BSG 2024 IBD guidelines protocol (standard operating procedures)

BMJ Open Gastroenterology

BSG 2024 IBD guidelines protocol (standard operating procedures)

Ana-Maria Darie ,[1] Vasiliki Sinopoulou,[2] Verma Ajay,[3] Klaartje Bel Kok,[4] Kamal V Patel ,[5] Jimmy Limdi,[6] Naila Arebi,[7] Philip Smith,[8] Shahida Din ,[9] Said Din,[10] Matthew Shale,[11] Sreedhar Subramanian,[12] Polychronis Pavlidis,[13] Rachel Cooney ,[14] Dennis McGonagle,[15] Newton A C S Wong,[16] Gordon W Moran,[17] Morris Gordon[2]

To cite: Darie A-M, Sinopoulou V, Ajay V, *et al*. BSG 2024 IBD guidelines protocol (standard operating procedures). *BMJ Open Gastro* 2023;**10**:e001067. doi:10.1136/bmjgast-2022-001067

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjgast-2022-001067).

GWM and MG contributed equally.
A-MD and VS contributed equally.

A-MD and VS are joint first authors.
GWM and MG are joint senior authors.

Received 17 November 2022
Accepted 2 January 2023

For numbered affiliations see end of article.

Correspondence to
Dr Ana-Maria Darie;
ana-maria.darie@nuh.nhs.uk

ABSTRACT

Introduction In the past 5 years, there have been several advances in the management of inflammatory bowel disease (IBD). We aim for a new guideline to update the most recent guideline published in 2019. We present the prospective operating procedure and technical summary protocol in the manuscript.

Methods 'Grading of Recommendations Assessment, Development and Evaluation' (GRADE) will be followed in the development of the guideline, approach as laid out in the GRADE handbook, supported by the WHO. The guideline development group is formed by a variety of disciplines, across both primary and secondary care that took part in an online Delphi process and split into key areas. A final consensus list of thematic questions within a 'patient, intervention, comparison, outcome' format has been produced and agreed in the final phase of the Delphi process.

There will be a detailed technical evidence review with source data including systematic reviews appraised with AMSATAR 2 tool (Assessment of multiple systematic reviews), randomised controlled trial data that will be judged for risk of bias with the Cochrane tool and observational studies for safety concerns assessed through the Robins-I tool. Based on the available evidence, some of the recommendations will be based on GRADE while others will be best practice statements.

A full Delphi process will be used to make recommendations using online response systems.

This set of procedures has been approved by the Clinical Services and Standards Committee, the British Society of Gastroenterology executive board and aligned with IBD UK standards.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Since publication of the last British Society of Gastroenterology (BSG) guideline in 2019, substantial advances have been made in the management of the inflammatory bowel disease (IBD).

WHAT THIS STUDY ADDS

⇒ The prospective publishing of this document is part of that process of systematic guideline production. This manuscript describes the prospectively agreed methods and operating procedures that will be followed to produce the IBD BSG new guidelines, also how the guideline development group (GDG) was created and the process related to organisation, planning and training of GDG. This technical summary protocol describes the process in generating the patient, intervention, comparison, outcome thematic questions. This set of procedures has been approved by the Clinical Services and Standards Committee, the BSG executive board and aligned with IBD UK standards.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ We aim for this guideline to update the recent guidelines and to cover all aspects of IBD including but not exclusive to Crohn's disease, ulcerative colitis, indeterminate colitis, IBD unclassified and primary sclerosing cholangitis-related IBD.

INTRODUCTION

In the past 5 years, there have been several advances in the management of inflammatory bowel disease (IBD). To this effect, the British Society of Gastroenterology (BSG) Clinical Services and Standards Committee (CSSC) has commissioned a new guideline for the management of IBD. This is aimed to update the most recent guideline published in 2019.[1]

This document describes the prospectively agreed methods and operating procedures that will be followed to produce these new guidelines. The final guideline will contain the official recommendations of the BSG on all aspects of IBD care. This set of procedures has been approved by the CSSC, the BSG executive board and aligned with IBD UK standards. No funding has been received from any outside organisation—commercial or otherwise—to produce this document, with some support provided for members time as part of their employment at public higher educational institutions or within

their roles as National Health Service funded health professionals.

We aim for this guideline to cover all aspects of IBD including but not exclusive to Crohn's disease (CD), ulcerative colitis, indeterminate colitis, IBD unclassified and primary sclerosing cholangitis-related IBD. While primarily designed for use within the UK, the guideline will no doubt have international scope and utility for both healthcare professionals and people living with IBD and, therefore, is presented in a full systematic and transparent fashion. The prospective publishing of this document is part of that process of systematic guideline production.

METHODS

The development of this guideline follows the procedures of the 'Grading of Recommendations Assessment, Development and Evaluation' (GRADE) approach as laid out in the GRADE handbook, supported by the WHO handbook for guideline development.[2] The guideline development group (GDG) used the GIN-McMaster guideline development checklist,[3] an 18-point process map to support the steps in a GRADE compliant guideline development process.[4]

Organisation, planning and training

In April 2021, the BSG appointed a content and field expert as a guideline chair through a competitive process. To meet the best evidence guidance for such leadership, a non-voting GRADE and synthesis methodologist was appointed as co-chair.[5] Administrative support will be provided from both host higher education institutions of the co-chairs and access to a Cochrane and NICE expert information specialist will be arranged through these institutions. The chairs proposed their approach to the BSG and editor of the journal Gut and received approval. A timeline for the process was prepared with core milestones.

