Systematic Review of Volume and Methodological Quality of Randomized Trials in Acute Pancreatitis

Nejo Joseph, MBChB,† Gordon Liu, MBChB,*† Chris Varghese, MBChB,†‡ Wei Lim, MBBS,§ William Xu, MBChB,†|| Cameron I. Wells, MBChB,¶ Shayne Tobias, MBBS,# Gabriele Capurso, PhD,** Enrique de Madaria, PhD,††‡‡ Asbjørn Drewes, PhD,§§ Marc G. Besselink, PhD,||||¶¶ John Windsor, FRACS,†## and Sanjay Pandanaboyana, FRCS, MPhil||***†††*

Background: This systematic review assessed the volume and methodological quality of randomized controlled trials (RCTs) in relation to management of acute pancreatitis (AP).

Materials and Methods: The PubMed, MEDLINE, and CENTRAL databases were systematically searched for RCTs published across 3 time periods: <1996 (P1), 1996–2008 (P2), and >2008 (P3). RCT quality was assessed using the Cochrane Risk of Bias (RoB) 2 tool and sample size recalculation, and for spin (interpretation of nonstatistically significant results as relevant, making the study appear to be positive).

Results: Overall, 263 RCTs with 23,232 patients with AP were included. The average number of RCTs per year increased from 1.4, 6.0, to 10.6 in P1, P2, and P3, respectively. The RoB assessment showed *low*, *some*, and *high concerns* in overall RoB in 21%, 56%, and 24% of all RCTs. Selective reporting bias improved over time. Sample size calculation reporting significantly increased through the 3 time periods (17%, 38%, and 47%; $P < 0.001$). Spin was identified in 68 RCTs (26% of all RCTs).

Conclusion: The quantity and quality of published RCTs relating AP management has increased over time, however significant shortcomings of methodological quality persist. Significant improvements in the conduct and reporting of randomized trials in AP are required to improve the evidence base in this field.

Key Words: acute pancreatitis, quality assessment

(*Pancreas* 2025;54: e82–e88)

The incidence and hospital admission rates for acute pancreatitis (AP) have continued to increase over the last 30 years.[1–3] AP represents a significant health and cost burden necessitating ongoing improvements in its clinical management. Randomized controlled trials remain the key modality to accrue new evidence and drive improvements in clinical management and outcomes of AP.[4,5]

Progress in the quality of evidence available has been made in certain management domains for AP, including for nutritional support,[6] antibiotic use,[7] and treatment of necrotizing pancreatitis,[8] but ongoing shortage of level 1 evidence exists in AP analgesia and other domains of AP management.[9,10] As increasing numbers of trials are initiated in AP, it is prudent to formally assess the quantity of trials published to date across management domains, and their methodological quality, to inform and improve the conduct and reporting of future trials.

We considered the domain and subject matter of the trials and assessed their quality through 3 parameters: (*a*) risk of bias, (*b*) sample-size reporting, and (*c*) use of spin. A review of the quality of RCTs conducted in pancreatic cancer surgery using the Cochrane risk of bias (ROB) tool showed a trend toward reduced risk of bias over time.[11] Reviews of RCTs investigating other biomedical domains have previously shown inadequate and often inaccurate reporting of sample size calculations.[12] The presence of spin (the "use of specific reporting strategies, from whatever motive, to highlight that the experimental treatment is beneficial, despite a statistically non-significant difference for the primary outcome, or to distract the reader from statistically non-significant results"[13]) is another detriment to reporting of trial results. This quality parameter has been formally assessed in other medical disciplines with up to 75% trials being affected by spin; however, spin has not been previously analyzed for RCTs of AP management.[14,15]

Detailed guidelines exist for the design of RCTs and validated assessment tools for the reporting of methodological quality.[16] The aim of this study was to systematically review the quantity and quality of RCTs relating to the management of AP over time. Quality of trials was assessed using the RoB-2 tool and reviewing the completeness of sample size reporting and Spin. By doing so, we expect to identify gaps in evidence that will guide future research in AP.

MATERIALS AND METHODS

A systematic review was conducted in accordance with PRISMA 2020 guidelines (Appendix 1, http://links.lww.com/MPA/B231).[17] MEDLINE, Embase, and CENTRAL were

From the *Faculty of Medical and Health Sciences, and †Surgical and Translational Research Centre, Faculty of Medical and Health Sciences, University of Auckland; ‡Middlemore Hospital, Counties Manukau District Health Board, Auckland, New Zealand; §Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom; ||Whangarei Hospital, Te Tai Tokerau, Whangarei; ¶Department of General Surgery, Te Toka Tumai Auckland, Auckland, New Zealand; #School of Medicine and Medical Sciences, University College Dublin, Dublin, Ireland; **Pancreatico-Biliary Endoscopy and Endosonography Division, Pancreas Translational and Clinical Research Centre, San Raffaele Scientific Institute IRCCS, Vita-Salute San Raffaele University, Milan, Italy; ††Gastroenterology Department, Dr. Balmis General University Hospital; ‡‡Department of Clinical Medicine, Miguel Hernández University, ISABIAL, Alicante, Spain; §§Department of Gastroenterology, Centre for Pancreatic Diseases, Aalborg University Hospital, Aalborg, Denmark; ||||Amsterdam UMC, location University of Amsterdam, Department of Surgery; ¶¶Amsterdam Gastroenterology Endocrinology Metabolism, Amsterdam, the Netherlands; ##HBP/Upper GI Unit, Te Toka Tumai Auckland, Auckland, New Zealand; ***HPB and Transplant Unit, Freeman Hospital, Newcastle upon Tyne Hospitals NHS Foundation Trust; and †††Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom.

