

Central Lancashire Online Knowledge (CLOK)

Title	Minimum sample size estimates for trials in inflammatory bowel disease: A systematic review of a support resource
Type	Article
URL	https://clock.uclan.ac.uk/id/eprint/40263/
DOI	https://doi.org/10.3748/wjg.v27.i43.7572
Date	2021
Citation	Gordon, Morris, Lakunina, Svetlana, Sinopoulou, Vasiliki and Akobeng, Anthony (2021) Minimum sample size estimates for trials in inflammatory bowel disease: A systematic review of a support resource. World Journal of Gastroenterology, 27 (43). pp. 7572-7581. ISSN 1007-9327
Creators	Gordon, Morris, Lakunina, Svetlana, Sinopoulou, Vasiliki and Akobeng, Anthony

It is advisable to refer to the publisher's version if you intend to cite from the work.
<https://doi.org/10.3748/wjg.v27.i43.7572>

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/>

World Journal of *Gastroenterology*

World J Gastroenterol 2021 November 21; 27(43): 7402-7581



FRONTIER

- 7402** Antimicrobial peptides and the gut microbiome in inflammatory bowel disease
Gubatan J, Holman DR, Puntasecca CJ, Polevoi D, Rubin SJ, Rogalla S
- 7423** Idiopathic chronic pancreatitis: Beyond antioxidants
Mehta RM, Pandol SJ, Joshi PR

OPINION REVIEW

- 7433** COVID-19 as a trigger of irritable bowel syndrome: A review of potential mechanisms
Settanni CR, Ianiro G, Ponziani FR, Bibbò S, Segal JP, Cammarota G, Gasbarrini A

REVIEW

- 7446** Overview of the microbiota in the gut-liver axis in viral B and C hepatitis
Neag MA, Mitre AO, Catinean A, Buzoianu AD
- 7462** Hepatocellular carcinoma locoregional therapies: Outcomes and future horizons
Makary MS, Ramsell S, Miller E, Beal EW, Dowell JD

MINIREVIEWS

- 7480** Recent advances in artificial intelligence for pancreatic ductal adenocarcinoma
Hayashi H, Uemura N, Matsumura K, Zhao L, Sato H, Shiraishi Y, Yamashita YI, Baba H
- 7497** Hepatitis B: Who should be treated?-managing patients with chronic hepatitis B during the immune-tolerant and immunoactive phases
Kawanaka M, Nishino K, Kawamoto H, Haruma K

ORIGINAL ARTICLE

Basic Study

- 7509** Liver injury changes the biological characters of serum small extracellular vesicles and reprograms hepatic macrophages in mice
Lv XF, Zhang AQ, Liu WQ, Zhao M, Li J, He L, Cheng L, Sun YF, Qin G, Lu P, Ji YH, Ji JL
- 7530** Genome-wide map of N⁶-methyladenosine circular RNAs identified in mice model of severe acute pancreatitis
Wu J, Yuan XH, Jiang W, Lu YC, Huang QL, Yang Y, Qie HJ, Liu JT, Sun HY, Tang LJ
- 7546** Survivin-positive circulating tumor cells as a marker for metastasis of hepatocellular carcinoma
Yu J, Wang Z, Zhang H, Wang Y, Li DQ

Prospective Study

- 7563** Immunoglobulin G in non-alcoholic steatohepatitis predicts clinical outcome: A prospective multi-centre cohort study

De Roza MA, Lamba M, Goh GBB, Lum JHM, Cheah MCC, Ngu JHJ

SYSTEMATIC REVIEWS

- 7572** Minimum sample size estimates for trials in inflammatory bowel disease: A systematic review of a support resource

Gordon M, Lakunina S, Sinopoulou V, Akobeng A

ABOUT COVER

Editorial Board Member of *World Journal of Gastroenterology*, Serag M Esmat, MD, Professor, Department of Internal Medicine, Faculty of Medicine, Cairo University, Cairo 11562, Egypt. seragesmat@kasralainy.edu.eg

AIMS AND SCOPE

The primary aim of *World Journal of Gastroenterology* (WJG, *World J Gastroenterol*) is to provide scholars and readers from various fields of gastroenterology and hepatology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online. WJG mainly publishes articles reporting research results and findings obtained in the field of gastroenterology and hepatology and covering a wide range of topics including gastroenterology, hepatology, gastrointestinal endoscopy, gastrointestinal surgery, gastrointestinal oncology, and pediatric gastroenterology.

INDEXING/ABSTRACTING

The WJG is now indexed in Current Contents®/Clinical Medicine, Science Citation Index Expanded (also known as SciSearch®), Journal Citation Reports®, Index Medicus, MEDLINE, PubMed, PubMed Central, and Scopus. The 2021 edition of Journal Citation Report® cites the 2020 impact factor (IF) for WJG as 5.742; Journal Citation Indicator: 0.79; IF without journal self cites: 5.590; 5-year IF: 5.044; Ranking: 28 among 92 journals in gastroenterology and hepatology; and Quartile category: Q2. The WJG's CiteScore for 2020 is 6.9 and Scopus CiteScore rank 2020: Gastroenterology is 19/136.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: Yan-Xia Xing; **Production Department Director:** Xiang Li; **Editorial Office Director:** Ze-Mao Gong.