A plan for a mixed approach to training and support will be made. A planned hierarchy of involvement will consist of the core leadership team and the rest of the GDG. GDG members will support technical reviewing as well as guideline recommendation voting, with others supporting voting alone. The core leadership team will be offered bespoke GRADE training workshops through a collaboration with Professor Schünemann and Dr Miranda Lamgendam at the Department Epidemiology and Data Science, Amsterdam University Medical Centers, University of Amsterdam, Netherlands.

The wider GDG will have additional resources made available supplemented with drop-in sessions to allow bespoke advice and education.

GDG membership

The GDG was formed in September 2021 following a general call, followed by targeted invitations through BSG media channels. This was to ensure a balance of professionals from a variety of disciplines, across both primary and secondary care. These included the following: gastroenterology, gastrointestinal surgery, paediatric gastroenterology, radiology, dermatology, rheumatology, pathology, pharmacy, clinical psychology, dietetics, specialist IBD nursing, charities including Crohn's & Colitis UK and contributions from service user representatives. A national call out from Crohn's and Colitis UK facilitated identification of people living with IBD who joined the GDG. The group met online to outline the roles and operating procedures.

The GDG members will be directed to general guidance in line with general guidance on established national processes from National Institute for Health and Care Excellence 2012 (NICE).[6] This will confirm that as members of the team, all GDG members agree to be coauthors of the full guideline, to maintain the confidentiality of open discussion and debate within the guideline process, as well as to the confidentiality of the content of the guideline prior to publication.

Members will be asked to declare all conflicts of interest in line with BSG guidelines. The BSG will appoint a conflicts of interest chair to oversee the process.

GDG priority setting and identifying target audience

The GDG were invited to take part in an online Delphi process in February 2022. To manage this, the GDG was split into key areas, as per the 2019 BSG guidelines.[1] These subgroups included: diagnostics and classification, ulcerative colitis, CD, special considerations and service delivery. The identity and content of the subgroups will be changed based on the nature of the thematic questions identified and brought forward for evidence synthesis.

The subgroups were given a list of all core thematic topics and specific thematic questions from the previous BSG IBD guidelines.[1]

The results were analysed to identify topics for the new guidelines based on consideration of two major factors:

- Need for update due to emerging or new evidence not apparent when the last guideline was published.
- Priority for stakeholders for specific clinical or patient factors.

A final consensus list of thematic questions within a 'patient, intervention, comparison, outcome' (PICO) format has been produced and agreed in the final phase of the Delphi process (online supplemental file 2). An example PICO is given below in box 1.

Final subgroups will be formed based on the thematic question grouping. Each of the subgroups will have a chair who will represent the subgroup in the core

Box 1 Example PICO to be used in the present BSG IBD guideline methodology.

P=chronic active ulcerative colitis.
I=mesalazine PO.
C=placebo.
O=corticosteroid-free remission.

leadership team. They, together with collaborators from within each of the subgroups, will work closely with the core leadership team to produce technical reviews for the thematic questions within their subgroup. These would form the basis for consideration of recommendations within subgroup discussion prior to wider GDG voting. Each subgroup leads will be offered bespoke GRADE training. Both core leadership and subgroup meetings will be held throughout the process.

Discussions were held within the wider GDG, within the CSSC and with the editor of Gut regarding the most appropriate outputs from the process for our target audiences. The following was decided:

- The prospective publishing of a guideline operating procedure and technical summary protocol in an open access journal (this manuscript).
- The publishing of a succinct main guideline that summarises key recommendations, the certainty of underpinning evidence and the strength of the recommendations all within the main published journal output.
- An accompanying larger detailed technical evidence review that will include all the underpinning primary evidence, secondary synthesis quality and analysis data, also published within the main journal output.
- A final user-focused guideline that summarises the key recommendations and the evidence justification of these recommendations in a manner that is informed, coproduced and for the intended use by people living with IBD.

These outputs will offer systematic, high-quality and high utility output for all of our audiences.

PICO thematic question generation

The generation of questions has been guided by the GRADE guidelines.[7] As this is an update guideline, the process of prioritisation has been based on the thematic and PICO questions from the previous guideline. This allows the process to focus on generation of novel questions in areas that have emerged and refinement of existing questions.

The generation of new questions is a key component of the Delphi prioritisation, asking for any new or uncovered areas to be presented in free text, with justification of the specific question. Analysis of these results has allowed new candidate questions to be considered by each of the GDG subgroups. With new questions added in a standard PICO format.