Received for publication January 22, 2024; accepted April 24, 2024.

Address correspondence to: Sanjay Pandanaboyana, FRCS, MPhil, HBP and Transplant Unit, Freeman Hospital, Freeman Road, Newcastle Upon Tyne, NE7 7DN, UK (e-mail: sanjay.pandanaboyana@ncl.ac.uk).

The authors declare no conflict of interest.

This research did not receive any specific grant from funding agencies in the public, commercial, or for non-profit sectors.

Ethics committee approval: Not applicable.

Data Availability: All template data forms, data extracted from included studies, data used in analysis, and analytic code are available upon reasonable request.

Supplemental digital contents are available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.pancreasjournal.com).

Copyright © 2024 Wolters Kluwer Health, Inc. All rights reserved.

DOI: 10.1097/MPA.0000000000002397

Copyright © 2025 Wolters Kluwer Health, Inc. All rights reserved.

searched up to September 2022. No restrictions were placed on the date of publication. The search strategy comprised medical subject heading (MeSH) terms and text words (tw) combined with Boolean operators "AND" and "OR." The final search strategy was as follows: "Acute pancreatitis" OR "Acute" AND ("pancrea* inflammation"). An a priori protocol was developed; however, this review did not meet the requirements for prospective registration in PROSPERO.

Eligibility Criteria

All randomized controlled trials (RCTs) assessing the treatment of acute pancreatitis (AP) were included. RCTs investigating the prevention of AP, or the treatment of chronic pancreatitis or pancreatic malignancies were excluded. Manuscripts in languages other than English, trial protocols, non-RCTs, quasi-RCTs, and conference abstracts were also excluded. Four reviewers screened the titles and abstracts of all retrieved references for inclusion, such that each record was reviewed independently by at least 2 reviewers. Disagreements were resolved by consensus with a third independent reviewer. The full texts of all articles with consensus agreement for inclusion by at least 2 reviewers were retrieved and assessed.

Data Collection

Data extraction was conducted using a prespecified proforma that was hosted on Google Forms (available on request). The following predefined data were acquired: year of publication, journal, general topic, region of publication, grading of AP severity, tool used for AP severity grading, sample size, spin techniques, sample size calculation variables, analysis population, and quality features as described below.

Assessment of Risk of Bias

The Cochrane Risk-of-Bias 2 (RoB-2) tool as described by the current version of the Cochrane handbook was used to assess the methodological quality of the included trials.[18] The tool evaluates bias in the following dimensions: randomization process, allocation of participants, maintenance of participants, measurement of outcome, and reporting of results. Four reviewers assessed risk of bias, and each study was reviewed independently by at least 2 reviewers. Disagreements were resolved by consensus with a third independent reviewer.

"Spin" Analysis

We use the definition of spin proposed by Boutron and colleagues as "use of specific reporting strategies, from whatever motive, to highlight that the experimental treatment is beneficial, despite a statistically nonsignificant difference for the primary end point, or to distract the reader from statistically nonsignificant results."[13] This definition was used to evaluate the presence of spin in the title, abstract, or main text of the trial. There are 8 intentional strategies with spin, which seeks to highlight the benefits of treatment by emphasizing: (1) nonsignificant trends in the primary end point; (2) secondary end points; (3) subgroup analyses; (4) secondary analysis of primary end point, such as by changing the analysis population or measuring the treatment effect; (5) intragroup comparisons, including before and after treatment; (6) a lack of any mention of a treatment's unclear safety profile; (7) safety alone in the absence of any significant results for primary end point; and (8) for any other situations deemed as spin by the reviewers, denoted as "other." Spin was independently assessed by multiple authors with conflicts being resolved by consensus opinion.

Sample Size Recalculation

Two reviewers (N.J. and G.L.) recalculated the sample size for each RCT that provided complete and correct descriptions of the sample for sample size calculation and statistical analysis of the end points. Recalculation of the reported sample sizes was done using the software G*Power version 3.1.9.7.[19] If multiple comparisons were conducted, a Bonferroni correction was applied. If attrition was estimated, it was included in the sample size recalculation. The sample size calculation of an RCT was considered accurate when the reported sample size could be reproduced to ±2 persons per trial arm of the sample size as declared by the trial authors. Sample size calculations that failed to report the following details were considered incomplete: 1- or 2-tailed test and choice of statistical test(s).