NAME OF JOURNAL

World Journal of Gastroenterology

ISSN

ISSN 1007-9327 (print) ISSN 2219-2840 (online)

LAUNCH DATE

October 1, 1995

FREQUENCY

Weekly

EDITORS-IN-CHIEF

Andrzej S Tarnawski, Subrata Ghosh

EDITORIAL BOARD MEMBERS

<http://www.wjgnet.com/1007-9327/editorialboard.htm>

PUBLICATION DATE

November 21, 2021

COPYRIGHT

© 2021 Baishideng Publishing Group Inc

INSTRUCTIONS TO AUTHORS

<https://www.wjgnet.com/bpg/gerinfo/204>

GUIDELINES FOR ETHICS DOCUMENTS

<https://www.wjgnet.com/bpg/GerInfo/287>

GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH

<https://www.wjgnet.com/bpg/gerinfo/240>

PUBLICATION ETHICS

<https://www.wjgnet.com/bpg/GerInfo/288>

PUBLICATION MISCONDUCT

<https://www.wjgnet.com/bpg/gerinfo/208>

ARTICLE PROCESSING CHARGE

<https://www.wjgnet.com/bpg/gerinfo/242>

STEPS FOR SUBMITTING MANUSCRIPTS

<https://www.wjgnet.com/bpg/GerInfo/239>

ONLINE SUBMISSION

<https://www.f6publishing.com>

Minimum sample size estimates for trials in inflammatory bowel disease: A systematic review of a support resource

Morris Gordon, Svetlana Lakunina, Vasiliki Sinopoulou, Anthony Akobeng

ORCID number: Morris Gordon 0000-0002-1216-5158; Svetlana Lakunina 0000-0002-3180-6336; Vasiliki Sinopoulou 0000-0002-2831-9406; Anthony Akobeng 0000-0001-5418-3728.

Author contributions: Gordon M conceived the study, contributed to design, analysis and writing; Lakunina S led completion, analysis and write up; Sinopoulou V and Akobeng A contributed to analysis, reviewed and edited the write up.

Conflict-of-interest statement: None to declare.

PRISMA 2009 Checklist statement: The authors have read the PRISMA 2009 Checklist, and the manuscript was prepared and revised according to the PRISMA 2009 Checklist.

Country/Territory of origin: United Kingdom

Specialty type: Gastroenterology and hepatology

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review report's scientific quality classification
Grade A (Excellent): 0
Grade B (Very good): B, B

Morris Gordon, Svetlana Lakunina, Vasiliki Sinopoulou, Biomedical Evidence Synthesis and Translation to Practice Unit, School of Medicine, Preston PR1 7BH, United Kingdom

Anthony Akobeng, Division of Gastroenterology, Sidra Med and Res Ctr, Doha 26999, Qatar

Corresponding author: Morris Gordon, MBChB, PhD, Professor, Biomedical Evidence Synthesis and Translation to Practice Unit, School of Medicine, HA340, Harrington Building, 135A Adelphi St, Preston PR1 7BH, United Kingdom. mgordon@uclan.ac.uk

Abstract

BACKGROUND

Of 25% of randomised controlled trials (RCTs) on interventions for inflammatory bowel disease (IBD) have no power calculation.

AIM

To systematically review RCTs reporting interventions for the management of IBD and to produce data for minimum sample sizes that would achieve appropriate power using the actual clinical data.

METHODS

We included RCTs retrieved from Cochrane IBD specialised Trial register and CENTRAL investigating any form of therapy for either induction or maintenance of remission against control, placebo, or no intervention of IBD in patients of any age. The relevant data was extracted, and the studies were grouped according to the intervention used. We recalculated sample size and the achieved difference, as well as minimum sample sizes needed in the future.

RESULTS

A total of 105 trials were included. There was a large discrepancy between the estimated figure for the minimal clinically important difference used for the calculations (15% group differences observed *vs* 30% used for calculation) explaining substantial actual sample size deficits. The minimum sample sizes indicated for future trials based on the 25 years of trial data were calculated and grouped by the intervention.

CONCLUSION

A third of intervention studies in IBD within the last 25 years are underpowered, with large variations in the calculation of sample sizes. The authors present a sample size estimate resource constructed on the published evidence base for

Grade C (Good): C, C

Grade D (Fair): 0

Grade E (Poor): 0

Open-Access: This article is an open-access article that was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution NonCommercial (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 and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

Received: March 4, 2021**Peer-review started:** March 4, 2021**First decision:** June 3, 2021**Revised:** June 30, 2021**Accepted:** October 31, 2021**Article in press:** October 31, 2021**Published online:** November 21, 2021**P-Reviewer:** Liefferinckx C, Vasudevan A, Yuksel I**S-Editor:** Fan JR**L-Editor:** A**P-Editor:** Fan JR

future researchers and key stakeholders within the IBD trial field.