Key areas of focus for refinement of all questions have been considered by the subgroups, guided by the framing question guidance.[7] These core elements of refinement around PICO questions and their specific application have been presented in draft form to the GDG and all feedback considered, with the final list below:

- UK regulatory approval is mandatory for any pharmacological or similar intervention to be considered. However, UK NICE approval is not considered.
- Multiple treatment arms will be considered. To allow consideration of non-placebo comparators and standard therapies, network meta-analysis will be deployed in key targeted areas, as decided by the GDG subgroups and when sufficient volume of homogeneous studies is likely to be included.
- For studies of induction of remission, the key outcomes will be clinical remission (including corticosteroid-free), clinical response, endoscopic remission, endoscopic response, biochemical remission and biochemical response, withdrawals due to adverse events or serious adverse events.
- The magnitude of effect of any intervention (against placebo or a comparator) for induction of clinical remission of importance will be defined at 10%.
- For studies on maintenance of remission, the key outcomes will be clinical relapse, loss of clinical response, endoscopic and biochemical relapse, withdrawals due to adverse events and serious adverse events. Maintenance of remission studies will only be considered if patients were in clinical or endoscopic remission at baseline. Preplanned sensitivity analysis will include studies with patients in response, but this will not be the core analysis.
- The magnitude of effect of any intervention (against placebo or comparator) for maintenance of clinical remission of importance will be defined at 10%.
- For all such analysis, planned subgroup analysis will include whether patients received concomitant therapies (more than 50% in a group) with immunosuppression, corticosteroids or biological therapy (not anti-inflammatory, antibiotics or other classes of therapy), whether patients were biologically naïve and in the case of maintenance studies, as stated above, whether they are in a state of remission or response at baseline.
- For studies not targeting induction or maintenance of remission, outcomes to be considered will include quality of life (using study defined scales), morbidity (including escalation of therapy or hospitalisation), abdominal pain and fatigue.
- For all safety outcomes, no predefined levels for magnitude of effect of safety concerns will be made, as this will be dependent on the specific effect. All events with randomised controlled trials (RCTs) will be recorded and presented descriptively, as well as through meta-analysis. However, specific reporting will be made for all the following adverse events in both trials and observational studies: mortality, malignancy, sepsis, organ targeted inflammatory effects (pancreatitis, nephritis, hepatitis) and thromboembolic events will be specifically reported for all interventions.
- Therapy delivered in primary, secondary or tertiary care, as well as self-administered will be considered but setting will be described in the technical summary to allow any clarifying statements to be made regarding the context of evidence.

GRADE recommendations and good practice/research statements

The GDG determined which PICOs are likely to conducive to evidence synthesis and GRADE assessment of outcomes. This included all forms of evidence (observational and RCT). The GDG does not believe there is enough evidence to make a GRADE assessment for certain areas of interest. The GDG will have a number of options following GRADE guideline production methodology.[8]

The first will be to make GRADE recommendations recognising these limitations, justifying the factors that led to this (these may be strong or conditional). The second option is when the GDG will find it difficult or impossible to formally summarise and GRADE the evidence, but guideline panel members are confident that there is unequivocal benefit or harm from the proposed PICO response, a good practice statement may, under strict criteria, instead.[9]

At this stage, the GDG has identified PICOs that are likely to fall under the good practice statement context, due to lack of studies. If through the technical review, evidence becomes apparent that allows synthesis, this will be used, but if not, statements will be produced after GRADE recommendations for all such outcomes and delphi agreement sought as per such GRADE recommendations.

Both elements will be clearly labelled throughout the end guideline.

Considering importance of outcomes and interventions, values, preference or utilities to consumers

The list of these elements that have guided the PICO formation has been presented to our partners in Crohn's & Colitis UK and people living with IBD members. The key principle has been to balance volume of evidence and completeness with the utility to people living with IBD and stakeholders while aligning the BSG recommendations to those produced by the IBD standards.

Evidence selection

Criteria for considering studies for this review

Types of studies

A three-phase approach will be employed for searching for studies.

1. Systematic reviews will be included. Potential reviews will be those completed since 2018[1] (the previous guideline). Potential reviews will be assessed using the AMSTAR 2 tool (A MeaSurement Tool to Assess systematic Reviews) .[10] When multiple reviews are found on the same topic, the highest rated review will be included. Subgroups will determine if the AMSTAR rated reviews are of sufficient quality to be included. If they are not up to date (completed in the last 12 months) or any additional studies found within the wider search (or rejected systematic reviews) will be added and the meta-analysis ran again to update the results. If risk of bias or GRADE ratings are not included, these will be completed using the approach below. Cochrane systematic reviews will be prioritised for inclusion, but with the same stipulations of updating analyses to include all studies.
2. RCTs that assess the interventions of interest will be included for consideration. Only studies that use conventional dose regimens in at least one treatment arm will be considered for inclusion. Phase 1 studies will not be included. Studies must be randomised; quasi-randomised or non-randomised studies will not be included. These studies will be extracted and analysed as per the methods below and where appropriate, combined with the systematic reviews above.
3. For safety concerns only, observational studies will be included.

Types of participants

Trials enrolling adult participants (>16 years of age) with IBD as defined by conventional clinical, radiological, endoscopic or histological criteria will be considered for inclusion. The disease activity is mention in table 1.

Types of interventions

Trials using the following interventions for induction or maintenance of remission delivered by any route will be included. The types of interventions for ulcerative colitis are mention in table 2 and the types of intervention for CD are mention in table 3.

For studies focusing on symptom management, outside of induction or maintenance of remission, any intervention will be included.

Types of outcome measures

- For studies of induction of remission, the key outcomes will be clinical remission (including corticosteroid-free), clinical response, endoscopic remission, endoscopic response, biochemical remission and biochemical response, withdrawals due to adverse events or serious adverse events. Histological remission data will be included as a secondary outcome, as it is expected that data will be sporadic in RCTs for this outcome.
- For studies on maintenance of remission, the key outcomes will be clinical relapse, loss of clinical response, endoscopic and biochemical relapse, withdrawals due to adverse events and serious adverse events. Histological remission data will be included as a secondary outcome, as it is expected that data will be sporadic in RCTs for this outcome.