Statistical Analysis

The RCTs were divided into terciles according to date of publication: before 1996 (period 1 [PI]), 1996–2008 (P2), and 2008 onward (P3). In calculations requiring yearly averages for each period, 1965 was taken as the first year of P1 as this was the publication year of the oldest trial in this dataset. The 3 periods were selected on the basis of milestones in the conduct and reporting of RCTs: the publication of the CONSORT statement in 1996[16] and the incorporation of the mandatory registration of prospective clinical trials into the Declaration of Helsinki in 2008.[20]

Categorical data were presented as number and percentage and analyzed using χ^2 test, and ordinal tests of association were used for analysis of proportions as appropriate. Continuous data were presented as medians, quartiles, and ranges. R version 4.2.0 (R Foundation for Statistical Computing, Vienna, Austria) was used for analyses. The following packages on R were used in analysis: *tidyverse*, *finalfit*, and *ggplot*. Two-sided $P < 0.050$ was considered significant.

Recalculation of Sample Size

Relative deviation (%) of the reported and recalculated sample sizes was calculated via:

$$[\text{sample size(recalculated)} - \text{sample size(reported)}]/\text{sample size(recalculated)} \times 100$$

Evidence Mapping

The methods of Hüttner and colleagues[21] were used for evidence mapping. Bubble plots were created, mapping all RCTs by reported severity of pancreatitis against type of intervention. The sample size of the trials is represented by bubble size, and the geographical regions are color coded.

RESULTS

Trial Characteristics

In total, 5380 publications were identified by the systematic search of 3 databases. After removing duplications and full-text review, 263 RCTs were eligible for inclusion (Fig. 1). The median number of patients included was 64 (range, 11–508). The median numbers of participants were similar across periods (66.5, 58, and 67; $P = 0.09$). The number of RCTs increased with time (42, 72, then 149). The average number of publications per year for each period also increased: 1.4, 6.0, and 10.6 (Fig. 2).

Most RCTs were conducted in Asia with (n = 124, 47%), Europe (92, 35%), and North America (n = 40, 15%). The remaining 7 RCTs were conducted in South America (n = 5) and Australia/New Zealand (n = 2). The proportion of RCTs

© 2024 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2025 Wolters Kluwer Health, Inc. All rights reserved.

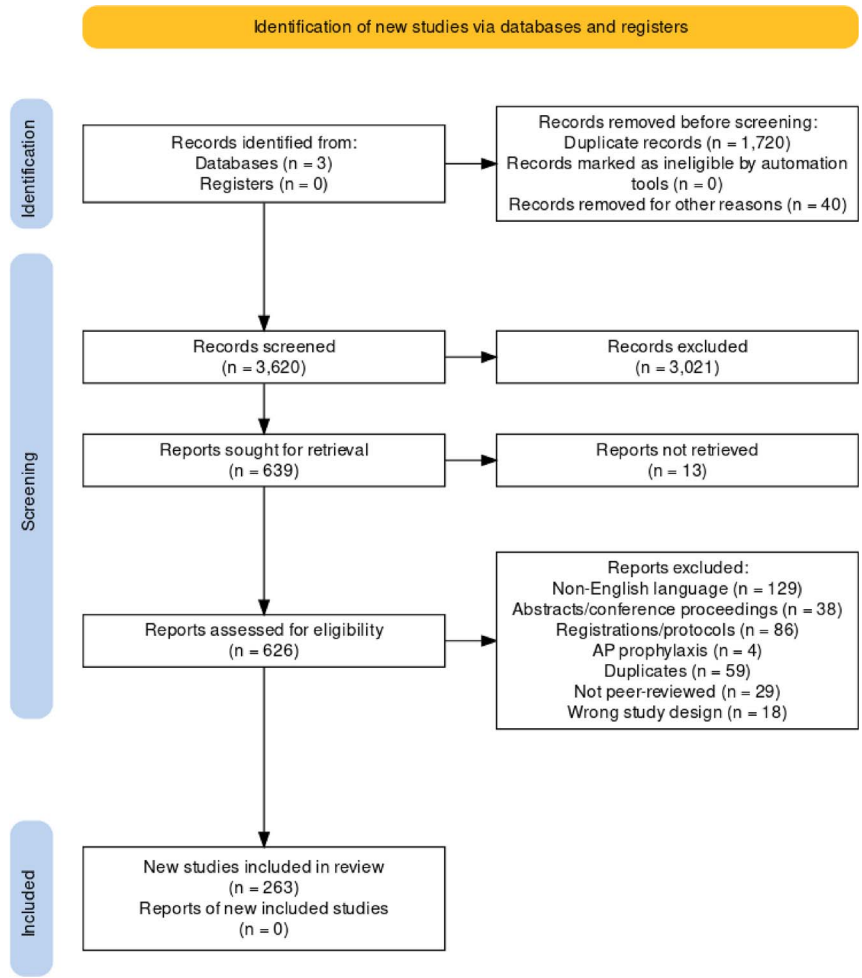

FIGURE 1. PRISMA flow diagram showing selection of articles for review. Figure made using an online reference tool from Haddaway et al. "PRISMA20 20: an R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis." Campbell Systematic Reviews, 18, e1230. https://doi.org/10.1002/cl2.1230.

conducted in Asia increased over time from 7.1% to 26% to 68% (Table 1). In contrast, the percentage of RCTs conducted in Europe decreased from 71% to 53% to 16%. In North America, the proportion of RCTs decreased from 31% to 19% and then to 11% during the same periods ($P < 0.001$).

The RCTs were published in 120 different journals. The journals that published the most RCTs were *Pancreas* (n = 19), *World Journal of Gastroenterology* (n = 15), *Annals of Surgery* (n = 10), and *Pancreatology* (n = 10).

