

Central Lancashire Online Knowledge (CLOK)

| | |
|----------|---|
| Title | The Quality of reporting Inflammatory Bowel Disease Randomized Control Trials: A systematic review |
| Type | Article |
| URL | https://clock.uclan.ac.uk/50968/ |
| DOI | https://doi.org/10.1136/bmjgast-2023-001337 |
| Date | 2024 |
| Citation | Gordon, Morris, Khudr, Jamal, Sinopoulou, Vassiliki, Lakunina, Svetlana, Rane, Aditi and Akobeng, Anthony (2024) The Quality of reporting Inflammatory Bowel Disease Randomized Control Trials: A systematic review. <i>BMJ Open Gastroenterology</i> , 11 (1). |
| Creators | Gordon, Morris, Khudr, Jamal, Sinopoulou, Vassiliki, Lakunina, Svetlana, Rane, Aditi and Akobeng, Anthony |

It is advisable to refer to the publisher's version if you intend to cite from the work.
<https://doi.org/10.1136/bmjgast-2023-001337>

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLOK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

Quality of reporting inflammatory bowel disease randomised controlled trials: a systematic review

Morris Gordon ,^{1,2} Jamal Khudr ,^{1,3} Vassiliki Sinopoulou ,¹ Svetlana Lakunina,¹ Aditi Rane,¹ Anthony Akobeng^{4,5}

To cite: Gordon M, Khudr J, Sinopoulou V, *et al.* Quality of reporting inflammatory bowel disease randomised controlled trials: a systematic review. *BMJ Open Gastroenterol* 2024;**11**:e001337. doi:10.1136/bmjgast-2023-001337

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/bmjgast-2023-001337>).

Received 22 December 2023
Accepted 2 April 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹School of Medicine, University of Central Lancashire, Preston, UK

²Blackpool Families Division, Blackpool Victoria Hospital, Blackpool, UK

³Plastic Surgery, Manchester University NHS Foundation Trust, Manchester, UK

⁴Gastroenterology, Sidra Medicine, Doha, Ad Dawhah, Qatar

⁵School of Medicine, Weill Cornell Medical College-Qatar, Doha, Qatar

Correspondence to

Dr Morris Gordon;
MGordon@uclan.ac.uk

ABSTRACT

Objective Our objective was to perform a systemic evaluation of the risk of bias in randomised controlled trial (RCT) reports published on inflammatory bowel disease (IBD).

Design We assessed the risk of bias using the Cochrane tool, as indicators of poor methodology or subsequently poor reporting. We systematically selected, with dual independent judgements, all studies published on IBD with no time limits and assessed the methodological quality of included studies again using independent dual ratings.

Results 563 full texts were included after selection and review. No abstract publications were free of any source of bias. Full-text publications still fared badly, as only 103 full-text papers exhibited a low risk of bias in all reporting domains when excluding blinding. RCTs published in journals with higher impact factor (IF) were associated with an overall reduced rate of being at high risk. However, only 6% of full RCT publications in journals with an IF greater than 10, published in the past 5 years, were free of bias.

The trend over time is towards improved reporting in all areas. Trials published by larger author teams, in full-text form and by industry and public sponsorship were positively correlated with a lower risk of bias. Only allocation concealment showed a statistically significant improvement with time ($p=0.037$).

Conclusion These findings are consistent with those of other specialties in the literature. While this unclear risk of bias may represent poor reporting of methods instead of poor methodological quality, it leaves readers and future secondary researchers with significant questions regarding such key issues.

INTRODUCTION

Randomised controlled trials (RCTs) are the study of choice for evaluating the efficacy of interventions in the management of inflammatory bowel disease (IBD).^{1,2} Their methodological approach to reduce the risk of bias (RoB) ensures that the true effects of the intervention are reported in a manner that can best represent clinical reality.

In considering published RCTs, it is difficult to distinguish poor writing from

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ All randomised trials published should report using the Consolidated Standards of Reporting Trials guidance, released in 1995. This is more important now as the use of Grading of Recommendations, Assessment, Development and Evaluations (GRADE) as a system to support guideline decision-making means poor method reporting of individual studies impacts the certainty of overall outcomes used to make recommendations.

WHAT THIS STUDY ADDS

⇒ When comparing studies published in the past 5 years in inflammatory bowel disease (IBD) to those before, statistically significant improvement was only noted in one out of the six risk of bias domains. The highest impact journals in the last 5 years still publish less than 1 in 10 IBD trials at low risk of bias in all areas.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Reporting of risk of bias must be a priority for all researchers and peer reviewers as poor reporting threatens the certainty of all decision-making by guidelines committees when using the GRADE framework to make decisions internationally. Journals should strongly endorse adherence to reporting guidelines to authors and highlight these six key criteria as mandatory for clear reporting for all randomised controlled trials.

poor research quality. All studies should report using the Consolidated Standards of Reporting Trials (CONSORT) statement's minimum methodological standards of reporting that were initially released in 1995.³ While the CONSORT statement is a requirement for major journals, especially those with the top 1% of impact factors (IF), many journals do not mandate this which often leaves readers with uncertain interpretations. While peer review can and should address this and be expected to adhere to a three-decade-old standard, there are long-standing concerns

with the validity and reliability of this process within the field of academia.^{3,4}

Quality is a sometimes abstract concept, often a subjectively understood term. It has been further complicated in the last decade with the hugely significant emergence of a consensus on rating the quality of a whole evidence base for each outcome through the use of Grading of Recommendations, Assessment, Development and Evaluations (GRADE).⁵ For individual RCTs, quality is typically described as 'design and RCT conduct, to prevent systematic errors, or bias'.⁶ Bias occurs when the results of a study do not represent the truth because of the inherent limitations in the design or conduct of a study.⁷

The Cochrane RoB tool is employed to judge key elements within evidence reviews.^{8,9} Selection bias, which includes appropriate randomisation, also includes the difficult-to-understand principle of allocation concealment (AC) which leads to overestimates of the treatment effect of close to 40% if poorly reported.¹⁰ This is best understood with an example. In an RCT comparing biological IBD treatment to a placebo, even with an appropriate randomisation schedule, if the allocation schedule is available during recruitment, a researcher aware that the next patient will receive a biologic might avoid a patient with difficult venous access. This can create an imbalance between groups. This is not the same as blinding, which can be impractical, but concealing allocation schedules, which is always feasible and crucial to avoid bias. Other sources of bias do exist that are not specifically accounted for within the Cochrane criteria and are usually accounted for and identified through overall assessments of available data. The other bias (OB) section within the Cochrane criteria allows for documentation of the above concerns; however, the other six criteria have universal agreement as the most specific sources of bias with the highest influence on reported outcomes.