Such studies will only be considered if patients were in clinical or endoscopic remission at baseline. Preplanned sensitivity analysis will include studies with patients in response, but this will not be the core analysis.

Search methods for identification of studies

Electronic searches

A Cochrane and NICE expert Information Specialist will search the following sources from 1 January 2018 until date:

Table 1 Disease activity will be defined through the following modalities

	Crohn's disease	UC	Pouchitis
Active disease (clinical scores)	Crohn's Disease Activity Index (CDAI)[24] >150 Harvey Bradshaw Index (HBI)[25] >5	MAYO* score for Ulcerative Colitis Index[26] >2	Pouchitis Disease Activity Index[27] >4
Active disease (endoscopic scores)	Simple Endoscopic Score for Crohn's Disease (SES CD)[28] >3 Crohn's Disease Index of Severity (CDEIS)[29] >3	MAYO score >1 Ulcerative Colitis Endoscopic Index of Severity (UCEIS)[30] >2	
Active disease (biochemical scores)	C reactive protein (CRP)[31] >5 mg/L, faecal calprotectin (FCP)[31] >250 µg/g	CRP >5 mg/L, FCP >250 µg/g	
Remission (clinical scores)	CDAI <150, HBI <4	MAYO score <2 and no individual subscore >1	
Remission (endoscopic scores)	SES CD: 0–2 CDEIS <3	UCEIS <1 MAYO score ≤1	
Remission (biochemical scores)	CRP <5, FCP <250	CRP <5, FCP <250	
Response (clinical scores)	CDAI-70, CDAI-100[32] HBI drop by 3 or more points[33]	Reduction of baseline MAYO score by ≥3 points and a decrease of 30% from the baseline score with a decrease of a least one point on the rectal bleeding subscale or an absolute rectal bleeding score of 0 or 1 (21)	
Response (endoscopic scores)	50% drop in the SES-CD[34]	Decrease in MAYO endoscopic score ≥1 grade or a decrease in UCEIS ≥2 points[35]	
Response (biochemical scores)	This is not defined		

Note: For studies defining these indices with other methods, they will not be included in main analysis, but will be included in wider sensitivity analysis.
MAYO score: The Mayo Clinic score is a tool used to evaluate the severity of ulcerative colitis (UC) in patients. It is based on a patient's clinical symptoms, endoscopic findings, and laboratory results. The score ranges from 0 to 12, with higher scores indicating more severe disease. The Mayo score takes into account factors such as the extent of colon involvement, the presence of ulcers, and the degree of inflammation seen on endoscopy. It is used to determine the severity of the disease, to monitor the response to treatment, and to guide decisions about therapy.
*MAYO and HBI score denotes derivations of the scores like modified MAYO, partial MAYO score or rectal bleeding and stool frequency UC patient reported outcomes and abdominal pain and stool frequency CD patient-reported outcomes. Specific definitions for active disease for these derivations are not described here.
UC, ulcerative colitis.

- Cochrane Central Register of Controlled Trials (CENTRAL) via Cochrane Library.
- Embase via Ovid SP.
- MEDLINE via Ovid SP.

For each of the searches, a strategy was devised and included in online supplemental file 1.

Searching other resources

All studies included in the previous BSG guideline will be considered.

Manual searches of reference lists from potentially relevant publications will be performed to ensure that studies that may have been missed by the computer-assisted search are identified.

We will also perform hand searches of the conference proceedings of major gastroenterology meetings (eg, digestive disease week; the annual meeting of the European Crohn's and Colitis Organisation; and United European Gastroenterology Week) from 2019 to 2024.

Evidence synthesis

Selection of studies

At least two pairs of authors will independently assess all publications identified by the search strategy based on the inclusion and exclusion criteria. Any disagreement among authors will be resolved through discussion until consensus is reached. If there is a disagreement, a third author will be consulted.

Data extraction and management

Two pairs of reviewers will independently extract data. Any disagreements will be resolved by discussion with a third author. A standard data extraction form will be

Table 2 Types of intervention for ulcerative colitis

Type	Subclass
5-aminosalicylic acids (5-ASA)	Mesalazine
	Sulphasalazine
Corticosteroids	Hydrocortisone
	Methylprednisolone
	Prednisolone
	Budesonide Multi matrix system (MMX)
	Beclomethasone dipropionate
Antibiotics	
Probiotics	
Synbiotics	
Thiopurine	Azathioprine Mercaptopurine Thioguanine
Methotrexate	
Calcineurin inhibitors	Ciclosporin
	Tacrolimus
Advanced medical therapies	
Biologics-TNF	Adalimumab
	Infliximab (intravenous and subcutaneous)
	Golimumab
Biologics-anti integrin	Vedolizumab (intravenous and subcutaneous)
Cytokine inhibitors	Ustekinumab
	Risankizumab
Janus kinase (JAK) inhibitors	Tofacitinb
	Filgotinib
	Upadacitinib
S1P modulators	Ozanimod
Dietary interventions	Specialist or exclusion diets
Faecal transplantation	
Surgical therapies	
Alternative therapies	

