

A systematic review of the quality of reporting and risk of bias for randomized crossover trials in digestive disease journals

Qian Zhou , Zhi-hang Chen, Jin-xin Zhang and Sui Peng

Abstract

Background: The quality of randomized crossover studies on digestive diseases is unclear. We aimed to review crossover trials in digestive disease journals and evaluate their reporting quality and risk of bias.

Methods: We searched the PubMed, Web of Science, and Scopus databases for all crossover trials in 39 digestive journals between January 2011 and September 2021. Reporting adherence was based on the CONSORT 2010 statement: extension to randomized crossover trials published in July 2019. A newly released Cochrane risk of bias tool 2.0 extension for crossover trials was applied to assess the risk of bias.

Results: In total, 173 studies were included in the analysis, and 16.2% were published following the CONSORT statement extension. The crossover design was not only widely used in drug efficacy trials (48.6%) but also in endoscopic ultrasound trials (23.7%) and dietary studies (17.9%) in the field of digestive diseases. The overall reporting adherence was 37.6% for full texts and 43.4% for abstracts. The proportions of trials with low, some concerns, and high risk of bias were 13.9%, 15.6%, and 70.5%, respectively. The difference in reporting adherence and high risk of bias between pre- and post-CONSORT was not significant. Having a sample size plan, defining primary end points, and pre-registration showed higher reporting adherence and lower risk of bias than those who did not.

Conclusion: These findings demonstrated the inadequate quality of randomized crossover trials for digestive diseases. Compliance with the CONSORT extension for crossover trials must be strengthened and improved (PROSPERO CRD: 42021248723).

Keywords: CONSORT extension, crossover trials, digestive disease, risk of bias

Received: 27 July 2021; revised manuscript accepted: 29 November 2021.

Introduction

Digestive diseases, such as gastroesophageal reflux disease, irritable bowel syndrome, and inflammatory bowel disease, are widespread worldwide, represent a heavy healthcare burden, and account for several deaths.^{1–4} In recent years, an increasing number of clinical trials exploring therapeutic efficacy on digestive diseases have been registered and performed. In particular, crossover trial designs play an important role in digestive disease studies due to their advantage of accelerating clinical translation by reducing the required sample size.^{5,6}

Crossover trials are experiments in which participants, namely, patients or healthy volunteers, are given two or more sequential treatments in random order separated by a washout period.^{6,7} The most common design is the AB/BA design, in which participants are assigned randomly to the two sequence groups A (first)–B (second) and B (first)–A (second), and the two treatments are compared at the individual rather than group level. Each participant received all treatments, rather than a single treatment, as in parallel group trials. A cross-sectional study reported that crossover

Ther Adv Gastroenterol

2022, Vol. 15: 1–13

DOI: 10.1177/
17562848211067874

© The Author(s), 2022.
Article reuse guidelines:
sagepub.com/journals-
permissions

Correspondence to:

Qian Zhou
Department of Medical
Statistics, Clinical Trials
Unit, The First Affiliated
Hospital, Sun Yat-sen
University, Guangzhou
510080, China.
[zhouq49@mail.sysu.
edu.cn](mailto:zhouq49@mail.sysu.edu.cn)

Sui Peng
Clinical Trials Unit, The
First Affiliated Hospital,
Sun Yat-sen University,
Guangzhou 510080, China

Department of
Gastroenterology, The
First Affiliated Hospital,
Sun Yat-sen University,
Guangzhou, China
pengsui@mail.sysu.edu.cn

Zhi-hang Chen
Department of Liver
Surgery, The First
Affiliated Hospital, Sun
Yat-sen University,
Guangzhou, China

Jin-xin Zhang
Department of Medical
Statistics, School of Public
Health, Sun Yat-sen
University, Guangzhou,
China



designs were used in a significant proportion of randomized clinical trials (116 of 526).⁸

In the field of digestive diseases, using a crossover design offers great advantages. First, most digestive disease conditions are consistent between treatment periods, but the effects of treatment do not last. This situation allows the same patient to receive different drugs during different periods. Second, slight modifications of different dosages are usually compared in clinical trials in relation to drugs or food for digestive diseases. For example, when investigating the efficacy of acid inhibition of omeprazole, a dosage of 40 mg was chosen for the experimental group and 20 mg for the control group.⁹ Crossover designs efficiently detect a slight impact resulting from slight dosage modifications. Therefore, the control group is generally a standard treatment rather than a placebo. Crossover designs permit opportunities for head-to-head trials on the basis of reducing sample size.¹⁰ Furthermore, some digestive studies have focused on comparing different endoscopies, and many studies using crossover designs have saved more samples than parallel designs.¹¹ The fact that the endoscopy effect is not persistent in the following period like with drugs is an advantage of using a crossover design.