Key Words: Inflammatory bowel disease; Crohn's disease; Ulcerative colitis; Gastroenterology; Statistics; Sample size

©The Author(s) 2021. Published by Baishideng Publishing Group Inc. All rights reserved.

Core Tip: This work has identified a large variation in the estimated minimal clinically important difference (MCID) between study groups in inflammatory bowel disease trials in the literature, with no standard to support study designers or reviewers. We have provided a resource to support sample size estimation based on observed MCID in the literature over the last 25 years.

Citation: Gordon M, Lakunina S, Sinopoulou V, Akobeng A. Minimum sample size estimates for trials in inflammatory bowel disease: A systematic review of a support resource. *World J Gastroenterol* 2021; 27(43): 7572-7581

URL: <https://www.wjgnet.com/1007-9327/full/v27/i43/7572.htm>

DOI: <https://dx.doi.org/10.3748/wjg.v27.i43.7572>

INTRODUCTION

Sample size estimation (SSE) is an extremely important calculation for designing a clinical trial. Failure to produce an appropriate calculation may lead to imprecise results[1]. If a sample size is too large, statistically significant outcomes may be theoretically detected that may not be clinically relevant (type 1 error). This, however, is rarely a concern as studies are rarely overpowered to balance the study power with the cost. On the other hand, if a sample size is too small then a clinically significant outcome may not be detected statistically (type 2 error)[2,3]. The reporting of SSE in randomised controlled trials (RCTs) is a standard requirement according to the consolidated standards of reporting trials (CONSORT) statement which was introduced as a guide to conducting RCTs in 1996[4].

In a previous systematic review[5], we showed that 25% of RCTs on interventions for inflammatory bowel disease (IBD) have no power calculation (PC). A third of those who report PC do not achieve their target sample size. Based on those results, we decided to conduct a further systematic review.

We set out to systematically review RCTs on interventions for the IBD management, extract the vital parameters needed for sample size calculations, and synthesise the data to demonstrate whether trials across the field are adequately powered. We also set out to use the actual clinical data across these comparisons to synthesise data for minimum sample sizes that would achieve appropriate power to support future researchers designing trials and performing SSEs.

MATERIALS AND METHODS

This review was performed in alignment with Cochrane guidelines[6] in April 2020 and reported in line with the Preferred Reporting Items for Systematic Reviews and Meta-analyses statement[7].

Eligibility criteria

We followed the sampling methodology described within our systematic review protocol (uploaded within our institutional repository)[8] used for our previous review of the reporting of sample size calculations[5].

In brief, we included RCTs investigating either induction or maintenance therapy with biologics, immunomodulators, and microbiome against control, placebo, or no intervention. We conducted a comprehensive search of the Cochrane IBD Specialized Trials Register, CENTRAL, Cochrane library of IBD reviews for primary RCTs. The search terms are presented in [Supplementary material](#).

We included RCTs published since 1996 (after the publication of the CONSORT statement). We excluded reports lacking clear information on the number of participants; cluster RCTs; pilot or feasibility studies; studies with mixed population of people with and without IBD; studies on secondary analyses of follow-up data collection after discontinuation of treatment. We excluded abstracts as these rarely allow space for such information to be presented. As we wanted to assess the established evidence for a PC of treatment for the IBD, we excluded RCTs describing all interventions where work may be at phase 3 (pharmacological: *e.g.* ustekinumab, golimumab, tofacitinib) or not under the three core headings (biologic, immunomodulators or anti-inflammatories).

Complying to the above search strategy, two authors (SL and MG) identified RCTs titles that appeared to be applicable. These were independently screened and in cases of disagreement, a third review author (VS) was involved to reach consensus. Two review authors independently extracted and recorded data on a predefined checklist. When disagreements occurred, a third review author was involved, and the consensus was reached.

We created an excel document to extract data regarding the trials. Firstly, we separated the studies into 8 categories [Crohn's disease (CD)–clinical relapse, clinical remission, endoscopic relapse, endoscopic remission; ulcerative colitis (UC)–clinical relapse, clinical remission, endoscopic relapse, endoscopic remission]. Secondly, we grouped the studies according to the intervention used. One author extracted the data, and in case of any problems, the data was checked by the second author.

The extracted data although is not available publicly can be obtained *via* direct contact with authors. The references of the included studies can be found in [Supplementary material](#).