Table 3 Types of intervention for Crohn's disease

Intervention	
5-aminosalicylic acid (5-ASA)	Mesalazine
	Sulphasalazine
Corticosteroids	Hydrocortisone
	Methylprednisolone
	Prednisolone
	Budesonide-ileal release
Antibiotics	
Probiotics	
Synbiotics	
Thiopurine	Azathioprine Mercaptopurine Thioguanine
Methotrexate	
Calcineurin inhibitors	Ciclosporin Tacrolimus
Advanced medical therapies	
Biologics-TNF	Adalimumab
	Infliximab (intravenous and subcutaneous)
	Golimumab
Biologics anti-integrin	Vedolizumab (intravenous and subcutaneous)
Cytokine inhibitors	Ustekinumab
	Risankizumab
Janus kinases (JAK) inhibitors	Tofacitinb
	Filgotinib
	Upadacitinib
S1P modulators	Ozanimod
Dietary interventions	Specialist or exclusion diets
Surgical therapies	
Haematological treatments	
Alternative therapies	

When available, data on biosimilars will be grouped with that of the originator drug

used. The following data will be retrieved from included studies:

1. General information (title, journal, year, publication, type).
2. Participant characteristics (disease activity at inclusion; number and percentage of advanced medical therapy naive vs advanced medical therapy experienced patients; concomitant corticosteroid usage; number and percentage of patients who previously underwent surgery for the treatment of CD; disease duration; luminal or fistulising disease).
3. Primary and secondary outcomes.
4. Risk of bias information.

Assessment of risk of bias in included studies

The methodological quality of each included study will be independently assessed by two pairs of authors using the Cochrane risk of bias tool.[11] Factors to be assessed include:

1. Random sequence generation.
2. Allocation concealment.
3. Blinding of participants, outcome assessors and investigators.
4. Incomplete outcome data.

5. Selective outcome reporting.
6. Other potential sources of bias.

The studies will be judged to be of low, high and unclear risk bias based on these factors. Any disagreements regarding risk of bias assessment between the two authors will be resolved through discussion until consensus is reached. If consensus cannot be reached, a third author will be consulted to resolve the disagreement.

Measures of treatment effect

For the dichotomous outcomes, the treatment effect will be expressed as risk ratios with corresponding 95% CIs. For continuous outcomes, the treatment effect will be expressed as mean differences (MD) with 95% CIs. However, if the studies assessed the same continuous outcome differently, we will estimate the treatment effect using the standardised MD (SMD). We will present SMDs as SD units and interpret them as follows: 0.2 represents a small effect, 0.5 a moderate effect and 0.8 a large effect.

Unit of analysis issues

The participant will be the unit of analysis. For studies comparing more than two intervention groups, we will make multiple pair-wise comparisons between all possible pairs of intervention groups. To avoid double counts, we will divide shared intervention groups evenly among the comparisons. For dichotomous outcomes, we will divide both the number of events and the total number of participants. For continuous outcomes, we will only divide the total number of participants, and leave the means and SD unchanged.

We will include cross-over studies, but we will only pool their data if they were reported separately before and after cross-over, and we will only use pre-cross-over data. In the case of cluster-RCTs, we only use study data if the trial authors had used appropriate statistical methods in taking the clustering effect into account.

Dealing with missing data

The intention-to-treat principle (ie, all patients lost to follow-up are considered treatment failures) will be applied in the case of missing outcome data.

Assessment of heterogeneity

We will scrutinise studies to ensure that they are clinically homogeneous in terms of participants, interventions, comparators and outcomes. To test for statistical heterogeneity, we will use a χ^2 test. A $p<0.1$ gives an indication of the presence of heterogeneity. Inconsistency will be quantified and represented by the I^2 statistic. We will interpret the thresholds as follows.[12]

0%–40%: might not be important.
30%–60%: may represent moderate heterogeneity.
50%–90%; may represent substantial heterogeneity.
75%–100%: considerable heterogeneity.

We will examine possible explanations for heterogeneity when sufficient data are available, including factors such as participant characteristics (eg, age, sex), condition severity, healthcare system and country.

In the case of considerable statistical heterogeneity, we will investigate whether this can be explained on clinical grounds or risk of bias, in which case, we will aim to conduct sensitivity analyses. If no reasons for statistical heterogeneity are found, we will present the results narratively and in detail.

Evaluating the assumption underlying network meta-analysis

We will use a design-based decomposition of Cochran Q for assessing the homogeneity on the whole network, the homogeneity within designs and the homogeneity/consistency between designs. This approach also allows for an assessment of the consistency assumption after detaching the effect of single designs. Inconsistency will be located using the net heat plot.[13]

All analyses will be run with R statistical package (R Development Core Team) and the netmeta library.[14]

Assessment of reporting biases

To assess the presence of small-study effects in the network meta-analysis, we will use a funnel plot that accounts for the fact that each set of studies estimates a different summary effect.[15] We will focus on the comparisons of all active treatments against placebo, which may be more prone to small-study effects.

Publication bias will be investigated using funnel plots for each meta-analysis when there are at least four studies. Funnel plot asymmetry will be tested using rank a correlation test when there are at least 10 studies.[16]

Data synthesis

Direct comparisons of treatment effects

We will conduct pairwise meta-analyses by synthesising studies that compare the same interventions using a random-effects model.[17] For dichotomous outcomes, results will be expressed as a pooled RR with 95% CI. Where continuous scales of measurement are used to assess the effects of treatment, the MD will be used or the SMD if different scales have been used.