However, some studies have discussed the specific problems associated with using a crossover design.^{6,7,12–15} Among these, the most important is the carry-over effect that results from an insufficient washout period and residual treatment effects. The carry-over effect occurs when the effects of a drug or a treatment given during one period persist into the following one, thus interfering with the effects of a different, subsequent drug and causing carryover bias. In addition, improper use of statistical methods is another problem prevalent in crossover trials that leads to unconvincing, even wrong, conclusions. Therefore, some studies have emphasized the importance of reporting proper statistical methods in systematic reviews and meta-analyses.^{16,17} Furthermore, if a patient withdraws during the first period, their data for the following period cannot be collected, thus leading to an increased ratio of missing data and unavailable within-subject comparison. Thus, some issues specific to crossover designs must be considered when reporting and publishing the results of crossover trials. Owing to the different research purposes, crossover designs also have issues that are specific

to digestive diseases. Therefore, understanding the quality of reporting and risk of bias in studies of digestive diseases will be greatly beneficial for investigators in this field to improve the study design and research quality.

Previous reviews have assessed the quality of reporting of crossover trials in several conditions, including chronic pain,^{18,19} open-angle glaucoma, and intraocular hypertension.²⁰ The Consolidated Standards of Reporting Trials (CONSORT) 2010 statement: extension to randomized crossover trials was published 2 years ago to facilitate better reporting of crossover trials.⁷ However, the citation rate is not sufficient compared to the number of crossover trials.²¹ Recently, the revised Cochrane risk of bias tool for randomized trials [Cochrane risk of bias tool 2.0 (ROB 2.0)], which is used to evaluate the risk of bias for crossover trials, has also been released.²² This systematic review aimed to assess the reporting quality of trial design characteristics according to the CONSORT 2010 extension and risk of bias in crossover trials based on ROB 2.0 in the field of digestive diseases. We hope that, by understanding the current state of crossover trials of digestive diseases, we can improve the reporting quality and reduce the risk bias in future trials.

Materials and methods

Study design

This study was a systematic review aiming to assess the quality of reporting and risk of bias in randomized crossover trials in digestive disease research. The study was performed based on three literature databases (PubMed, Web of Science, Scopus) search according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations, as shown in Supplementary Table 1.²³ Our protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO: CRD 42021248723).

We restricted our review to digestive journals with a journal impact factor (JIF) larger than 3.0, based on the 2019–2020 Journal Citation Reports. Supplementary Table 2 lists the 39 journals reviewed. The median JIF was 4.16 [interquartile range (IQR): 3.53, 7.63]. These journals covered nearly 60% of all digestive journals with nonzero JIFs. We also searched journals specific for

cancers, such as *Gastric Cancer*, *Liver Cancer*, and *Journal of Hepatocellular Carcinoma*, but no cross-over article was found; thus, we did not include these journals. All studies using randomized crossover designs in pre-specified journals published between January 2011 and September 2021 were included. Articles using parallel-controlled designs, observational studies, pilot studies, animal or cell studies, and systematic reviews were excluded. One of the authors (Z.H.C.) screened all titles and abstracts using the predefined eligibility criteria, and 50% of the titles and abstracts were screened independently by at least one coauthor. Full articles were then screened by one author using the predefined eligibility criteria. All eligible articles were screened twice. The differences between authors were resolved through discussion. None of the articles excluded at first were subsequently found to be eligible.

Search strategy and data sources

Reports of randomized crossover trials in journals of digestive diseases were identified by searching three databases (PubMed, Web of Science, and Scopus; from January 2011 to September 2021). The search strategy included two parts: one for crossover trials ('(((cross-over OR cross over OR crossover) Not (cross-sectional))) Not case cross') and the other for pre-specified journals listed in Supplementary Table 2. These two parts were combined using the Boolean operator AND. All searches were performed by researchers Q.Z. and Z.H.C. The reports retrieved from the three databases were duplicated first.

