	
Crohn's disease: management Evidence review for post-surgical maintenance of remission[8]	National Institute for Health and Care Excellence, UK	2019	**CDG author reports** The guideline cites a source methodology document, stating that: 'The quality is then summarised by individual study and, if using the GRADE approach, by outcome'[11] **Judgement of methodology presented** In this evidence review, GRADE was used to assess individual studies. Authors note that due to this, inconsistency cannot be ranked and so this is not considered across the evidence base. Imprecision rating considered only on Confidence intervals, no consideration of event numbers or other elements. Publication bias not considered. Risk of bias of individual studies, not outcome level. Evidence profiles are presented, but as stated are not compliant with GRADE.	NO
ECCO Guidelines on Therapeutics in Crohn's Disease: Medical Treatment[7]	European Crohn's and Colitis Organisation, Europe	2020	**CDG author reports** Based on the GRADE workflow and selection of committee members considered knowledge of the GRADE methodology. All panellists received adequate training in GRADE before starting the process. **Judgement of methodology presented** Online appendices with detailed information of the GRADE process and judgements and GRADE evidence profiles.	YES
ECCO Guidelines on Therapeutics in Crohn's Disease: Surgical Treatment[10]	European Crohn's and Colitis Organisation, Europe	2020	**CDG author reports** 'These Guidelines abide by the GRADE methodology in terms of framing clinically relevant questions to draw evidence-based statements and recommendations. However, due to the peculiarities of the surgical literature, appraisal of the systematically researched literature was conducted according to the Oxford methodology (grading from evidence level [EL] 1: systematic review of randomised	NO

			controlled trials to EL 5: expert opinion).’	
			Judgement of methodology presented	
			No use of GRADE in synthesis or evidence profiling. No clear justification for not using GRADE, except the suggestion the evidence base was poor	
Management of Paediatric Ulcerative Colitis, Part 1: Ambulatory Care—An Evidence-based Guideline From ECCO and ESPGHAN[9]	European Society of Paediatric Gastroenterology, hepatology and Nutrition and European society of Crohn’s and Colitis, Europe		**CDG author reports**	NO
			‘grading of evidence according to the Newcastle-Ottawa assessment scales for case-control and cohort studies (19) and according to the Cochrane Handbook for clinical trials (20) (Supplemental Table 2: tables of evidence with grading’	
			Judgement of methodology presented	
			No overall outcome or statement level GRADE ranking of evidence given. Whilst GRADING is stated, this does not appear to be an explicit statement on the use of GRADE, as individual study level judgements are made in the evidence profiles. Additionally, these judgements are not consistent not consistent with Cochrane handbook for clinical trials.	

References

1. Guyatt GH. Evidence-based medicine. ACP J Club March/April 1991:A-16.
2. Kumar A. Systematic reviews and gastroenterology. Gastroenterology. 1995; 108;4: A22 DOI 10.1016/0016-5085(95)22754-7
3. Dulai G, Guha S, Kahn K, et al. Preoperative prevalence of Barrett's esophagus in esophageal adenocarcinoma: A systematic review. Gastroenterology 2002; 122:26–33
4. Useem J, Brennan A, LaValley M, et al. Systematic Differences between Cochrane and Non-Cochrane Meta-Analyses on the Same Topic: A Matched Pair Analysis. PLoS ONE 2015;10(12): e0144980. doi:10.1371/journal.pone.0144980
5. GRADE Working Group. GRADE home page. Available at: http://www.gradeworkinggroup.org/#pub Accessed on 13th May 2020.
6. Gjuladin-Hellon T, Gordon M, Iheozor-Ejiofor Z, et al.Oral 5-aminosalicylic acid for maintenance of surgically-induced remission in Crohn's disease. Cochrane Database Syst Rev. 2019; 6:CD008414. doi: 10.1002/14651858.CD008414.pub3
7. Torres J, Bonovas S, Doherty G, et al, on behalf of the European Crohn's and Colitis Organisation [ECCO], ECCO Guidelines on Therapeutics in Crohn's Disease: Medical Treatment, Journal of Crohn's and Colitis 2020;14:4–22
8. NICE Crohn's guideline update. Available at: https://www.nice.org.uk/guidance/ng129/evidence/may-2019-evidence-review-6777581005?tab=evidence Accessed on 13th May 2020
9. Turner D, Ruemmele F, Orlanski-Meyer O, et al. Management of Paediatric Ulcerative Colitis, Part 1: Ambulatory Care—An Evidence-based Guideline From European Crohn's and Colitis Organization and European Society of Paediatric Gastroenterology, Hepatology and Nutrition. Journal of Pediatric Gastroenterology Nutrition. 2018;67:257-291. doi: 10.1097/MPG.0000000000002035.
10. Adamina M, Bonovas S, Raine T, et al, on behalf of European Crohn's and Colitis Organisation [ECCO], ECCO Guidelines on Therapeutics in Crohn's Disease: Surgical Treatment, Journal of Crohn's and Colitis, 2020;14:155–168
11. Developing NICE guidelines Available at: https://www.nice.org.uk/media/default/about/what-we-do/our-programmes/developing-nice-guidelines-the-manual.pdf Accessed on 13th May 2020
12. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
13. AGA. AGA Clinical Practice Guidelines on the Management of Mild-to-Moderate Ulcerative Colitis Available at: https://www.gastrojournal.org/article/S0016-5085(18)35407-6/fulltext Accessed on 13th May 2020
14. Sultan S, Falck-Ytter, Y, Inadomi, J.M. The AGA institute process for developing clinical practice guidelines part one: grading the evidence. Clin Gastroenterol Hepatol. 2013; 11: 329–332
15. Lichtenstein G, Loftus E, Isaacs K, F et al. ACG Clinical Guideline: Management of Crohn's Disease in Adults. American Journal of Gastroenterology 2018;113:481–517 doi: 10.1038/ajg.2018.27

16. BSG IBD guidelines Available at: https://www.bsg.org.uk/resource/bsg-consensus-guidelines-ibd-in-adults.html Accessed on 13th May 2020
17. Grade working Group. Criteria for applying or using GRADE. Available at: https://gradeworkinggroup.org/docs/Criteria_for_using_GRADE_2016-04-05.pdf Accessed on 20th June 2020