Adverse effects will be tabulated and assessed with descriptive techniques, as they are likely to be different for the various agents used. Where possible, the pooled risk difference (RD) with 95% CI will be calculated for each adverse effect, either compared with no treatment or another agent.

Separate analyses will be conducted based on whether the patient population at the time of study entry was biological naive or exposed; and had a disease duration of less than or equal to 10 years, or more than 10 years.

Network meta-analysis

Network meta-analysis is a method of synthesising information from a network of trials addressing the same questions but involving different interventions. Joint analysis of data within a network framework allows novel inferences on treatment comparisons that have not been previously addressed directly in any studies, and it increases precision for comparisons with few data.[18–22]

We will perform network meta-analysis in STATA (StataCorp: Release V.14.2) using the graph-theoretical method.[21]

Subgroup analysis and investigation of heterogeneity

The presence of heterogeneity among studies will be assessed using the χ^2 test (a p<0.10 will be regarded as statistically significant) and the I^2 statistic.[23] Data will not be pooled for meta-analysis if a high degree of heterogeneity is observed (eg, I^2>75%). Sensitivity analyses will be conducted as appropriate to investigate heterogeneity.

In the presence of statistical and clinical heterogeneity, and if sufficient studies are available, we will consider the following potential sources of heterogeneity:

1. Mistakes and inconsistencies in data extraction and entry.
2. Population.
3. Intervention (ie, dose, frequency or route).
4. Risk of bias.
5. Funding source.
6. Study design (ie, duration of treatment, duration of follow-up and number of participants).
7. Date of publication.

If we find significant inconsistency, we will investigate possible sources. Specifically, we will investigate the distribution of prespecified clinical and methodological variables that we suspect may be potential sources of either heterogeneity or inconsistency in each comparison specific group of trials. If sufficient studies are available we will consider the following potential sources of heterogeneity and network inconsistency:

1. Different drug doses and routes of administration.
2. Patients with exposure to immunosuppressive drugs (ie, thiopurines, methotrexate).
3. Patients with more severe disease activity.
4. Patients who are biological-naive.
5. Disease location and extent.
6. Studies that used clinical and objective measures to determine patient eligibility (C reactive protein, faecal calprotectin, endoscopy).

Sensitivity analysis

If possible, we will perform the following sensitivity analyses to examine the impact of the following variables on the pooled effect estimate:

1. Random-effects versus fixed-effect modelling.
2. Studies with low risk of bias versus studies with unclear or high risk of bias.
3. Relevant lost to follow-up (studies with >25% lost to follow-up).
4. Trials with a sample size of less than 50 participants.
5. Studies for maintenance where patients in response were included to be added.
6. Inclusion of studies defining disease activity, remission or relapse with none described or none validated approaches.

GRADE judgements of certainty

We will present summary of findings tables for all comparisons, which include our primary outcomes.

Based on risk of bias, inconsistency, imprecision, indirectness and publication bias, we will grade the certainty of the evidence for each outcome as high, moderate, low or very low (described below). We will justify all decisions to downgrade the certainty of studies using footnotes, and we will make comments to aid the reader's understanding of the review where necessary.

GRADE Working Group grades of evidence:

- High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.
- Moderate certainty: we are moderately confident in the effect estimate; the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
- Low certainty: our confidence in the effect estimate is limited; the true effect may be substantially different from the estimate of the effect.
- Very low certainty: we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of effect

Development of recommendations and strength of recommendations

An evidence technical review meeting will be run to allow the GDG members to discuss and explore the technical summaries and GRADE recommendations for each of the PICO outcomes. Particular areas of focus will be those with little to no evidence and the almost certainly lower quality safety data.

For areas where no observational or higher evidence or sufficient to allow appropriate GRADE judgements was found, a small working group will be convened to produce proposed good practice statements. These will be produced based on the previous BSG IBD guideline statements with amendments as needed.

Consensus process

A full Delphi process will be used to make recommendations using online response systems.

In round 1, the GDG will be asked to make recommendations for each of the PICOs. This will be open-ended and will allow the members to make any recommendations they wish. Each of the interventions mentioned in the protocol will be specifically noted. Members will also be asked to explicitly state if they believe no recommendation should be made based on the evidence available and if so, open-text responses for a potential good practice statement will be encouraged. Finally, the good practice statements for none GRADE PICOs will be presented for open text comment and amendments.

The core leadership team will collate and analyse the data and organise the PICOs submitted for round 2. In round 2, the GDG members will be presented with the consensus recommendations and statements and asked for each if they wish to make a free-text amendment. If

they do not wish to amend, they will be asked to judge the recommendation as strong-conditional or as a good practice statement.

At this point members will also be asked to consider key and specific research themes and recommendations and provide free-text mention of these.

Before round 3, a virtual or face-to-face round table will be run to allow discussion of areas of convergence and specific divergence. This will focus on the following key areas:

- Novel recommendations when compared with previous guidelines.
- Recommendations that do not align with current NICE guidance.
- Safety recommendations.
- Areas of conditional or no recommendation proposed.
- Research recommendations

Further round tables may be indicated, depending on the nature and depth of discussions.