Data collection and definition

Data on publication details, including the following three categories, were extracted. First, the reporting of trial characteristics included the name of the first author, year of publication (2011–2021), population (healthy volunteers, patients), positions (intestinal, gastroesophageal, hepatic-biliary-pancreatic, others), intervention (food, drugs, endoscopy-relevant, others), registration (yes, no), result (positive, negative), and justification of crossover designs. Second, the reporting of trial design characteristics included study design (two-way, three-way, four-way crossover, others), number of treatments (2–6), number of periods (2–5), the use of a washout period (yes, no), center (single multiple), hypothesis testing (superiority, noninferiority), sample

size pre-estimation (yes, no), number of study participants (including number planned, randomized, and analyzed), primary outcome (specified continuous end point, specified binary outcome, not specified), and blinding (open label without providing reasons, open label with reasons, single-blinded, double-blinded). Third, the statistical methods for primary outcome data analyses included various statistical methods separated by continuous and binary outcomes accounting for paired data analysis (yes, no), missing data process (yes, no), and handling of carry-over or period effect (yes, no, no with reasons). Data extraction was performed by researchers Z.H.C. and Q.Z., separately, using an Excel spreadsheet with pre-specified items to be extracted (Excel for Mac 2011; Microsoft, Redmond, WA, USA). The differences between authors were resolved through discussion.

Assessment of reporting quality and risk of bias

The reports were assessed by one of the authors (Z.H.C.) and double checked by another author (Q.Z.). Reporting quality was assessed according to the CONSORT 2010 statement: extension to randomized crossover trials.⁷ A total of 37 items were required by the reporting checklist and were presented by a bar chart of the percentages calculated by dividing the number of articles reported by the total number. Furthermore, we calculated the reporting adherence rate for each article by dividing the number of items reported by the number of items required according to CONSORT and evaluated the overall quality of reporting using median and IQR. Abstract adherence to the CONSORT 2010 extension was also evaluated and classified as 'yes' if nine or more items of the 16 required for the abstracts by the reporting checklist indicated yes. To consider the lag in applying the guidance, we defined 'after CONSORT' as 6 months after the statement's publication.

The risk of bias was evaluated using ROB 2.0 and additional considerations for crossover trials.²² ROB 2.0 consisted of five domains and an overall judgment. Each domain focuses on different aspects of a trial, including randomization, period and carry-over effect, missing outcome data, measurement of outcome, and selection of the reported results. In accordance with the original statement requirements, we used ROB 2.0 for randomized crossover trials to assess the risk of

bias based on the trial primary outcome. If the primary outcome was not defined in the trial, or more than two primary outcomes were attested, assessments were performed based on each reported outcome; then, a comprehensive assessment was made for this article.

Statistical analysis

Summary data were presented as the proportion of article abstracts and full texts reporting the features of interest. Categorical variables were described using frequencies and proportions, while continuous variables as mean and standard deviation (SD) or median with IQR. Chi-square test or Fisher's exact test was used to analyze the categorical data. Continuous data were compared using the *t*-test or Wilcoxon rank-sum test. Wilcoxon signed-rank test was performed to compare the difference in sample size between true observed and estimated or randomized numbers. Subgroup analysis was performed according to the risk of trial bias. The study was inappropriate for pool data analysis; thus, an integrated analysis was not performed. All statistical analyses and figures were performed using R software (R Core Team, 2021).²⁴ Statistical significance was set at $p < 0.05$.

Results

Trial characteristics

Database searching revealed that 2548 hits, including 622 from PubMed, 912 from Web of Science, 963 from Scopus database, and 51 from other sources, from 29 of the 39 academic journals were eligible for inclusion (Figure 1; Supplementary Table 3). If two or more trials were included in one publication, the one representing the primary study purpose and a larger sample size was extracted. Thus, 173 randomized crossover trials were extracted. The numbers of included crossover trials ranging from 5 to 22 between 2011 and 2021 (9 months in 2021) are presented in Figure 2. No monotonic trend over time was observed. Table 1 shows the overall characteristics of the included crossover trials. Twenty-eight studies were published 6 months after the publication date of the CONSORT statement extension for crossover trials. The total sample size included for analysis was 10,477, and the median value was 25 (IQR: 16–55). Most of the studies (86.7%) were designed as AB/BA crossover trials. Furthermore, they were mainly

conducted in gastroesophageal (38.7%) and intestinal (42.2%) positions.