Extracted data included

(1) Number of events and participants originally assigned to each group; (2) Characteristics of participants; (3) The proportion that we calculated according to the number of events and participants ($x = n/N$), in which n is a number of events and N is a number of participants); (4) The difference achieved that we calculated according to the proportions of two groups (proportion 1–proportion 2); (5) Intervention and control details; (6) Presence of SSE and calculation details [minimal clinically important difference (MCID) used for PC, power, significance level, target sample size]; and (7) Outcomes (the number of patients recruited and completing study; the number of treatment success/failures; and the difference achieved).

We used the studies in which intervention was compared to the control or placebo. We grouped those studies according to the interventions, type of treatment (induction, maintenance), and outcomes (relapse, remission) and calculated mean difference and mean MCID where it was possible.

After resolving all the inconsistencies with data extraction regarding the use of sample size calculations for the studies with achieved difference of less than 10%, we produced two tables (Tables 1 and 2). We recalculated sample size for those groups using the power of 80%, probability of type 1 error 0.05, and the achieved difference. We used those parameters as they were the most commonly used amongst the studies. The parameters we used were two independent groups, dichotomous outcomes. In group 1 we have put the rate reported by the study of the intervention drug, and in group 2 we have put the rate of the placebo.

The smalllest MCID that was reported by the studies was 10%, thereby, we decided to not reproduce PC for those studies with the achieved difference of less than 10%. We also calculated the mean sample deficit in percentage based on the target sample size and achieved sample size reported by the studies.

After receiving the sample size of participants, we made a decision whether the study is underpowered, and if yes, then by how many people.

Data synthesis

We produced descriptive statistics regarding the sample sizes for the studies grouped according to the interventions (Tables 1 and 2).

Ethical statement

As all data included already existed within the published scholarly output, no ethical approval was sought.

Table 1 Overall summary of power calculations and sample size deficits

	Total studies	Studies with power calculation	Studies with difference of 10% and less ¹	How many studies didn't achieve target sample size	Mean sample size underpowered (range)	Mean sample size needed	How many studies are underpowered ²
CD induction	39	26	12	6	28 (2-70)	231	11
CD maintenance	25	19	9	3	52 (7-79)	300	10
UC induction	27	19	8	3	22 (1-55)	219	4
UC maintenance	16	10	0	1 + 1 didn't report	21	196	7

¹Those studies were not included in analysis.²Either didn't achieve their target sample size, or their achieved sample size is less than mean sample size needed.

UC: Ulcerative colitis; CD: Crohn's disease.

RESULTS

A total of 7451 potential citations were screened and 308 full texts assessed for eligibility. There were 209 texts excluded, 106 because they were published prior to the release of the CONSORT statement and 103 because they did not match our inclusion outcome. This left a total of 99 trials included, with 60 pertaining to CD and 39 to UC. The full details are shown in [Figure 1](#).

The mean proportion of patients achieving clinical remission reported within the placebo groups of induction studies was 34.34% in CD trials and 26.79% for UC. For endoscopic remission, 0% in CD and 29.6% for UC. The mean proportion of patients achieving clinical relapse for maintenance studies were 55% for CD and 46.79% for UC. For endoscopic relapse, 78.85% in CD, and 28.7% in UC.

Within CD induction studies, 26 out of 41 (63.4%) reported a PC and 19 of 26 (73.1%) in maintenance studies. Within UC induction studies, 22 out of 31 (71%) reported a PC and 10 of 17 (58.8%) in maintenance studies.

When considering the MCID that those studies reporting a PC employed for this calculation, within CD induction studies the mean difference was 33% (range 20%-50%) and 27% difference for maintenance studies (15%-40%). Within UC induction studies the mean was 26% (range 19%-40%) and 27% for maintenance studies (18%-40%). The MCIDs these studies reported rarely matched the actual differences achieved by these studies. In fact, the discrepancy between this estimated figure for the MCID used for the PC and the actual differences seen were a mean of 22.8% higher in CD induction studies, 13.8% higher in maintenance studies, 15.7% higher in UC induction studies, and 10.2% higher in maintenance studies.

These discrepancies are proportionally large and in the context of PCs are clearly substantial and led to large numbers of studies being underpowered. These are summarised in [Table 1](#). Study specific data with further details is available upon request.

[Table 2](#) gives the results of our sample size calculations at the intervention specific level that employed the actual achieved clinical differences from previous studies, using the power of 80% and the probability of type 1 error 0.05. This shows the minimum sample sizes that would be indicated for RCTs compared with placebo to use. Within comparisons where the mean difference was less than 10%, no calculation has been given as this would be a very high indicative figure.