The topics of the RCTs were in a wide range of management domains, including nutrition (n = 71, 28%), pharmacological treatment (55, 22%), surgical or endoscopic management of necrosis (37, 14%), antibiotic treatment or prophylaxis for infection of necrosis (27, 11%), and fluid management (16, 6%). RCT

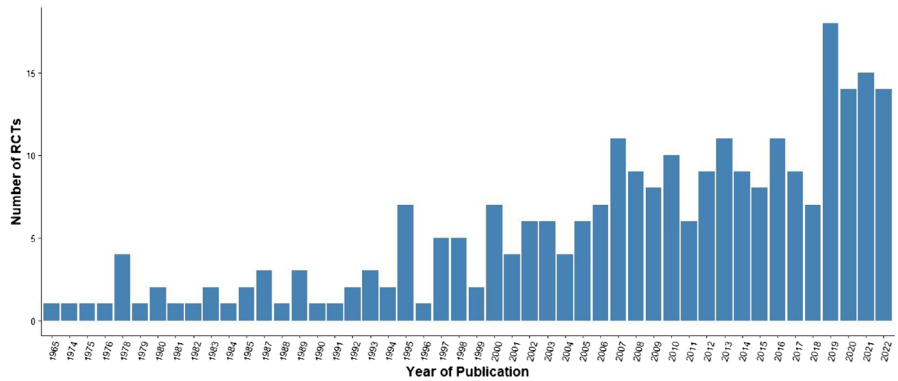

FIGURE 2. Time trends in the number of RCTs in AP management.

© 2024 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2025 Wolters Kluwer Health, Inc. All rights reserved.

TABLE 1. Trial Characteristics Across P1 (Before 1996), P2 (1996–2008), and P3 (After 2008)

	P1	P2	P3	Total	*P*
Region of publication					
Europe	30	38	24	92	$P < 0.001$
Asia	3	19	102	124	
North America	9	14	27	40	
South America	0	1	4	5	
Australia and New Zealand	0	0	2	2	
Severity					
Mild	1	3	20	24	$P < 0.001$
Moderately severe/moderate	0	3	5	8	
Severe	10	44	79	133	
Combination of 2 grades	5	3	16	24	
All/Not specified	26	19	29	74	
Topic					
AP Antibiotic prophylaxis/ treatment	6	10	11	27	$P = 0.037$
AP feeding	1	30	40	71	
AP fluid management	0	1	15	16	
Infection control	7	7	23	37	
Herbal medicine	0	2	9	11	
Pharmacological treatment of AP	19	11	25	55	
Other	9	11	26	46	

AP indicates acute pancreatitis.

topics significantly changed over time ($P = 0.037$; Table 1), with AP fluid management RCT'S increasing and antibiotic treatment/ prophylaxis RCT's decreasing over time.

The severity of AP in patients recruited for the RCTs was variable (Table 1). Trials only including patients with severe disease increased from 24% to 61% between P1 and P2 and then decreased to 53% in P3. The Atlanta criteria and revised Atlanta criteria were only used to classify severity in 51 (19.4%) trials. The known etiology of patients showed a significant association with region and period of publication ($P < 0.001$; Supplementary Table 1, http://links.lww.com/MPA/B232). Hypertriglyceridemia-induced AP increased in proportion with each time period and with the greatest occurrence in Asia; however, most patients included in the RCTs in all regions had biliary AP.

Evidence Mapping

An evidence map was created with the overall RoB grade plotted against the topic of the RCTs (Fig. 3). Herbal medicine has small bubbles and a significant gap indicating a small-pooled sample size. There are also no trials published in this field with a low risk of bias. Most trials with a high risk of bias were published in Asia, which was also found to be a statistically significant association ($P = 0.02$; Supplementary Table 2, http://links.lww.com/MPA/B233). Trials with low risk and some concerns of bias were more equally dispersed by region of publication. European studies provided the largest contribution to pooled populations of studies with low risk of bias in every field of study excluding herbal and pharmacological treatment of AP.

Risk of Bias Assessment

Overall, 21% (54/263) of trials were considered to have low risk of bias, but there were some concerns in 56% (146/263) and high risk of bias in 24% (63/263). More than half of the trials (143/263, 54%) clearly reported an adequate method and concealment of randomization and were therefore classified as being of low risk, 111 (42%) trials had some concerns, and 9 (3%) had a high risk of randomization bias. Bias due to deviations from intended interventions (allocation bias) was determined to be low in 173 (66%) trials of some concerns in 69 (26%) and high in 21 (8%) trials. Bias in missing data was low in a vast majority of RCTs (245/263, 93%), with only 10 (4%) and 8 (3%) trials having high risk of bias and some concerns, respectively, due to missing outcome data. Most trials also had a low risk of bias in the measurement domain of the ROB-2 tool (220/263, 84%). Some concerns were raised for this domain in 23 (9%) trials, and 20 (8%) RCTs had high risk of bias in measurement of outcome. Selective reporting was considered low in 79 (30%) trials. Most trials