Adherence to reporting standards

Figure 2 depicts the change in overall adherence to reporting by year for both full texts and abstracts. The reporting adherence of trial characteristics requested by the CONSORT statement for each item is shown in Figure 3(a). Adherence to the reporting standards was poor (<50% adherence) for 22 of the 37 CONSORT items. Overall, publications adhered to between 0% and 100% of the CONSORT items, with a median of 37.6% (IQR: 14.5–63.6%). Five items were reported in 90% or more of the 173 studies and 11 items in less than 20%. A flowchart for patient enrolment was present in 71 of the 173 (41.0%) studies. We divided the 37 items into seven parts: title, abstract, introduction, method, results, discussion, and others. The reporting adherence to the introduction (99.1%) was the highest, followed by abstract (49.7%), and the adherence to title was the lowest, followed by results (32.4% and 35.8%, respectively). Sensitivity analysis was performed by excluding three items ('Methods-Change from protocol', 'Methods-Changes to outcomes', and 'Methods-Similarity of interventions') from the data because of the very low percentage of study protocols available. The median adherence rate increased to 40.6% (IQR: 21.9–62.5%). The reporting adherence of abstracts requested in the CONSORT statement of each item is shown in Figure 3(b). The median adherence percentage for the abstracts was 43.4% (IQR: 26.0–85.4%). Reporting adherence of the 'Conclusions' was the most frequent, followed by 'Methods-Objective' and 'Methods-Interventions' (97.1%, 95.4%, and 94.8%, respectively). Nonadherence to reporting 'Funding Information' was the most frequent, followed by 'Methods-Participants' and 'Results-Outcome' (3.5%, 16.2%, and 20.8%, respectively).

Risk of bias assessment

The overall risk of bias was assessed using ROB 2.0. Most of the studies (70.5%) were classified as having a high risk of bias, 13.9% as low risk, and 15.6% as posing some concerns (Figure 4). The measurement of the outcome domain was most commonly rated at high risk of bias (as opposed to the missing outcome data domain). Reporting adherence was significantly negatively correlated with the risk of bias (Spearman correlation

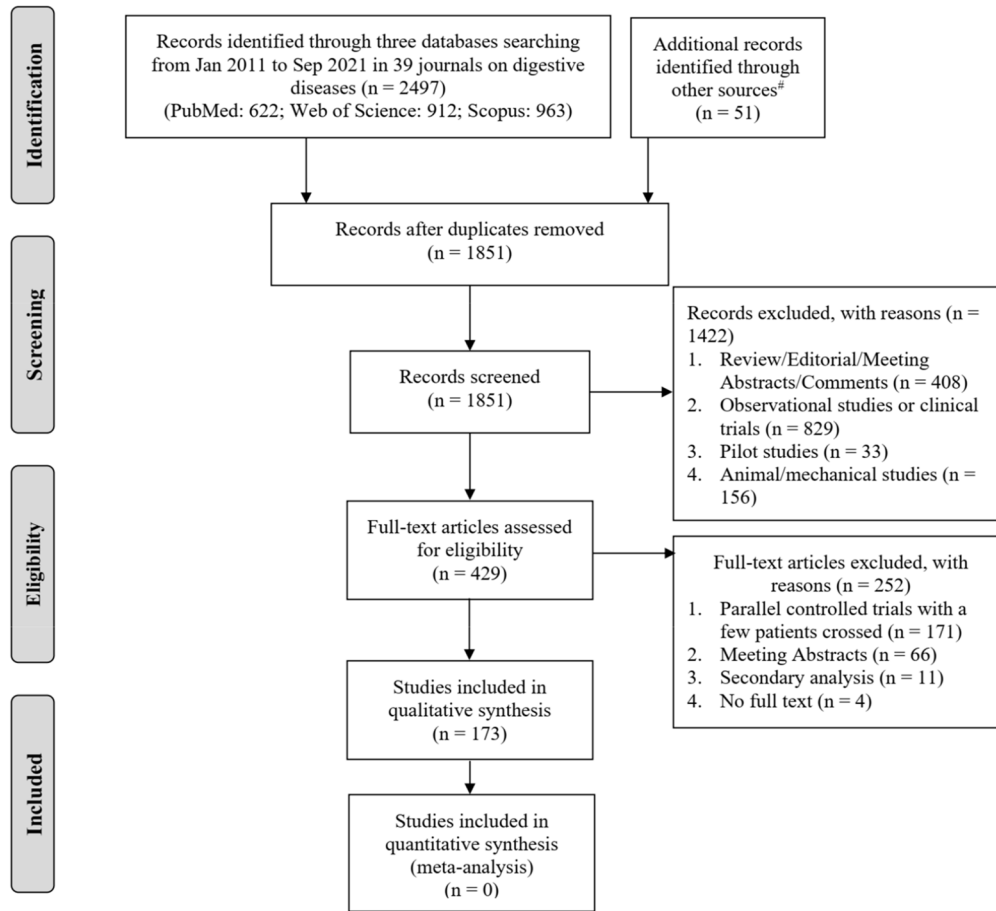


Figure 1. PRISMA flow chart of the study. In addition, we consulted the *Hepatobiliary Surgery and Nutrition* website, as no record was found from the three literature databases. The information is presented in other sources (#).