DISCUSSION

Within this review, it has been demonstrated that there is no clear basis or accepted standard for current practice for MCID estimation when producing a PC for a primary RCT within IBD. This has led to huge variations in suggested figures for recruitment. These trials present practical and logistical challenges to organisers, with potential inconvenience to patients, as well as the cost to those funding such research. Having an accurate figure for calculations is important to ensure this investment of resource is used most efficiently and effectively. It is key to note that we are not commenting at

Table 2 Proposals for minimum clinically important difference and associated power calculations for future studies

Ulcerative colitis-comparison	Ulcerative colitis-difference achieved (Group 1–Placebo)	Ulcerative colitis-Minimum sample size needed based on data	Crohn's disease-comparison	Crohn's disease-difference achieved (Group 1–Placebo)	Crohn's disease-Minimum sample size needed based on data
Induction studies					
Outcome–clinical remission			Outcome–clinical remission		
Vedolizumab <i>vs</i> Placebo	14.8%	190	Glutamine-enriched diet <i>vs</i> Placebo	-11.1	634
			Azathioprine <i>vs</i> Placebo	-3.6%	NA
			6-MP <i>vs</i> Placebo	5%	NA
Fecal Transplant <i>vs</i> Control	20.3%	150	6-MP <i>vs</i> Placebo	5%	NA
Budesonide <i>vs</i> Placebo	6.5%	NA	Interventional diet <i>vs</i> Control diet	20.9%	160
Type 1 IFNs <i>vs</i> Placebo	5.9%	NA	Elemental diet <i>vs</i> Non elemental diet	1.6%	NA
Etrolizumab <i>vs</i> Placebo	13.4%	140	N6/N9 rich feeds <i>vs</i> non N6/N9 rich food	-1.1%	NA
			Low dose naltrexone <i>vs</i> Placebo	9%	NA
5-ASA <i>vs</i> Placebo	11.8%	422	GM-CSF <i>vs</i> Placebo	7.8%	NA
Outcome–endoscopic remission			Brakinumab <i>vs</i> Placebo	8.5%	NA
Vedolizumab <i>vs</i> Placebo	37.7%	182	Ustekinumab <i>vs</i> Placebo	8.6%	NA
			Natalizumab <i>vs</i> Placebo	14.8%	310
Fecal Transplant <i>vs</i> Control	26.4%	160	Methotrexate <i>vs</i> Placebo	-14.8%	350
Budesonide <i>vs</i> Placebo	13.9%	NA	Antibiotics <i>vs</i> Placebo	10%	780
Methotrexate <i>vs</i> Placebo	46.7%	NA	Outcome–endoscopic remission		
Etrolizumab <i>vs</i> Placebo	7.7%	NA	Low dose naltrexone <i>vs</i> Placebo	22.2%	60
5-ASA <i>vs</i> Placebo	53.7%	306			
Maintenance studies					
Outcome–clinical relapse			Outcome–clinical relapse		
5-ASA <i>vs</i> Placebo	-16.4%	290	5-ASA <i>vs</i> Placebo, medically induced	3.1%	NA
Vedolizumab <i>vs</i> Placebo	-27.4	84	5-ASA <i>vs</i> Placebo, surgically induced	-5.4%	NA
Interventional diet <i>vs</i> Control diet	-3.6%	NA	Anti-TB <i>vs</i> Placebo	-23%	130
Probiotics <i>vs</i> Control	-16.7	154	Azathioprine <i>vs</i> Placebo, medically induced	-9.9%	NA
Azathioprine <i>vs</i> Placebo	-22.4	154	Azathioprine <i>vs</i> Placebo, surgically induced	-17.3%	254
Methotrexate <i>vs</i> Placebo	19.9%	194	6-MP <i>vs</i> Placebo, surgically induced	-10.9%	646
Rectal 5-ASA <i>vs</i> Placebo	-29%	90	Omega -3 fatty acids diet <i>vs</i> Control diet	-8.5%	NA
Curcumin <i>vs</i> Placebo	-9.6%	NA	Elemental diet <i>vs</i> No supplements	-29.4%	88