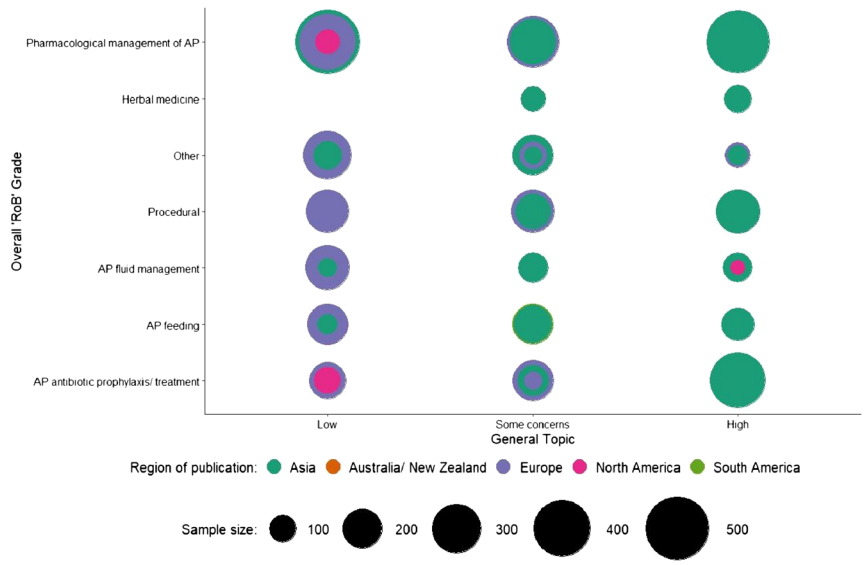

FIGURE 3. Evidence map showing the type of intervention by the quality of the study. The pooled sample size is represented by bubble size and color coded by region of publication.

© 2024 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2025 Wolters Kluwer Health, Inc. All rights reserved.

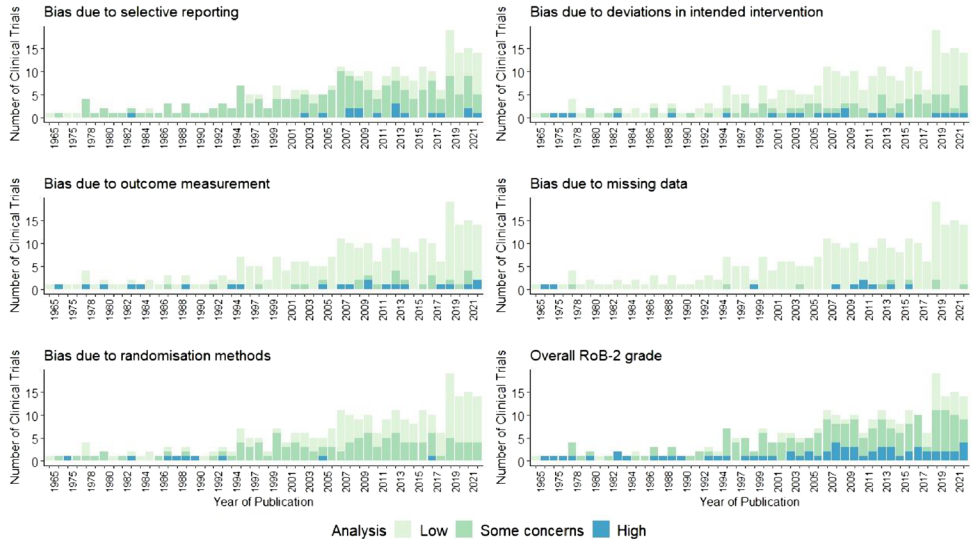

FIGURE 4. Time trend in quality of studies by the 3 parts of the study period.

had some concerns (167/263, 63%) primarily stemming from a lack of a prespecified analysis plan. There were 17 (6%) trials with high risk of selective reporting. There were no trends for some domains of the ROB-2 assessment: randomization bias ($P = 0.41$), bias due to deviations from intervention ($P = 0.42$), and bias due to missing outcome data ($P = 0.95$). However, some domains showed significant variation with time period of publication. Bias in measurement of outcome was significantly associated with the period of publication, with a higher proportion of trials in P3 (86%) having low bias than in P1 (76%) ($P = 0.009$). Bias from selective reporting of results was low in 10% (n = 4), 19% (n = 14), and 41% (n = 61) of trials in P1, P2, and P3, respectively ($P < 0.001$) (Figs. 4, 5; Supplementary Table 3, http://links.lww.com/MPA/B234).

Sample Size Recalculation

Sample size calculations were only reported in 40% trials, and sufficient data were available to recalculate the sample size in 63% (67/106) of these trials (25% of the 263 trials). Among trials providing the complete data for sample size recalculation, the median percentage variation between reported and recalculated sample size was 0%, the lower quartile was −10.3%, and the upper quartile was 7.71%. The range was −87.0% to 98.8%.

The number of trials that provided a sample size calculation significantly increased through the 3 time periods 7 (17%) in P1, 28 (39%) in P2, and 71 (48%) in P3 ($P < 0.001$). The number of trials that provided sufficient data to recalculate the sample size did not change over time ($P = 0.86$). The accuracy of the recalculation of sample size did not change over time ($P = 0.34$).