A previous study on over 20 000 RCTs in Cochrane reviews until 2014 noted improved reporting trends in all these key items of reporting within studies published between 1997 and 2008.¹¹ In IBD, a limited evidence base informs the treatment of vast patient numbers. Recently, most IBD international guideline committees have adopted the GRADE approach for quality and recommendation assessments. Imperatively, GRADE evaluates bias risk at the outcome level, not per study. Thus, despite the presence of some studies at low RoB, overall evidence quality can be downgraded if other studies exhibit higher risk.⁵ This emphasises the increasing need for unbiased RCT reporting to distinguish genuine quality concerns from reporting issues.

We set out to examine how these key elements of RoB are reported within IBD RCTs and factors associated with a higher RoB through a systematic review of all IBD RCTs published since the CONSORT statement (1995).³ This will be completed using the Cochrane RoB tool to a whole sample of published RCTs.

MATERIALS AND METHODS

The protocol for the review has been uploaded to a repository (Repository ID: 33117).¹² An ethical screening tool was completed, and full ethical approval was not required.

Study selection

We performed a systematic electronic database search of all the RCTs published on IBD from the following databases: MEDLINE, Embase, CENTRAL and The Cochrane Register (online supplemental appendix 1). All studies on IBD were collected up to the date of the search (September 2020). All citations were imported into Rayyan and deduplicated for abstract screening in duplicate at the title and abstract level.

Inclusion criteria included RCTs involving patients with either Crohn's disease, ulcerative colitis or a combination of the conditions. Trials including all age groups and patients in any disease state were considered. Studies could involve any form of intervention compared with any other intervention, placebo, no treatment or usual care. Studies could include any outcome measures. Studies could include any outcome measures. All early phase II and phase III trials were included. Phase I, animal trials, quasirandomised trials and other study types were not included. No time or language restrictions were imposed.

Three authors (JK, VS and SL) independently reviewed all titles and abstracts, with two author judgements required before the progression of a paper to full-text assessment. Any disagreements were resolved by a fourth author (MG).

All journal articles chosen for full-text review were evaluated independently again by two authors (JK and AR) to assess inclusion criteria to consider for analysis, with the third and fourth authors (MG and VS) resolving differences. RCTs published in abstract form only, which met the eligibility criteria, were also included in this review.

Data extraction and quality assessment

From included RCTs (full text and abstract), data were simultaneously extracted independently by the above authors (JK and AR), and using a standardised form, an RoB judgement was recorded for each item, as discussed above. Additionally, key demographic and descriptive data were extracted, including the publication type (full text/abstract/letter), year of publication, journal source, language, number of authors, funding source and number of study centres. Disagreements were once again resolved through the involvement of a third author (VS and MG).

A matching algorithm and manual validation were used to standardise journal names and eliminate abbreviations. We visited the website of each listed journal, if available, and extracted its up-to-date IF. The IF centile and year-specific IF for each publication were then collected from Journal Citation Reports.

The publication characteristics were then combined with the collected RoB assessment for each RCT, and

as such were categorised as 'low', 'high' or 'unclear' for each key RoB item as per the Cochrane RoB tool.⁶ In the context of missing information, primary authors were not contacted.

Assessment of methodological quality and conduct rigour

Our assessment of RCTs' methodological quality was based on the premise that a key item is considered at unclear RoB when inadequate information is provided to allow judgement of high or low risk, in line with the tool.⁶

Analysis

The data were first analysed to assess the RoB for each key methodological item (low, high, unclear). We initially considered the whole sample of trials. Then, we compared the full text to the abstract overall. For subsequent analyses, abstracts were removed to avoid conflating limited reporting space with method quality.

For the full RCT publications, we then evaluated each criterion to determine the proportion of trials at low and high RoB. This was then used to determine the evolution of poor reporting over time, its association to IF, IF date, centile of IF, funding source, reported form, study centre and the number of authors. The synthesis was narrative and descriptive, in line with appropriate reporting guidance.¹³ We reported the overall proportions of key items using both bar graphs and pie charts and then explored the impact of other factors on these key items as above.

We calculated using SPSS (v27) descriptive statistics and conducted χ^2 tests for statistical differences as per online supplemental appendix 2.¹⁴ Comparison was made between the group of papers that had a low RoB to those with high and unclear RoB judgements as markers of poor reporting.

RESULTS

Selection process

After excluding duplicates, 1851 unique papers were examined and published in 167 journals (figure 1). After abstract and full-text screening, 563 RCTs were included, encompassing 362 unique full-text articles (median year of publication 2015, IQR 2006–2018) reported in 86 journals (table 1).

General characteristics

The IF for the journals can be seen in table 1. RCTs were most likely to be reported in journals with IF ≥ 10 (285 (50.6%)). A large majority were written in English. Of the RCTs included, 14 were published in journals without an IF (2.5%).

Methodological quality assessment

Overall assessment

For sequence generation (SG) and AC, the proportion of trials judged as 'low risk' and 'unclear risk' was 59.1% and 66.4%, respectively. The unclear RoB was similar for the blinding of participants and personnel (BPP) at 41.6% and the blinding of outcome assessors (BOA) at 51.7%.

For incomplete outcome data (IOD), selective reporting (SR) and OB, the figures were 39.4%, 25.8% and 35.9%, respectively (figure 2).

The proportion of trials at 'high risk' of bias of all trials was 2.5% for SG and 2.1% for AC. It was significantly worse for the components of blinding where it was at 17.1% for BPP and 12.1% for BOA. The proportion of studies at 'high risk' for IOD, SR and OB was 3.6%, 2.8% and 3.7%, respectively (figure 2).

Abstract-only versus full-text publications

Trials published in full-text form compared with those in abstract form had a reduced proportion of 'unclear risk' of bias for all six RoB items. When considering unclear RoB as a surrogate for poor reporting, and consequently poor methodology, significant disparities can be seen between full-text and abstract-form publications when assessing risk (figure 3). This was particularly evident for IOD data which experienced a 65% increase in high and unclear RoB when reported in abstract form.

All subsequent analysis is presented with RCTs published in abstract form removed from the sample.

Author team size

When comparing author groups with fewer than five authors to those with more, there were notable differences in the quality of trials (figure 4). Specifically, trials conducted by teams with less than five authors showed a significant increase in 'unclear risk' regarding SG and BPP. The increase was 35% ($p=0.014$) for SG and 19.9% ($p=0.021$) for blinding, indicating a higher proportion of unclear risk in smaller author teams.

Impact factor

Journals with high IF were associated with a lower proportion of trials at 'high' or 'unclear risk' compared with those with low or no IF (figure 5).

Statistical differences were seen in five out of the seven RoB components when comparing RCTs published in journals with an IF compared with those without one. Similar trends were evident between RCTs published in journals with an IF <5 to those >5 . No statistically significant differences were evident when comparing journals with an IF between 5 and 10 to those with an IF >10 .

This trend was evident when journals were examined by their IF centile, as trials with no IF were 40.4% more likely to have unclear or high judgements compared with those above the 90th centile.