Outcome–endoscopic relapse			Interventional diet <i>vs</i> Control diet	-2.5%	NA
Vedolizumab <i>vs</i> Placebo	-34	60	Antibiotics <i>vs</i> Placebo	-14.6%	360
			Methotrexate <i>vs</i> Placebo	-24.2%	128
5-ASA <i>vs</i> Placebo	-16.4%	290	Methotrexate <i>vs</i> Placebo	-24.2%	128
			Outcome–endoscopic relapse		
			5-ASA <i>vs</i> Placebo	2.7%	NA
			Azathioprine <i>vs</i> Placebo	-23%	130
			6-MP <i>vs</i> Placebo	-3.8%	NA
			Antibiotics <i>vs</i> Placebo	6.6%	NA
Induction studies					
Outcome–clinical remission			Outcome–clinical remission		
Vedolizumab <i>vs</i> Placebo	14.8%	190	Glutamine-enriched diet <i>vs</i> Placebo	-11.1	634
			Azathioprine <i>vs</i> Placebo	-3.6%	NA
			6-MP <i>vs</i> Placebo	5%	NA
Fecal Transplant <i>vs</i> Control	20.3%	150	6-MP <i>vs</i> Placebo	5%	NA
Budesonide <i>vs</i> Placebo	6.5%	NA	Interventional diet <i>vs</i> Control diet	20.9%	160
Type 1 IFNs <i>vs</i> Placebo	5.9%	NA	Elemental diet <i>vs</i> Non elemental diet	1.6%	NA
Etrolizumab <i>vs</i> Placebo	13.4%	140	N6/N9 rich feeds <i>vs</i> non N6/N9 rich food	-1.1%	NA
			Low dose naltrexone <i>vs</i> Placebo	9%	NA
5-ASA <i>vs</i> Placebo	11.8%	422	GM-CSF <i>vs</i> Placebo	7.8%	NA
Outcome–endoscopic remission			Brakinumab <i>vs</i> Placebo	8.5%	NA
Vedolizumab <i>vs</i> Placebo	37.7%	182	Ustekinumab <i>vs</i> Placebo	8.6%	NA
			Natalizumab <i>vs</i> Placebo	14.8%	310
Fecal Transplant <i>vs</i> Control	26.4%	160	Methotrexate <i>vs</i> Placebo	-14.8%	350
Budesonide <i>vs</i> Placebo	13.9%	NA	Antibiotics <i>vs</i> Placebo	10%	780
Methotrexate <i>vs</i> Placebo	46.7%	NA	Outcome–endoscopic remission		
Etrolizumab <i>vs</i> Placebo	7.7%	NA	Low dose naltrexone <i>vs</i> Placebo	22.2%	60
5-ASA <i>vs</i> Placebo	53.7%	306			
Maintenance studies					
Outcome–clinical relapse			Outcome–clinical relapse		
5-ASA <i>vs</i> Placebo	-16.4%	290	5-ASA <i>vs</i> Placebo, medically induced	3.1%	NA
Vedolizumab <i>vs</i> Placebo	-27.4	84	5-ASA <i>vs</i> Placebo, surgically induced	-5.4%	NA
Interventional diet <i>vs</i> Control diet	-3.6%	NA	Anti-TB <i>vs</i> Placebo	-23%	130
Probiotics <i>vs</i> Control	-16.7	154	Azathioprine <i>vs</i> Placebo, medically induced	-9.9%	NA

Azathioprine <i>vs</i> Placebo	-22.4	154	Azathioprine <i>vs</i> Placebo, surgically induced	-17.3%	254
Methotrexate <i>vs</i> Placebo	19.9%	194	6-MP <i>vs</i> Placebo, surgically induced	-10.9%	646
Rectal 5-ASA <i>vs</i> Placebo	-29%	90	Omega -3 fatty acids diet <i>vs</i> Control diet	-8.5%	NA
Curcumin <i>vs</i> Placebo	-9.6%	NA	Elemental diet <i>vs</i> No supplemets	-29.4%	88
Outcome–endoscopic relapse			Interventional diet <i>vs</i> Control diet	-2.5%	NA
Vedolizumab <i>vs</i> Placebo	-34	60	Antibiotics <i>vs</i> Placebo	-14.6%	360
5-ASA <i>vs</i> Placebo	-16.4%	290	Methotrexate <i>vs</i> Placebo	-24.2%	128
			Methotrexate <i>vs</i> Placebo	-24.2%	128
			Outcome–endoscopic relapse		
			5-ASA <i>vs</i> Placebo	2.7%	NA
			Azathioprine <i>vs</i> Placebo	-23%	130
			6-MP <i>vs</i> Placebo	-3.8%	NA
			Antibiotics <i>vs</i> Placebo	6.6%	NA

NA is put when the difference achieved is less than 10% (which is the least Minimal Clinically Important Difference used by the studies).