Presence of Spin

Overall, a quarter of trials (26%, 68/263) contained spin in at least one section, and only one trial contained spin in all sections (title, abstract, and main text). Spin was identified in the title of 8 trials (3%), in the abstracts of 44 trials (17%), and in the main text of 53 trials (20%). Spin techniques are presented

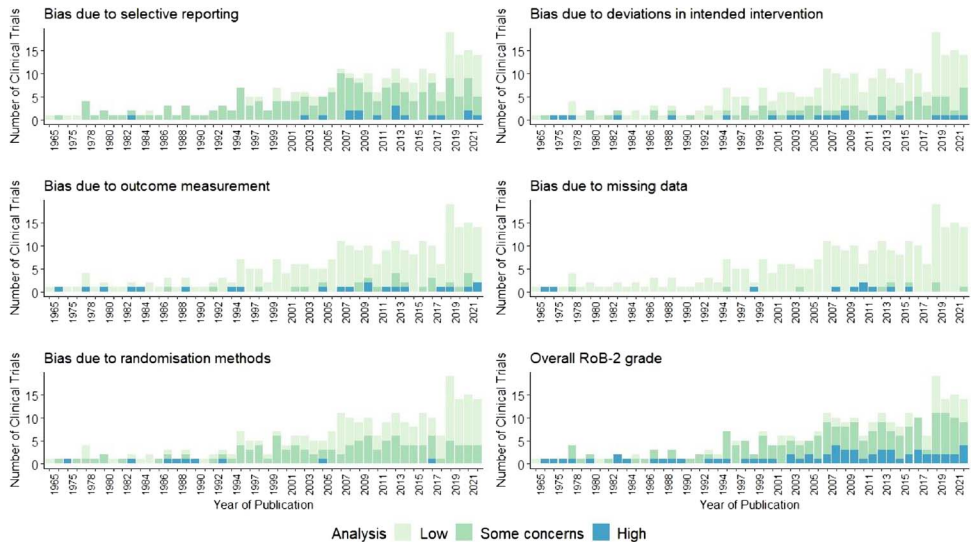

FIGURE 5. Time trend in quality of studies by year of study period.

© 2024 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2025 Wolters Kluwer Health, Inc. All rights reserved.

in Supplementary Table 2, http://links.lww.com/MPA/B233. There was misleading reporting by emphasizing trends in primary end point estimates, despite lacking significance, in 22 (8%) trials. There were other issues, including the use of secondary end point analyses in 27 (10%), no mention of safety of experimental treatment in 12 (5%), the use of safety as a justification for nonsignificant results in 6 (2%) trials, inappropriate use of subgroup analyses in 7 (3%) trials, and within-group comparisons in 4 (2%) being used to justify study hypothesis. Seventeen (7%) RCTs were also reported to have "other" forms of spin, which included not defining a primary outcome and reporting trends in secondary outcomes (Supplementary Table 4, http://links.lww.com/MPA/B235).

DISCUSSION

This systematic review comprehensively examines the quality and quantity of RCTs that investigate the management of acute pancreatitis. The results indicate an improvement of quality parameters and an increased annual volume of RCT's over time. It is notable that selective reporting of results has decreased, and the completeness of sample size calculations has increased. Spin affected a quarter of all trials. Notably, only 40% (n = 106) of the RCTs reported an a priori sample size, and of those RCTs, only 63% (n = 67) provided sufficient data for recalculation. Furthermore, the accuracy of sample size reporting showed no improvement with time (P = 0.34). There were other significant shortcomings in the quality of the literature overall, including a small median sample size of 63 patients, 55% with risk of bias concerns, and 24% with high risk of overall bias.

The stagnation in improvement of quality parameters in the present study may also reflect the significant changes in geographical regions where trials were conducted across time. Although Europe and North America showed some increases in the quantity of trials published across the time periods, their proportional share of all trials published overall diminished compared to Asia. Previous studies have found the quality of trials from Asia to be concerningly low, one study of surgical trials showed no improvement in any quality parameter during a 10-year study period.[22] The Chinese Cochrane Center found that only 207 (6.8%) of 2235 self-described RCTs appeared truly randomized after interviewing the authors of these trials.[23] These are consistent with evidence mapping showing an overrepresentation of trials burdened by concerning risk of bias within Asia. The growing number of RCTs conducted in Asia offers the largest opportunity to substantially improve trial quality and reporting to advance the management of AP. Furthermore, this region provides an opportunity to evaluate the generalizability of treatments beyond the confines of the Western world. However, it is crucial to acknowledge that quality concerns are particularly persistent in Asia and efforts should be taken to improve trial quality in this region.

The quality of RCTs is also influenced by the presence of "spin" or misleading reporting of statistically insignificant outcomes. This can lead to distorted or biased interpretation of results. Overall spin rates for the literature presented in this review are lower than other fields of study because there was no preselection of trials with nonsignificant results for analysis.[24] As with previous research in other biomedical disciplines, spin was predominantly observed in the abstract. In many trials with spin, significant secondary end points were emphasized in cases where the primary end points failed to demonstrate significant changes.[24] This is important because abstracts are more frequently read by clinicians and have a broader circulation, making them a crucial means of communicating research findings.[25] Spin has been identified as a pervasive detriment to honest reporting of trial results and should be carefully considered by peer reviewers and editors to improve the objectivity of trial data in the field.