Funding source

We found a clear difference in the proportion of trials at unclear risk by funding source (figure 6). Studies without a funding source or those that did not specify any were more likely to have a 'high' or 'unclear' RoB when compared with funded studies, especially when considering AC ($p=0.0029$) and BPP ($p=0.0034$). Public sponsorship-funded studies experienced a higher proportion of trials at 'unclear' and 'high risk', in comparison to studies with industry funding, when considering BOA

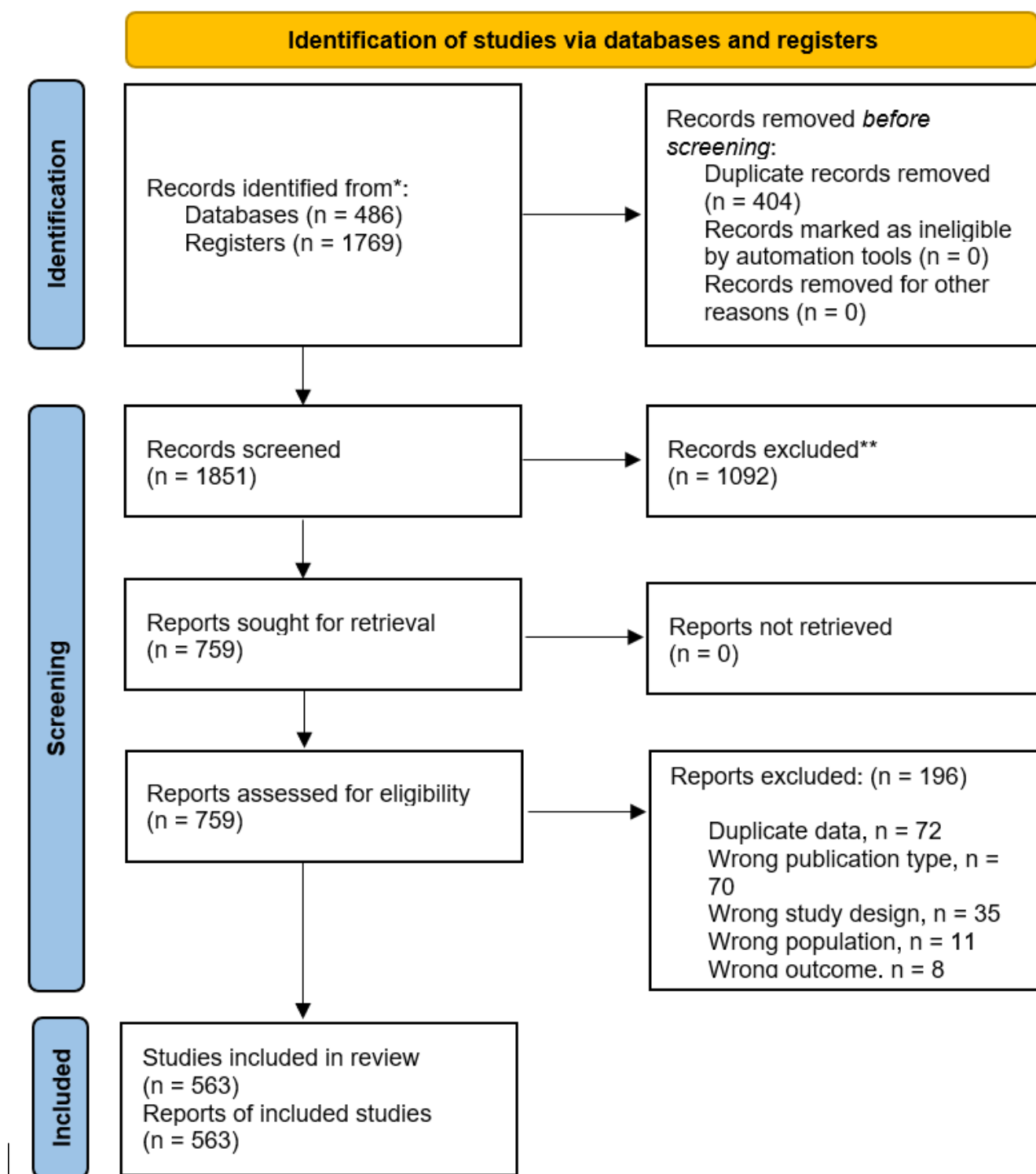


Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram outlining the screening process. *Totalled from different sources in each category. **All with independent dual decision making, no use of Artificial Intelligence tools

($p=0.0187$), IOD ($p=0.0425$) and SR ($p=0.0315$). No statistically significant difference was noted between industry and public sponsorship-funded studies when compared with those funded by the industry.

Multicentre versus single-centre trials

Studies conducted in a multicentre setting conveyed lower RoB in all methodological items compared with studies conducted at multiple centres (online supplemental

**Table 1** Characteristics of included RCTs

| Characteristic | n (%) unless otherwise specified |
|--|----------------------------------|
| Publication year | 2015 (2006–2018), 1974–2020 |
| Published in journal with impact factor | 549 (97.5) |
| Journal impact factor* | |
| ≥10 | 285 |
| 5–10 | 143 |
| <5 | 121 |
| Published in journal without an impact factor | 14 |
| 10 highest represented journals | |
| <i>Gastroenterology</i> | 133 (23.6) |
| <i>Journal of Crohn's and Colitis</i> | 65 (11.5) |
| <i>Gut</i> | 50 (8.9) |
| <i>Inflammatory Bowel Diseases</i> | 36 (6.4) |
| <i>American Journal of Gastroenterology</i> | 32 (5.7) |
| <i>Alimentary Pharmacology & Therapeutics</i> | 25 (4.4) |
| <i>United European Gastroenterology Journal</i> | 24 (4.3) |
| <i>New England Journal of Medicine</i> | 21 (3.7) |
| <i>Clinical Gastroenterology and Hepatology</i> | 15 (2.7) |
| <i>Lancet</i> | 13 (2.3) |
| *Median (IQR), min-max. RCT, randomised controlled trial. | |

appendix 3). For example, when looking at AC it is evident that 49.7% of trials conducted in single centres were at 'unclear risk' of bias compared with 33.8% of those in a single centre.

Evolution of the quality of RCT reporting

The proportion of RCTs at unclear RoB displayed mixed results. For AC, SG, BPP and BOA, they experienced a fall of 44%, 27% and 47%, respectively, in the 'unclear risk' categories when comparing data in RCTs from <1985 to those from 2016 to 2020 (online supplemental appendix 4). This is then contrasted by IOD, SR and OB which experienced a corresponding 7%, 2% and 6% increase in the proportion of trials at 'unclear bias' when examining RCTs published between the same periods.

Journal IF was not significant in the evolution of unclear RoB over time as similar inconclusive data were prevalent when comparing journals with different IF.

It is worth noting that when comparing the output from the most recent 3 years, when comparing the risk of 'unclear' or 'high' bias between studies published after 2017 to those prior, only AC showed a statistically significant improvement with time ($p=0.037$).