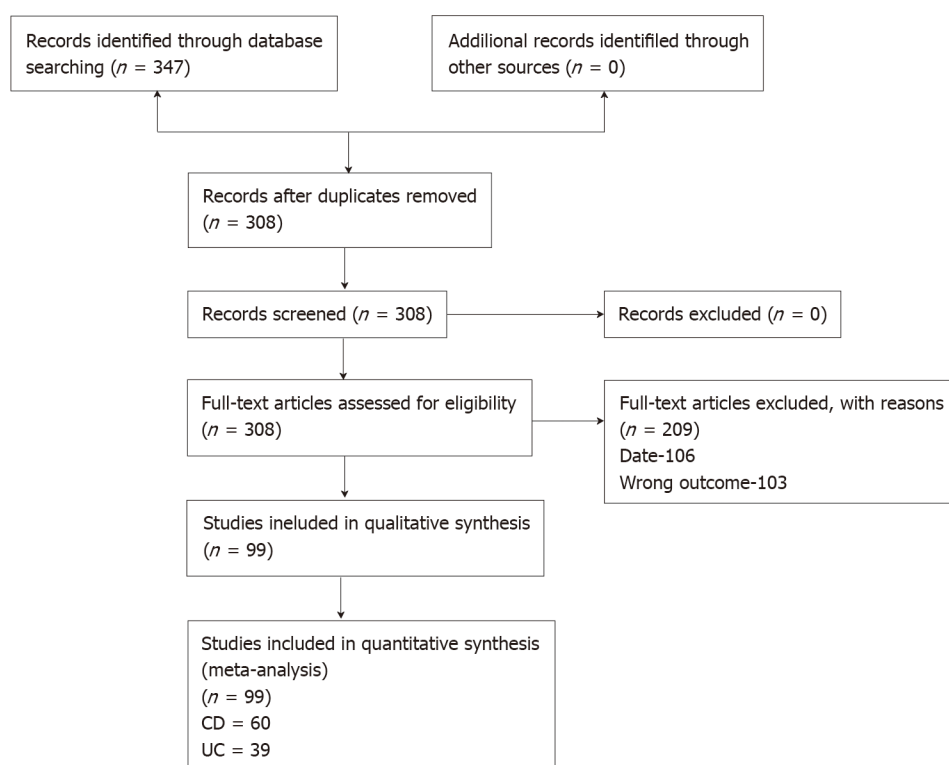


Figure 1 Study flow diagram. UC: Ulcerative colitis; CD: Crohn's disease.

the individual study level. It is inappropriate to look at the projected MCID and PC for a project, if calculated on a reasonable basis, to then retrospectively suggest that the findings of a lesser MCID mean it is underpowered. This not just statistically inappropriate, but methodologically flawed. However, these findings propose that the basis for such MCID estimations is at worst unclear and often can be seen as flawed.

There are further ethical issues these problems raise, such as being forced to give treatments to people without having a statistically proved effect or a high certainty result within the Grading of Recommendations Assessment, Development and

Evaluation analysis (due to reasons of imprecision from statistical sampling issues). The power of a study, therefore, has huge implications on the precision of estimates in the future analysis of data and in turn clinical practice guidelines. Within this review, 30% of studies appeared to be underpowered based on actual achieved clinical differences within the wider comparable evidence base, with mean sample size deficits up to 79 patients per trial. This does impact the overall certainty of the global evidence base within IBD, with precision a key limitation downgrading many outcomes within key guidelines across dozens of interventions.

Within this review, we present a resource for SSE not just for future study authors, but for study peer reviewers and most importantly professionals and the patients. This table gives an estimated PC result for a minimum sample size based on all existing studies within this period. Rather than being based on just single studies or clinical judgement, these represent estimates based on actual achieved clinical data and to our knowledge are the first time such a resource has ever been provided for researchers in the field or indeed for readers of future research. Additionally, for those wishing to calculate key statistics and measures of outcome from their primary studies, this paper provides a systematic and objective resource for baseline risk. This could be used for calculating numbers needed to treat or harm, for example.

This resource can be used by study designers to prevent PCs based on studies that offer a high MCID and as such a lower minimum sample size than is actually warranted. Conversely, it prevents unnecessary over recruitment. Funders can use this to appropriately budget and ensure viability of studies. Ethics boards and other governance groups will be able to consult this resource to support their consideration of research proposals.

There were a number of comparisons where the difference in practice was below 10% and it was deemed inappropriate to make a calculation in such cases, as no previous study has ever indicated an MCID below 10% as clinically significant to patients or practice. In these cases, consideration should be given to the overall figures presented in [Table 2](#) or minimum sample size and MCID in practice in a similar context.

We would also recommend that in practice, patients and key stakeholders should be involved in deciding on an MCID for a given intervention prior to a new study. They may indicate that in spite of any existing MCID evidence that such a difference is not significant enough to matter to those who are most impacted by the findings and such views must be reflected in the process of SSE. It is also worth noting that there will always be settings and contexts when deviation may be warranted, thereby, a resource is not prescriptive but rather presented as evidence-based guidance. We would, however, propose that such deviations can and should be justified to support transparency for the readings these trials report.

There are weaknesses and exceptions to these approaches. The search methods used limited the parameters of the search for pragmatic reasons. However, this does not represent any systematic bias, hence we do not believe it invalidates the findings, and in the future this resource can be updated prospectively. When the achieved difference was less than 10%, rather than reporting extremely large sample size calculations, no such calculation was made. Additionally, in studies comparing active agents, accurate estimates are needed based on the contexts as the hypothesis may not be of the inferiority or superiority but of no difference, which requires a different approach to calculations.

There were some limitations to this review. There are obvious issues of heterogeneity limiting the appropriateness of pooling the data, however, the only way to obtain the previously used MCID was through looking at the past studies. These are mainly related to missing or unclear information in primary studies regarding SSE and as authors were not contacted, assumptions were made for the basis of these calculations which could confer some inaccuracy in our estimations. We also limited our studies to those from after the CONSORT statement release as we felt this was a fair time from which to expect SSE to occur, but earlier studies could potentially have offered more insight. Finally, we have focussed on studies comparing treatment with placebo or no intervention. This was a pragmatic decision as many studies of agents choose to make this comparison, although often these do not reflect current standard clinical practice. In the cases of such comparisons, SSE may not have to be based on a MCID but instead assume clinical equivalency and therefore be informed differently. In essence, this guidance may not be relevant for these scenarios, although may inform statistical considerations within similar contexts. Finally, such a resource of course is likely to become inaccurate rapidly, with the need for updates, but as often no such resource is employed, we believe this is still an improvement on current practices.