Transparency in sample size calculations and reporting is essential in establishing the validity of the results. Sample size information inclusion is important in demonstrating the careful planning undertaken prior to trial commencement. Sample size calculations, when accurately reported, also indicate that trials are adequately powered to find statistically significant results. Poor sample size reporting and accuracy is widespread. In a recent study of sample size reporting in RCTs conducted in the field of age-related macular degeneration (AMD), 52% of trials did not report any information on sample size calculations.[15] In an RCT published in the field of dentistry, the difference between the reported and replicated sample sizes was found to range from −237.5% to 84.2%.[26] Omission of sample size calculation has been attributed to multiple factors including the strict word count limitations of most clinical journals and budgetary limitations.[15]

Most systematic reviews assessing the quality of RCTs in other medical disciplines have demonstrated significant improvement in at least one methodological quality parameter over time.[21,27,28] However, these trends are not absolute and consistent across all quality parameters. This may indicate true differences in quality enhancement in different biomedical fields; however, it may also be a result of multiple quality assessment tools being used in different reviews. For example, quality assessment of burns management using the Jadad score showed no significant association with time; however, the Jadad score only incorporates blinding of participants, randomization, and handling of losses to follow-up.[29] Other important quality indicators are not analyzed, and improvements in these parameters could have been missed.[27,29] Unlike the original RoB tool, RoB-2 considers trials to have low risk of bias in the absence of double blinding, conditional on all intended interventions reflecting care that would routinely be received outside the context of the trial. This redefinition of bias assessment and greater subjectivity of the definition influence risk of bias distribution and complicate comparisons to other quality assessment systematic reviews.

There are several limitations to this study, which should be considered. First, the ROB-2 form used for quality assessment has only moderate inter-assessor validity.[30] The relative subjectivity of certain fields is thought to be responsible for this. Inter-assessor validity of spin analysis has not previously been verified. Spin, by its definition, is a relatively subjective measure. This was mitigated by independent assessment by multiple authors, with any conflicts resolved by consensus. Consequential alterations to the Atlanta criteria in 2013 and the heterogeneity of classification systems used make severity predictions inconsistent.[31] Furthermore, only a small number of trials (19.4%) included Atlanta criteria for classifying the severity of disease. This further contributes to the heterogeneity of the trials included in this systematic review. The definition of acute pancreatitis itself varied across randomized controlled trials (RCTs) included in our systematic review and further limits the comparability of studies. Moving forward, using a standardized diagnostic criterion for AP from the best available evidence should be recommended for all RCTs. This will enhance their comparability and facilitate a more robust statistical synthesis of evidence in this field. The quality of non-English publications was not evaluated in this study, which suffers intrinsically from possible publication bias with studies with negative results possibly being less frequently published.

Critical appraisal of all areas of published medical research is important to identify knowledge gaps in the literature and areas where there is a need to improve methodological quality. This quality directly relates to the strength of recommendations in practice guidelines. Evidence mapping of this pooled data can show

© 2024 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2025 Wolters Kluwer Health, Inc. All rights reserved.

time and content trends and may help anticipate future research priorities.[32,33] Critical appraisal of other quality parameters including sample size recalculation accuracy and spin highlights shortcomings. Sample size recalculations were often inaccurate or incomplete. This may reflect a failure in the peer-review process. Journals could provide explicit instructions and examples illustrating guidelines for reporting sample size calculations and detailed information on sample size calculations. Editors and peer reviewers should be familiar with the prevalence and forms of spin in their area of research to ensure accurate and balanced interpretation of research findings and dissemination of research.

Although the quantity and quality of RCTs relating to the management of AP have increased over time, there remain significant shortcomings in methodological quality. Misleading reporting of results was identified in over a quarter of trials, and only 25% of trials reported reproducible sample sizes. Significant improvements in the conduct and reporting of randomized trials in AP are required to improve the evidence base in this field.