When blinding is excluded from the assessment tool due to its impracticality in specific study types, like surgical trials where attaining a low-risk status is pragmatically unattainable, we found that 103 (18.29%) papers exhibited a low RoB in all reporting domains. All of these low-risk papers were published in full-text form. When examining the breakdown between IF categories, 36 (6.3%) were published in journals with IF <5, 25 (4.4%) with IF 5–10 and 42 (7.5%) with IF >10. When examining the 127 papers that were published in the last 5 years in journals with an IF >10, eight trials (6.3%) were free of bias in all domains (excluding blinding).

X² evaluation of predetermined factors

There were statistically significant differences in reporting in journals with an IF compared with those without one when looking at all RoB categories except for AC and BPP (online supplemental appendix 5). Similarly, there were statistically higher chances of the studies being at low RoB when comparing studies published in journals with an IF >5 to those published in journals with an IF <5. In the final analysis, no statistically significant differences were present when comparing studies published in journals with an IF between 5 and 10 to those with an IF >10.

When considering funding, it is evident that having a funding source is associated with a statistically significant improvement in reporting AC and BPP. Having an industry source of funding demonstrates statistical improvement in BOA, IOD and SR when compared with studies funded by public sponsorship. Studies funded by both industry and public sponsorship sources had no differences when compared with those that were only funded by industry sources.

Looking at reporting over time, studies published in the last 3 years only exhibited a statistically significant improvement in AC when compared with those published before 2017.

DISCUSSION

This review indicates that, despite reporting improvements over time, many full-text published trials still are not at low RoB in essential reporting elements.

Our study revealed lower RoB levels when abstracts were excluded. This highlights concerns for conference organisers, guideline developers and others who rely on abstracts. Often, abstracts are the only trial form for years, sometimes never becoming full papers, impacting future GRADE-based decisions. This poses a question about the use of abstracts in systematic reviews due to

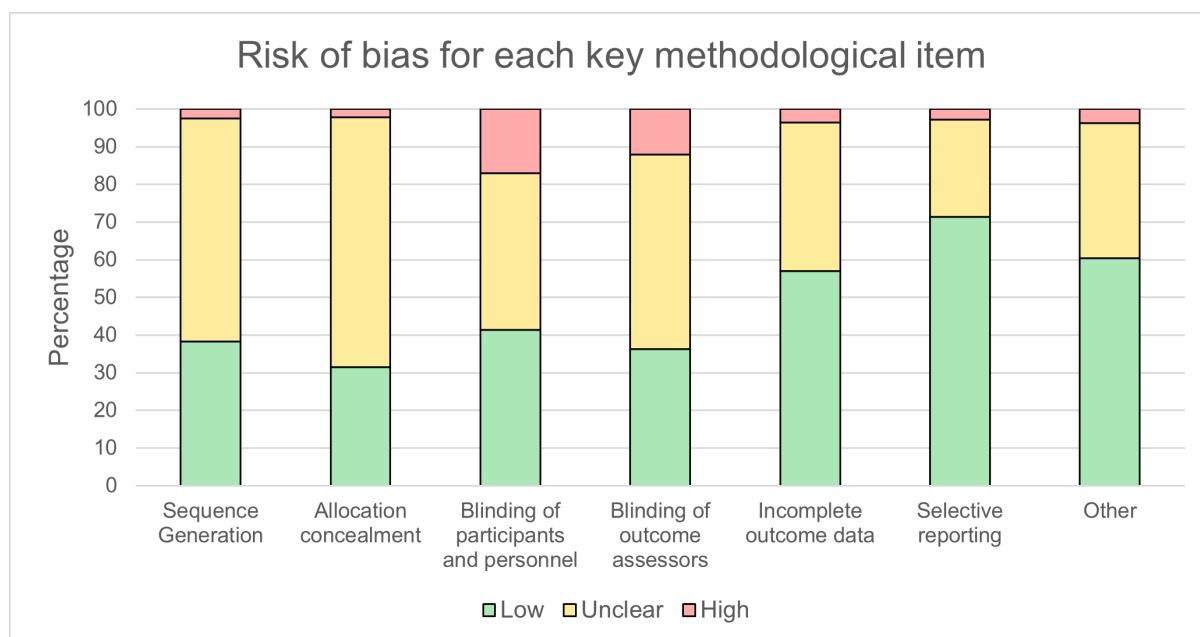


Figure 2 Risk of bias for included elements.