Then, a final round 3 Delphi survey will be sent with the proposed statements and recommendations, with the strength associated with each. Members will be asked to agree or disagree for each item.

This will allow a final recommendation list to include the level of agreement within the team. Any item that has agreement below 75% will not be included and as needed further round table and rounds of Delphi will be offered for these items until agreement can reach this minimum agreement threshold.

Writing of the guideline

The lead representatives from each subgroup together with the co-chairs of the GDG will form a writing group to prepare the full guidelines in line with the proposals stated in this protocol, with a succinct guideline with the recommendations and a separate but complimentary full technical review summary paper. These will be prepared in line with the appropriate Gut journal guidelines.

The full document will be sent to the GDG for internal peer review and amendment before going to the wider BSG executive for review. All indicated amendments will be made prior to the submission of the final guideline.

In parallel the core leadership team will work with CCUK (Crohn's and Colitis UK organization) representatives and people living with IBD to prepare a user-focused guideline.

Author affiliations
[1]Gastroenterology, Nottingham University Hospitals NHS Trust, Nottingham, UK
[2]School of Medicine, University of Central Lancashire, Preston, UK
[3]Digestive Disease, Kettering General Hospital NHS Foundation Trust, Kettering, UK
[4]Department of Gastroenterology, Barts Health NHS Trust, London, UK
[5]Department of Gastroenterology, St George's University Hospitals NHS Foundation Trust, London, UK
[6]Gastroenterology, Pennine Acute Hospitals NHS Trust, Manchester, UK
[7]Department of Inflammatory Bowel Disease, St Mark's Hospital, Imperial College London, London, UK
[8]Gastroenterology, University Hospitals NHS Foundation Trust, Liverpool, UK
[9]Gastroenterology, NHS Lothian Edinburgh, Western General Hospital, Clydebank, UK
[10]Gastroenterology, Derby Teaching Hospitals, NHS Foundation Trust, Derby, UK
[11]Gastroenterology, Queen's Medical Centre Nottingham University Hospital NHS Trust, Nottingham, UK
[12]Gastroenterology, Cambridge University Hospital NHS Foundation Trust, Cambridge, UK
[13]Gastroenterology, School of Immunology and Microbial Sciences, King's College London, London, UK
[14]Department of Gastroenterology, University Hospitals Birmingham NHS Trust, Birmingham, UK
[15]Department of Rheumatology, Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK
[16]Department of Cellular Pathology, Southmead Hospital, Bristol, UK
[17]Gastroenterology, NIHR Nottingham Biomedical Research Centre at Nottingham University Hospitals, Nottingham, UK

Collaborators Guideline Development Group (GDG): Nurulamin Noor, Marietta Iacucci, Ankur Srivastava, Keith Bodger, Abhishek Ray, Daniel Gaya, Mohammed Nabil Quraishi, Simon Borg-Bartolo, Tim Raine, Aamir Mohammed Saifuddin, Misha Kabir, Krishna Shah, Deborah Morris, Sailish Honap, Shellie Jean Radford, Mostafa Afifi, Paul Knight, Jeffrey Butterworth, Dharmaraj Durai, John N Gordon, Jude Misson, Victoria Jennings, Suudheer Kumar Vuyyuru, Mairi McLean, Cameron Braddy-Green, Ronit Das, Ella Mozdiak, Ailsa Hart, Jonathan Blackwell, Tun Gloria, Matthew Brookes, Uma Selvarajah, Laith Alrubaiy, Aditi Kumar, James Alexander, Emma Nowell, Ben Disney, Tanya Monaghan, Gill Townson, Antonia Churchhouse, Lucy Hicks, Tom Butler, Brianna Cook, Eleanor Liu, Paolo Giuffrida, Vida Cairnes, Nick Kennedy, Tony Tham, Pearl Avery, Nidhi Sagar, Glyn Scott, Christian Selinger, Lisa Whitley, Phil Harvey, Madhoor Ramdeen, Chris Lamb, Thomas Pinkney, Janindra Warusavitarne, Hughes Sarah Jane, Simon Lal, Leena Sinha, Jessica Chadwick, Rachel Ainley, Ruth Wakeman, Richard Hansen, Yeop Intan, Gasparetto Marco, Devadason David, Priya Narula, Gabriele de Marco, Savage Laura, Gohil Sapna, Sarah Cripps, Mark Follows, Christopher Clarke, Stuart Taylor, Roger Feakins, Maurice Loughrey discussed the results and comment on the manuscript.

Contributors GWM is the lead author that conceived the presented idea, wrote the manuscript together with MG and both were in charge of overall direction and planning. VS verified the analytical methods. A-MD drafted the manuscript and designed the tables. Coauthors: VA, KBK, KVP, JL, NA, ShD, PS, SaD, SM, SS, PP, RC, DM and NACSW agreed the review questions, approved the protocol and contributed to the final version of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests 'Yes, there are competing interests for one or more authors'.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

Author note Ana-Maria Darie and Vasiliki Sinopoulou are first authors and the joint senior authors are: Gordon W. Moran andGordon Morris

ORCID iDs
Ana-Maria Darie http://orcid.org/0000-0001-7896-5134
Kamal V Patel http://orcid.org/0000-0003-2611-4260
Shahida Din http://orcid.org/0000-0003-2855-3400