REFERENCES

1. Floyd A, Pedersen L, Nielsen GL, et al. Secular trends in incidence and 30-day case fatality of acute pancreatitis in North Jutland County, Denmark: a register-based study from 1981–2000. *Scand J Gastroenterol*. 2002;37:1461–1465.
2. Lindkvist B, Appelros S, Manjer J, et al. Trends in incidence of acute pancreatitis in a Swedish population: is there really an increase? *Clin Gastroenterol Hepatol*. 2004;2:831–837.
3. Hazra N, Gulliford M. Evaluating pancreatitis in primary care: a population-based cohort study. *Br J Gen Pract*. 2014;64:e295–e301.
4. van Dijk SM, Hallensleben NDL, van Santvoort HC, et al. Acute pancreatitis: recent advances through randomised trials. *Gut*. 2017;66: 2024–2032.
5. Steinberg WM. A step-up approach, or open necrosectomy for necrotizing pancreatitis. *N Engl J Med*. 2010;363:1286–1287 author reply 1287.
6. McClave SA, Chang WK, Dhaliwal R, et al. Nutrition support in acute pancreatitis: a systematic review of the literature. *JPEN J Parenter Enteral Nutr*. 2006;30:143–156.
7. Wittau M, Mayer B, Scheele J, et al. Systematic review and meta-analysis of antibiotic prophylaxis in severe acute pancreatitis. *Scand J Gastroenterol*. 2011;46:261–270.
8. Heinrich S, Schäfer M, Rousson V, et al. Evidence-based treatment of acute pancreatitis: a look at established paradigms. *Ann Surg*. 2006;243: 154–168.
9. Goodchild G, Chouhan M, Johnson GJ. Practical guide to the management of acute pancreatitis. *Frontline Gastroenterol*. 2019;10:292–299.
10. Stigliano S, Sternby H, de Madaria E, et al. Early management of acute pancreatitis: a review of the best evidence. *Dig Liver Dis*. 2017;49:585–594.
11. Flemyng E, Dwan K, Moore THM, et al. Risk of Bias 2 in Cochrane Reviews: a phased approach for the introduction of new methodology. *Cochrane Database Syst Rev [Internet]*. 2020;. Available at: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.ED000148/full. Accessed December 30, 2022.
12. Charles P, Giraudeau B, Dechartres A, et al. Reporting of sample size calculation in randomised controlled trials: review. *BMJ [Internet]*. 2009; 338. Available at: https://www.bmj.com/content/338/bmj.b1732. Accessed January 30, 2023.
13. Boutron I, Dutton S, Ravaud P, et al. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA*. 2010;303:2058–2064.
14. Ito C, Hashimoto A, Uemura K, et al. Misleading reporting (spin) in noninferiority randomized clinical trials in oncology with statistically not significant results: a systematic review. *JAMA Netw Open*. 2021;4: e2135765–e2135765.
15. Tulka S, Geis B, Baulig C, et al. Validity of sample sizes in publications of randomised controlled trials on the treatment of age-related macular degeneration: cross-sectional evaluation. *BMJ Open*. 2019;9:e030312.
16. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ [Internet]*. 2010;340. Available at: https://www.bmj.com/content/340/bmj.c869. Accessed January 19, 2023.
17. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Rev Esp Cardiol*. 2021;74:790–799.
18. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ [Internet]*. 2019;366. Available at: https://www.bmj.com/content/366/bmj.l4898. Accessed September 24, 2022.
19. Erdfelder E, Faul F, Buchner A. GPOWER: a general power analysis program. *Behav Res Methods Instrum Comput*. 1996;28:1–11.
20. Goodyear MDE, Krleza-Jeric K, Lemmens T. The Declaration of Helsinki. *BMJ*. 2007;335:624–625.
21. Hüttner FJ, Capdeville L, Pianka F, et al. Systematic review of the quantity and quality of randomized clinical trials in pancreatic surgery. *Br J Surg*. 2019;106:23–31.
22. Ahmed Ali U, van der Sluis PC, Issa Y, et al. Trends in worldwide volume and methodological quality of surgical randomized controlled trials. *Ann Surg*. 2013;258:199–207.
23. Wu T, Li Y, Bian Z, et al. Randomized trials published in some Chinese journals: how many are randomized? *Trials*. 2009;10:46.
24. Demarquette A, Perrault T, Alapetite T, et al. Spin and fragility in randomised controlled trials in the anaesthesia literature: a systematic review. *Br J Anaesth [Internet]*. 2023;130:528–535. Available at: http://dx.doi.org/10.1016/j.bja.2023.01.001.
25. Saint S, Christakis DA, Saha S, et al. Journal reading habits of internists. *J Gen Intern Med*. 2000;15:881–884.
26. Koletsi D, Fleming PS, Seehra J, et al. Are sample sizes clear and justified in RCTs published in dental journals? *PLoS One*. 2014;9:e85949.
27. Falagas ME, Grigori T, Ioannidou E. A systematic review of trends in the methodological quality of randomized controlled trials in various research fields. *J Clin Epidemiol*. 2009;62:227–231, 231.e1–9.
28. Park HJ, Kwon JH, Yoo JH. Quantity and Quality Assessment of Randomized Controlled Trials in the Atopic Dermatitis. *The Journal of Korean Medicine [Internet]*. 2014; Available at: https://www.koreascience.or.kr/article/JAKO201435053629381.page.
29. Appendix: Jadad scale for reporting randomized controlled trials. *Evidence-based Obstetric Anesthesia*. Oxford, UK: Blackwell Publishing Ltd; 2007:237–238.
30. Jørgensen L, Paludan-Müller AS, Laursen DR, et al. Evaluation of the Cochrane tool for assessing risk of bias in randomized clinical trials: overview of published comments and analysis of user practice in Cochrane and non-Cochrane reviews. *Syst Rev [Internet]*. 2016;5:80. Available at: https://link.springer.com/article/10.1186/s13643-016-0259-8.
31. Banks PA, Bollen TL, Dervenis C, et al. Classification of acute pancreatitis—2012: revision of the Atlanta classification and definitions by international consensus. *Gut*. 2013;62:102–111.
32. Miake-Lye IM, Hempel S, Shanman R, et al. What is an evidence map? A systematic review of published evidence maps and their definitions, methods, and products. *Syst Rev*. 2016;5:28.
33. O'Leary BC, Woodcock P, Kaiser MJ, et al. Evidence maps and evidence gaps: evidence review mapping as a method for collating and appraising evidence reviews to inform research and policy. *Environ Evid*. 2017;6:1–9.

© 2024 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2025 Wolters Kluwer Health, Inc. All rights reserved.