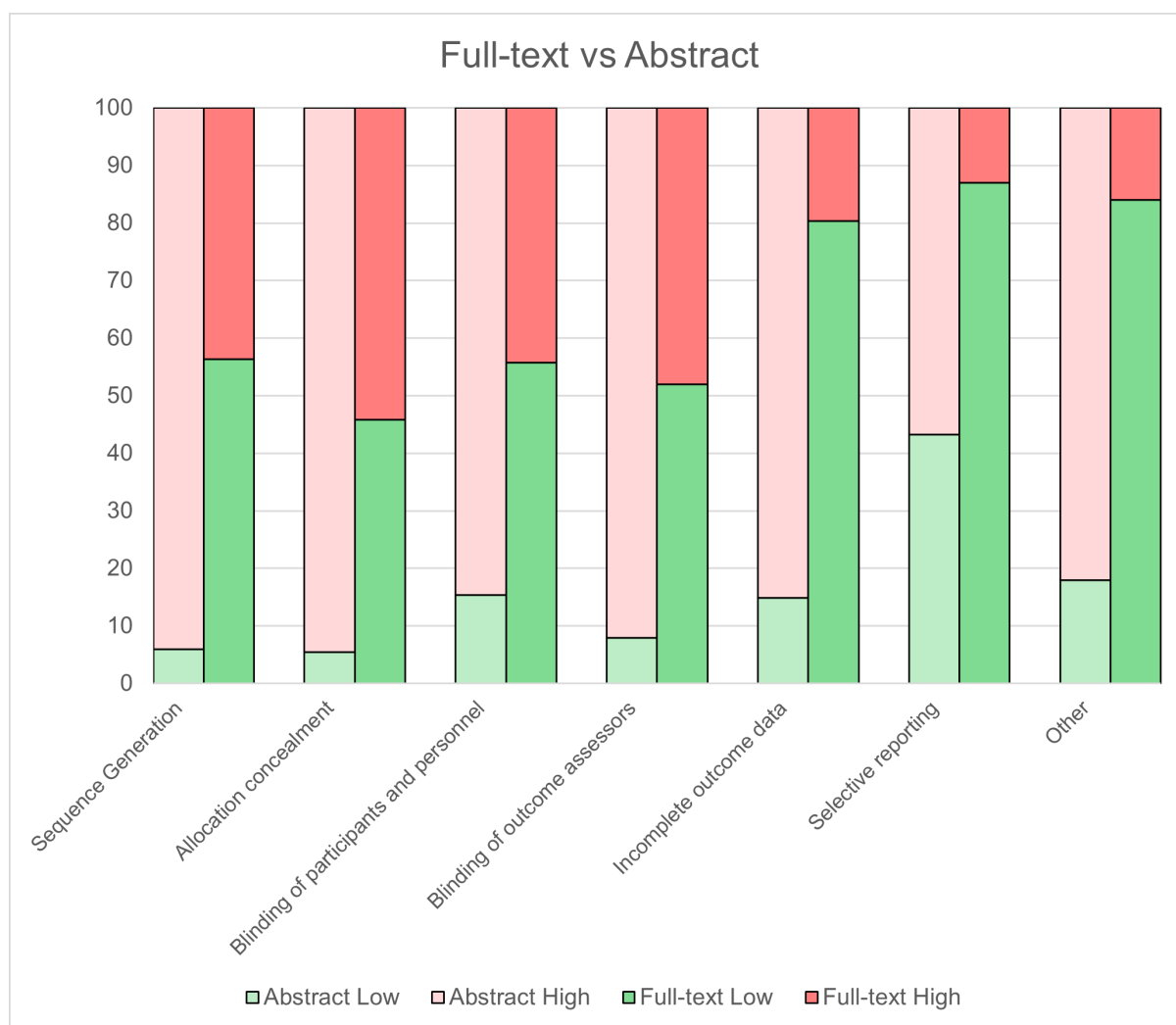


Figure 3 Risk of bias comparing full-text to abstract-only studies.

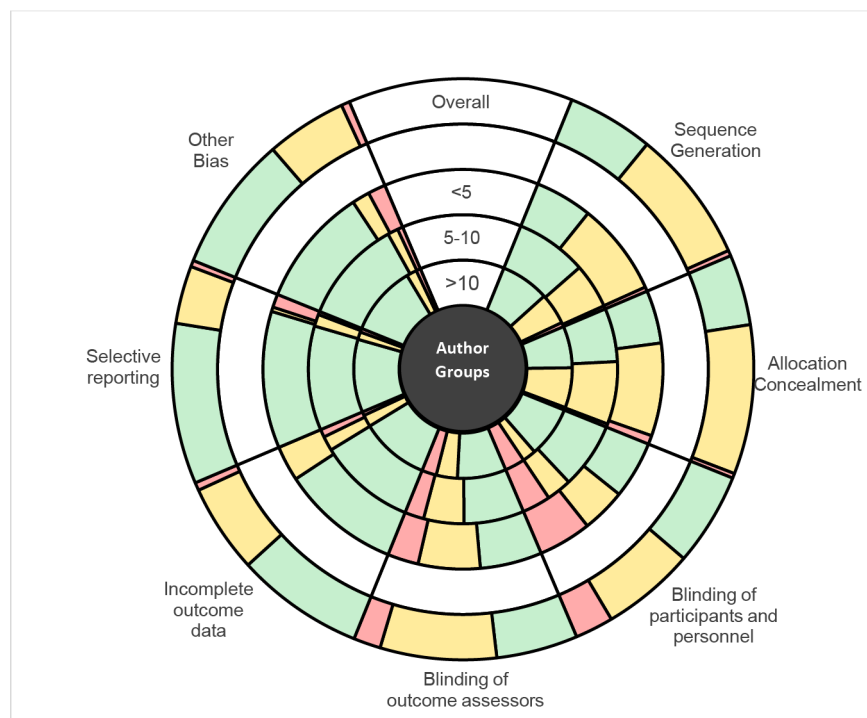


Figure 4 Risk of bias comparing author team size.

potential bias. With almost a third of RCTs remaining unpublished after abstract presentation, relying on this 'grey' literature introduces publication bias favouring positive results.^{15 16} To mitigate this bias, more thorough peer reviews for conference abstracts and a study design registry or detailed display tool for abstracts are essential.

Within the full-text trials only, the same patterns were seen with improvements over time, but low-risk reporting in all categories is still a rarity. This has huge consequences as it directly lowers the strength of the

recommendations during GRADE guideline development which is paramount.

Our study and those in the literature in other fields have shown that high rates of uncertainty exist in assessing the RoB, particularly for the SG and AC components.¹⁷ Studies reporting unclear or high risk of selection bias were more likely to reach positive conclusions and exaggerate intervention effects by an average of 9%.^{17 18} While the complete elimination of high RoB may not always be achievable due to the specific constraints of certain study

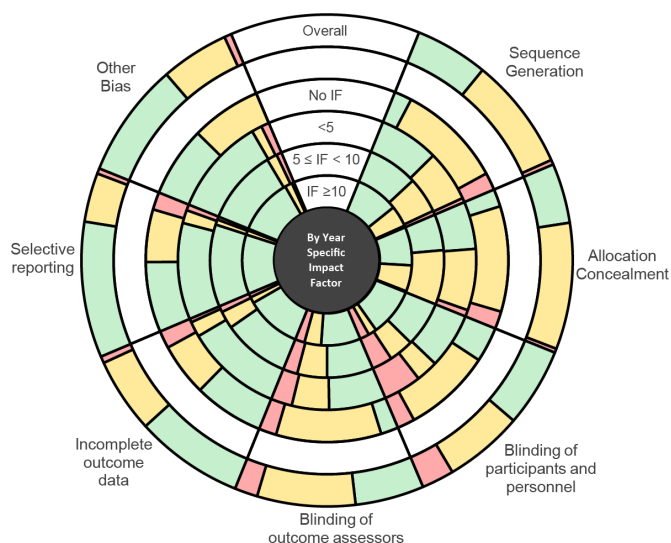


Figure 5 Risk of bias comparing journal impact factor (IF) of published studies.

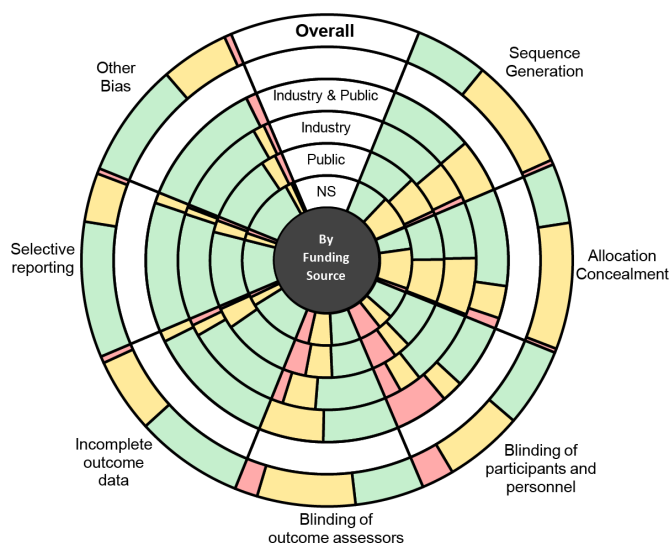


Figure 6 Risk of bias comparing funding source and methodological quality. NS, not specified.

designs, a significant portion of the unclear RoB can be attributed to inadequate reporting practices. It is critical to recognise that unclear RoB, being largely preventable, is unacceptable as it substantially limits the interpretation, application and certainty of available data. Additionally, while cautious interpretation of unclear RoB data by readers is essential, the implementation of the GRADE approach, which prioritises overall outcomes over individual study evaluations, typically leads to more guarded recommendations in evidence-based guidelines.⁵ This consideration gains further importance in light of the increasing integration of the GRADE methodology on an international scale, particularly in the UK, as evidenced by the forthcoming British Society of Gastroenterology guidelines.¹⁹

While the lack of blinding influence on effect measures has been highlighted,¹⁷ its challenging implementation in certain trials, such as surgery, emphasises the significance of AC for mitigating selection bias, given its feasibility and cost-effectiveness. While some trials suggest an overestimation of treatment effects due to inadequate AC,^{10 20–22} others do not.^{23–27} Regardless, instances of insufficient AC reporting should be reduced.