Future research is needed to potentially validate the calculations with clinical and patient input to ensure the SSE and MCID that the data informs has clinical, as well as statistical relevance. This could lead to a more triangulated resource that is statistically and evidentially sound, but also clinically sound and patient informed. This could conceivably lead to increases or decreases in minimally important differences to reflect complexity in specific clinical scenarios and interventional contexts.

CONCLUSION

In conclusion, a third of intervention IBD studies within the last 25 years are underpowered, with large variations in the approaches to calculating sample sizes and the minimum clinically important differences. The authors present a sample size estimate resource based on the published evidence base for future researchers and other key stakeholders within the IBD trial field.

ARTICLE HIGHLIGHTS

Research background

A third of randomised controlled trials (RCTs) on interventions for inflammatory bowel disease (IBD) have no adequate power calculation (PC).

Research motivation

A key element of PCs is an estimation of a minimally important clinical difference. The basis of these is capricious within the literature, with many not based on any existing or prior studies and as such can lead to massive shifts in PCs for similar studies, with concerns as to the underlying power.

Research objectives

We systematically reviewed RCTs reporting interventions for the management of IBD and to produced a resource for minimum clinically important difference using clinical data for the future researchers to use as a starting point.

Research methods

We included RCTs retrieved from Cochrane IBD trial register and CENTRAL investigating anti-inflammatory, immunomodulator and biologic therapies for either induction or maintenance of remission against control, placebo, or no intervention of IBD in patients of any age. The data was extracted and synthesized. We recalculated sample size and the achieved difference, as well as minimum sample sizes and presented in a tabular format.

Research results

Of 105 trials were included. A large discrepancy between the estimated figure for the minimal clinically important difference used for the calculations (15% differences observed *vs* 30% used for calculation) was observed explaining substantial actual sample size deficits. The minimum sample sizes indicated for future trials based on the 25 years of trial data were calculated and grouped by the intervention.

Research conclusions

There are large variations in the sample size calculations in the studies of interventions for IBD with a third of all studies being underpowered. The authors present a sample size estimate resource constructed on the published evidence base for future researchers and key stakeholders within the IBD trial field.

Research perspectives

The use of this resource will support research staff, ethics committees and journal editors in ensuring adequate sample sizing and powering of studies across the field.

REFERENCES

- 1 **Biau DJ**, Kernéis S, Porcher R. Statistics in brief: the importance of sample size in the planning and interpretation of medical research. *Clin Orthop Relat Res* 2008; **466**: 2282-2288 [PMID: [18566874](#) DOI: [10.1007/s11999-008-0346-9](#)]
- 2 **Nayak BK**. Understanding the relevance of sample size calculation. *Indian J Ophthalmol* 2010; **58**: 469-470 [PMID: [20952828](#) DOI: [10.4103/0301-4738.71673](#)]
- 3 **Cornish R**. Statistics: An Introduction To Sample Size Calculations. Mathematics Learning Support Centre, 2006: 1-5. [cited 10 January 2021]. Available from: <https://www.lboro.ac.uk/departments/mlsc/>
- 4 **Begg C**, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996; **276**: 637-639 [PMID: [8773637](#) DOI: [10.1001/jama.276.8.637](#)]
- 5 **Iheozor-Ejiofor Z**, Lakunina S, Gordon M, Akintelure D, Sinopoulou V, Akobeng A. Sample-size estimation is not reported in 24% of randomised controlled trials of inflammatory bowel disease: A systematic review. *United European Gastroenterol J* 2021; **9**: 47-53 [PMID: [33104495](#) DOI: [10.1177/2050640620967899](#)]
- 6 **Higgins JPT**, Thomas J. Cochrane Handbook for Systematic Reviews of Interventions. In: Chandler J, Cumpston M, Li T, Page MJ, Welch VA. Chichester: John Wiley and Sons, 2019
- 7 **Moher D**, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009; **6**: e1000097 [PMID: [19621072](#) DOI: [10.1371/journal.pmed.1000097](#)]
- 8 **Gordon M**, Lakunins S. [Protocol for systematic review 'The Reporting of Sample Size Estimation in Randomised Trials of Inflammatory Bowel Disease: A systematic review'.] [cited 10 January 2021]. Available from: <http://clock.uclan.ac.uk/33088>



Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA

Telephone: +1-925-3991568

E-mail: bpgoffice@wjgnet.com

Help Desk: <https://www.f6publishing.com/helpdesk>

<https://www.wjgnet.com>