Rachel Cooney http://orcid.org/0000-0003-3710-157X

REFERENCES

1 Lamb CA, Kennedy NA, Raine T, *et al*. British Society of gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. *Gut* 2019;68(Suppl 3):s1–106.
2 World Health Organization. *WHO handbook for guideline development*. 2014.
3 McMaster. GIN-mcmaster guideline development checklist. n.d. Available: https://cebgrade.mcmaster.ca/guidelinechecklistonline.html
4 Schünemann HJ, Wiercioch W, Etxeandia I, *et al*. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ* 2014;186:E123–42.
5 Fretheim A, Schünemann HJ, Oxman AD. Improving the use of research evidence in Guideline development: 3. group composition and consultation process. *Health Res Policy Syst* 2006;4:15.
6 Excellence NIoHac. The guidelines manual: appendix A - agreements and advice for guideline development group members. 2012. Available: https://www.nice.org.uk/process/pmg6/resources/the-guidelines-manual-appendix-a-agreements-and-advice-for-guideline-development-group-members-pdf-3304370657221
7 Guyatt GH, Oxman AD, Kunz R, *et al*. Grade guidelines: 2. framing the question and deciding on important outcomes. *J Clin Epidemiol* 2011;64:395–400.
8 Zeng L, Brignardello-Petersen R, Hultcrantz M, *et al*. Grade guidelines 32: grade offers guidance on choosing targets of grade certainty of evidence ratings. *J Clin Epidemiol* 2021;137:163–75.
9 Guyatt GH, Alonso-Coello P, Schünemann HJ, *et al*. Guideline panels should seldom make good practice statements: guidance from the grade working group. *J Clin Epidemiol* 2016;80:3–7.
10 Shea BJ, Reeves BC, Wells G, *et al*. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;358:j4008.
11 Higgins JPT, Altman DG, Gøtzsche PC, *et al*. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
12 Flemyng E, Dwan K, Moore TH, *et al*. Risk of bias 2 in Cochrane reviews: a phased approach for the introduction of new methodology. *Cochrane Database Syst Rev* 2020;10:ED000148.
13 Krahn U, Binder H, König J. A graphical tool for locating inconsistency in network meta-analyses. *BMC Med Res Methodol* 2013;13:35.
14 Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. *BMC Med Res Methodol* 2015;15:58.
15 Chaimani A. Accounting for baseline differences in meta-analysis. *Evid Based Ment Health* 2015;18:23–6.
16 Sterne JAC, Sutton AJ, Ioannidis JPA, *et al*. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:d4002bmj.d4002.
17 DerSimonian R, Laird N. Meta-Analysis in clinical trials. *Control Clin Trials* 1986;7:177–88.
18 Caldwell DM. An overview of conducting systematic reviews with network meta-analysis. *Syst Rev* 2014;3:109.
19 Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004;23:3105–24.
20 Salanti G, Higgins JPT, Ades AE, *et al*. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008;17:279–301.
21 Rücker G. Network meta-analysis, electrical networks and graph theory. *Res Synth Methods* 2012;3:312–24.
22 Rücker G, Schwarzer G. Reduce dimension or reduce weights? comparing two approaches to multi-arm studies in network meta-analysis. *Stat Med* 2014;33:4353–69.
23 Higgins JPT, Thompson SG, Deeks JJ, *et al*. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–60.
24 Best WR, Becktel JM, Singleton JW, *et al*. Development of a Crohn's disease activity index. *Gastroenterology* 1976;70:439–44.
25 Harvey RF, Bradshaw JM. A simple index of crohn's-disease activity. *Lancet* 1980;1:8167.
26 Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. *N Engl J Med* 1987;317:1625–9.
27 Irvine EJ. Usual therapy improves perianal Crohn's disease as measured by a new disease activity index. McMaster IBD Study Group. *J Clin Gastroenterol* 1995;20:27–32.
28 Daperno M, D'Haens G, Van Assche G, *et al*. Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD. *Gastrointest Endosc* 2004;60:505–12.
29 Mary JY, Modigliani R. Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective multicentre study. Groupe d'etudes thérapeutiques des affections inflammatoires Du tube digestif (GETAID). *Gut* 1989;30:983–9.
30 Travis SPL, 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.
31 Mosli MH, Zou G, Garg SK, *et al*. C-reactive protein, fecal calprotectin, and stool lactoferrin for detection of endoscopic activity in symptomatic inflammatory bowel disease patients: a systematic review and meta-analysis. *Am J Gastroenterol* 2015;110:802–19.
32 Thia KT, Sandborn WJ, Lewis JD, *et al*. Defining the optimal response criteria for the Crohn's disease activity index for induction studies in patients with mildly to moderately active Crohn's disease. *Am J Gastroenterol* 2008;103:3123–31.
33 Vermeire S, Schreiber S, Sandborn WJ, *et al*. Correlation between the Crohn's disease activity and harvey-bradshaw indices in assessing Crohn's disease severity. *Clin Gastroenterol Hepatol* 2010;8:357–63.
34 Ferrante M, Colombel J-F, Sandborn WJ, *et al*. Validation of endoscopic activity scores in patients with crohn's disease based on a post hoc analysis of data from sonic. *Gastroenterology* 2013;145:978–86.
35 Rutgeerts P, Sandborn WJ, Feagan BG, *et al*. Infliximab for induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 2005;353:2462–76.