Industry studies often show biased outcomes.²⁸ Industry-backed studies tend to overstate efficacy, while non-industry ones more frequently report harmful effects.^{29 30} Industry funding also correlates with a higher bias risk in SG than public sponsorship.³¹ In our dataset, industry-funded studies reported a 'low risk' of bias for key items when compared with publicly sponsored ones, possibly reflecting their understanding of these key items' significance. Other explanations might be their correlation with other factors that improve the RoB reporting such as larger author groups, multicentre settings and publishing in higher IF journals.

The evolution of reporting has been shown to improve over time throughout the literature and across different specialties and could be linked to the development of reporting guidelines, especially CONSORT.^{32–34} With a prominent upswing in reporting quality at the point of the introduction of the CONSORT statement in 1995 in our data (online supplemental appendix 4), a generalised improvement in reporting methodology is quite discernible over the 5 years after the release of the statement. Examining the data, it is interesting to note that the percentage of poorer quality papers has primarily remained constant over time since then and it is unclear why this progress has stagnated.

Our findings demonstrate that studies with a lower RoB tend to be published in high-impact journals, consistent with prior research.³⁵ Whether attributed to improved funding, abundant resources or larger samples remains undetermined. Notably, even recent cohorts in these highest impact journals were still rarely at low RoB in all areas, highlighting a missed opportunity for these highly cited journals to champion transparent and CONSORT-aligned IBD reporting. It is unclear why this is the case.

We further recommend that journals strongly endorse the adherence of reporting guidelines to RCT authors and highlight these six key criteria as mandatory for all RCTs among peer reviews.

This study's limitations include its focus on IBD RCTs, potentially limiting generalisability to other specialties or populations. We did not contact authors for clarification on unclear areas of RoB. While Cochrane is piloting a 2.0 bias assessment tool promising enhanced judgements, its core criteria remain largely unchanged. The introduction of a recent update to CONSORT in 2020 also further reinforces the importance of this item within avoiding uncertainty in reporting. Future research could consider using this updated tool.³⁶

CONCLUSION

The proportion of trials with inadequate methods and poor reporting within the field of IBD has reduced over time and in keeping with those of other specialties in the literature. However, almost half of the studies from the past 3 years carry potential bias risks, raising concerns for readers and researchers. Given GRADE's widespread use in guideline development, these inconsistencies undermine confidence in IBD decision-making, highlighting the need for rigorous bias risk reporting among researchers and reviewers.

X Jamal Khudr @jamal_khudr

Contributors MG: study concept and design; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content; study supervision. MG and JK share co-first authorship, but MG acts as a guarantor for the paper and its contents. JK and VS: acquisition of data; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content. SL and AR: acquisition of data; analysis and interpretation of data; drafting of the manuscript. AA: study concept and design.

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 None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. All data relevant to the study are included in the article or uploaded as supplementary information. The protocol for the review has been uploaded to a repository (Repository ID: 33117). Further methodological components and results are included in the appendix. Other data are available upon reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is

properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Morris Gordon <http://orcid.org/0000-0002-1216-5158>

Jamal Khudr <http://orcid.org/0000-0002-4148-0499>

Vassiliki Sinopoulou <http://orcid.org/0000-0002-2831-9406>

REFERENCES

- Sandler RS. Reporting randomized controlled trials in Gastroenterology: the CONSORT statement. *Gastroenterology* 2001;121:755.
- Elmunzer BJ. Increasing the impact of randomized, controlled trials in gastrointestinal Endoscopy. *Gastroenterology* 2015;149:521–5.
- Begg C. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637.
- Smith R. Peer review: a flawed process at the heart of science and journals. *J R Soc Med* 2006;99:178–82.
- Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
- Cochrane. COCHRANE CONSUMERS & COMMUNICATION REVIEW GROUP STUDY QUALITY GUIDE 2013.
- Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol* 2011;64:407–15.
- Higgins JPT. Cochrane collaboration. In: *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd edn. Hoboken, NJ: Wiley-Blackwell, 2020.
- Higgins JPT, Altman DG, Gotzsche PC, et al. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
- Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–12.
- Dechartres A, Trinquart L, Atal I, et al. n.d. Evolution of poor reporting and inadequate methods over time in 20920 randomised controlled trials included in Cochrane reviews: research on research study. *BMJ*;j2490.
- Khudr J, Iheozor-Ejiofor Z, Gordon M. The quality of reporting inflammatory bowel disease randomized control trials: A systematic review. *Gastroenterology* 2020;158:S–668.
- Wong G, Greenhalgh T, Westhorp G, et al. RAMESES publication standards: meta-narrative reviews. *BMC Med* 2013;11:20.
- IBM Corp. IBM SPSS Statistics for Windows, Version 28.0. Armonk, NY: IBM Corp, 2021.
- Scherer RW, Saldanha IJ. How should systematic reviewers handle conference abstracts? A view from the trenches. *Syst Rev* 2019;8:264.
- Scherer RW, Meerpohl JJ, Pfeifer N, et al. Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev* 2018;11:MR000005.
- Savovic J, Turner RM, Mawdsley D, et al. Association between risk-of-bias assessments and results of randomized trials in Cochrane reviews: the ROBES meta-epidemiologic study. *Am J Epidemiol* 2018;187:1113–22.
- Cramer H, Langhorst J, Dobos G, et al. Associated factors and consequences of risk of bias in randomized controlled trials of yoga: A systematic review. *PLOS ONE* 2015;10:e0144125.
- Darie A-M, Sinopoulou V, Ajay V, et al. BSG 2024 IBD guidelines protocol (standard operating procedures). *BMJ Open Gastroenterol* 2023;10:e001067.
- Egger M, Juni P, Bartlett C, et al. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? empirical study. *Health Technol Assess* 2003;7:1–76.
- Herbison P, Hay-Smith J, Gillespie WJ. Adjustment of meta-analyses on the basis of quality scores should be abandoned. *J Clin Epidemiol* 2006;59:1249–56.
- Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses *The Lancet* 1998;352:609–13.
- Als-Nielsen B, Chen W, Lotte Gluud L, et al. Are trial size and reported methodological quality associated with treatment effects? observational study of 523 randomised trials in. 2004.
- Balk EM, Bonis PAL, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002;287:2973–82.
- Bialy L, Vandermeer B, Lacaze-Masmonite T, et al. A meta-Epidemiological study to examine the association between bias and treatment effects in neonatal trials: the association between bias and treatment effects in neonatal trials. *Evid-Based Child Health Cochrane Rev J* 2014;9:1052–9.
- Chaimani A, Vasiladis HS, Pandis N, et al. Effects of study precision and risk of bias in networks of interventions: a network meta-Epidemiological study. *Int J Epidemiol* 2013;42:1120–31.
- Hartling L, Hamm MP, Fernandes RM, et al. Quantifying bias in randomized controlled trials in child health: A meta-Epidemiological study. *PLoS ONE* 2014;9:e88008.
- Lundh A, Lexchin J, Mintzes B, et al. Industry sponsorship and research outcome. *Cochrane Database Syst Rev* 2017;2:MR000033.
- Bero L, Anglemeyer A, Vesterinen H, et al. The relationship between study sponsorship, risks of bias, and research outcomes in Atrazine exposure studies conducted in non-human animals: systematic review and meta-analysis. *Environ Int* 2016;92–93:597–604.
- Abdel-Sattar M, Krauth D, Anglemeyer A, et al. The relationship between risk of bias criteria, research outcomes, and study sponsorship in a cohort of Preclinical Thiazolidinedione animal studies: a Meta-Analysis. *Evid Based Preclin Med* 2014;1:e00005:11–20.
- Crocetti MT, Amin DD, Scherer R. Assessment of risk of bias among pediatric randomized controlled trials. *PEDIATRICS* 2010;126:298–305.
- Moher D, Jones A, Lepage L. CONSORT group (Consolidated standards for reporting of trials). use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001;285:1992–5.
- Piggott M, McGee H, Feuer D. Has CONSORT improved the reporting of randomized controlled trials in the palliative care literature? A systematic review. *Palliat Med* 2004;18:32–8.
- To MJ, Jones J, Emara M, et al. Are reports of randomized controlled trials improving over time? A systematic review of 284 articles published in high-impact general and specialized medical journals. *PLoS One* 2013;8:e84779.
- Bala MM, Akl EA, Sun X, et al. Randomized trials published in higher vs. lower impact journals differ in design, conduct, and analysis. *J Clin Epidemiol* 2013;66:286–95.
- Sterne JAC, Savović J, Page MJ, et al. Rob 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898.

Appendix 1

Search strategy

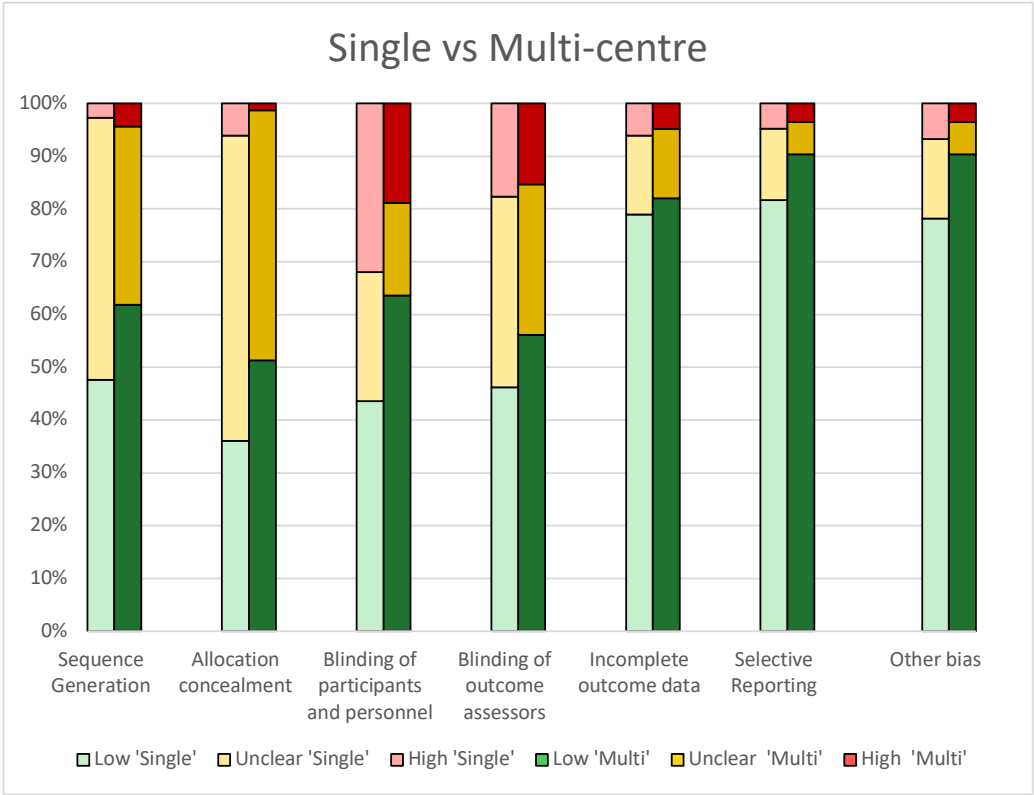
(random[Title/Abstract] OR factorial[Title/Abstract] OR crossover[Title/Abstract] OR cross over[Title/Abstract] OR cross-over[Title/Abstract] OR placebo[Title/Abstract] OR single blind[Title/Abstract] OR double blind[Title/Abstract] OR triple blind[Title/Abstract] OR assign[Title/Abstract] OR allocate[Title/Abstract] OR randomized controlled trial[Title/Abstract]) AND (Crohn disease[Title/Abstract] OR Crohn*[Title/Abstract] OR Ulcerative colitis[Title/Abstract] OR ulcerative col*[Title/Abstract] OR IBD[Title/Abstract] OR Inflammatory bowel disease*[Title/Abstract])

Appendix 2

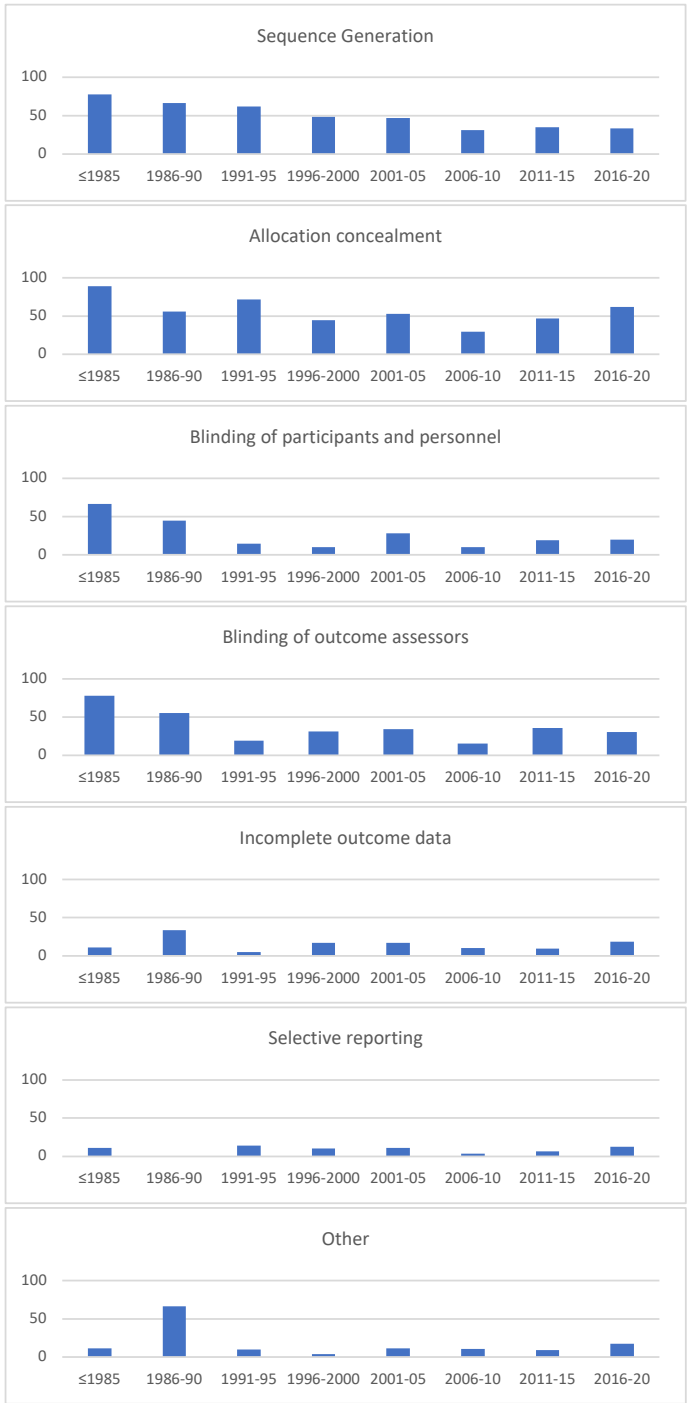
Analytical strategy

SPSS descriptive statistics and Chi-squared tests were conducted for statistical differences between pre-determined factors including studies published within journals with an IF to those with no IF, studies published in journals with an IF <5 to those with an IF >5, studies published in journals with an IF between 5-10 to those with an IF >10; studies with no-specified funding source to those with one, studies funded by public sponsorship to those funded by the industry, studies funded by the industry to those funded by both the industry and public sponsorship; studies published within the last three years to those published prior; and studies published by author groups with less than five authors to those with greater than five authors

Appendix 3 Risk of Bias comparing study centre setting



Appendix 4 Unclear risk of bias and progression of reporting over time



Appendix 5

Chi-squared comparison between pre-determined different factors

| | | | | | | |
|---------------|---|----------|---------------------|----------|------------------------|--------|
| Impact Factor | SG for NIF to IF | *0.0122 | SG for IF <5 to >5 | **0.0019 | SG for IF 5-10 to >10 | 0.7779 |
| | AC for NIF to IF | 0.0576 | AC for IF <5 to >5 | *0.0380 | AC for IF 5-10 to >10 | 0.2957 |
| | BPP for NIF to IF | 0.1550 | BPP for IF <5 to >5 | *0.0284 | BPP for IF 5-10 to >10 | 0.4271 |
| | BOA for NIF to IF | *0.0228 | BOA for IF <5 to >5 | 0.1252 | BOA for IF 5-10 to >10 | 0.8664 |
| | IOD for NIF to IF | *0.0241 | IOD for IF <5 to >5 | 0.4051 | IOD for IF 5-10 to >10 | 0.6523 |
| | SR for NIF to IF | **0.0018 | SR for IF <5 to >5 | *0.0376 | SR for IF 5-10 to >10 | 0.2218 |
| | OR for NIF to IF | *0.0104 | OR for IF <5 to >5 | 0.6130 | OR for IF 5-10 to >10 | 0.8053 |
| | p-value when comparing papers published in journals with different impact factors | | | | | |

| | | | | | | |
|---------|---|---------|----------------------------|---------|-------------------------|--------|
| Funding | SG for NS to Funded | 0.0583 | SG for Public to Industry | 0.2278 | SG for Industry to I+P | 0.9709 |
| | AC for NS to Funded | *0.0029 | AC for Public to Industry | 0.8529 | AC for Industry to I+P | 0.1559 |
| | BPP for NS to Funded | *0.0034 | BPP for Public to Industry | 0.0502 | BPP for Industry to I+P | 0.1576 |
| | BOA for NS to Funded | 0.0644 | BOA for Public to Industry | *0.0187 | BOA for Industry to I+P | 0.6194 |
| | IOD for NS to Funded | 0.8951 | IOD for Public to Industry | *0.0425 | IOD for Industry to I+P | 0.4394 |
| | SR for NS to Funded | 0.1222 | SR for Public to Industry | *0.0316 | SR for Industry to I+P | 0.9260 |
| | OR for NS to Funded | 0.3804 | OR for Public to Industry | 0.0574 | OR for Industry to I+P | 0.4980 |
| | p-value comparing papers with different funding sources | | | | | |

| | | |
|---------------------|---|---------|
| Reporting over time | SG for <2017 to >2017 | 0.1152 |
| | AC for <2017 to >2017 | *0.0372 |
| | BPP for <2017 to >2017 | 0.1557 |
| | BOA for <2017 to >2017 | 0.0866 |
| | IOD for <2017 to >2017 | 0.9310 |
| | SR for <2017 to >2017 | 0.8664 |
| | OR for <2017 to >2017 | 0.4742 |
| | p-value when comparing papers published in the past 3-years to those previous | |

| | | |
|--------------|---|---------|
| Author group | SG for <5 to >5 | *0.0140 |
| | AC for <5 to >5 | 0.1544 |
| | BPP for <5 to >5 | *0.0206 |
| | BOA for <5 to >5 | 0.1222 |
| | IOD for <5 to >5 | 0.5509 |
| | SR for <5 to >5 | 0.7845 |
| | OR for <5 to >5 | 0.2338 |
| | p-value when comparing papers published by different author group sizes | |

Legend

SG - Sequence generation

AC - Allocation concealment

BPP - Blinding of participants and personnel

BOA - Blinding of outcome assessors

IOD - Incomplete outcome data

SR - Selective reporting

OR - Other bias

IF - Impact factor

NIF - No IF

NS - Not specified

I+P - Industry & Public

* - p <0.05

** - p <0.01