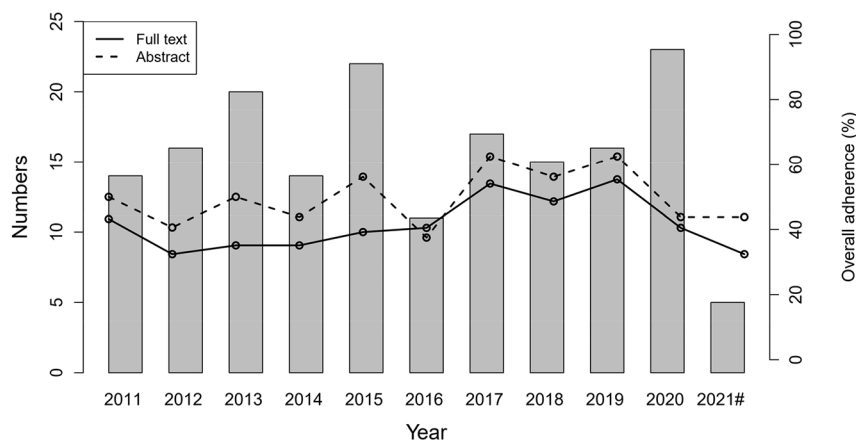


Figure 2. Number of randomized crossover trials and overall reporting adherence for both full text and abstract by year of publication. The 2021 period ranged from January to September.

Table 1. Characteristics of the included randomized crossover trials in digestive disease journals.

Characteristic	Frequency (%) or Median (IQR) (N = 173 trials in total)
Publication year	
Before CONSORT for crossover trial (2010–2019)	145 (83.8)
After CONSORT for crossover trial (2020–2011) ^a	28 (16.2)
Journal impact factor (n = 39 journals for searching)	4.160 (3.534–7.628)
Journal impact factor (n = 29 journals for analysis)	4.774 (3.549–7.549)
Population	
Patients	115 (66.5)
Healthy volunteers	58 (33.5)
Study design	
Two-way (AB/BA)	150 (86.7)
Three-way (ABC/BCA/CAB)	12 (6.9)
Four-way (ABCD/BCDA/CDAB/DABC)	7 (4.0)
Others	4 (2.3)
Justification of crossover design	
No	165 (95.4)
Yes	8 (4.6)
Number of treatments	
2	153 (88.4)
3	13 (7.5)
4	5 (2.9)
5	1 (0.6)
6	1 (0.6)
Number of periods	
2	153 (88.4)

(Continued)

Table 1. (Continued)

Characteristic	Frequency (%) or Median (IQR) (N = 173 trials in total)
3	13 (7.5)
4	6 (3.5)
5	1 (0.6)
Positions of condition	
Gastroesophageal	67 (38.7)
Intestinal	73 (42.2)
Hepatic–biliary–pancreatic	13 (7.5)
Others	20 (11.6)
Intervention	
Drugs	84 (48.6)
Endoscopy relevant	41 (23.7)
Food	31 (17.9)
Others	17 (9.8)
Control	
Null	11 (6.4)
Placebo	54 (31.2)
Standard therapy/typical diet/positive control	89 (51.4)
Different dose control	19 (11.0)
Registered in a clinical trial registry	
No	79 (45.7)
Yes	94 (54.3)
Hypothesis testing	
Superiority	167 (96.5)
Noninferiority	6 (3.5)
Results	
Positive	138 (79.8)
Negative	35 (20.2)
Total sample size	10477

(Continued)

Table 1. (Continued)

Characteristic	Frequency (%) or Median (IQR) (N = 173 trials in total)
Median sample size	25 (16–55)
Performing sample size estimation	
No	69 (39.9)
Yes	104 (60.1)
Washout	
No-without reason	95 (54.9)
No-with reason	5 (2.9)
Yes	73 (42.2)
Washout period (day)	14 (7, 21)
Primary end point definition	
No	41 (23.7)
Yes-continuous outcome	85 (49.1)
Yes-binary outcome	47 (27.2)
Masking	
Open label without reason	74 (42.8)
Open label with reason	6 (3.5)
Single-blinded	24 (13.9)
Double-blinded	69 (39.9)
IQR, interquartile range. ^a Considering the lag of the application of the guidance, we defined ‘after CONSORT’ as 6 months after the publication date (July 2019) of the CONSORT statement: extension for crossover trials.	

coefficient: -0.393 , $p < 0.001$). However, studies with a high risk of bias were $>50\%$ in both $\geq 50\%$ and $<50\%$ adherence to reporting groups (56.4% and 77.1%, respectively).

Subgroup analysis of adherence to reporting

To investigate the difference in reporting adherence in different categorical factor subgroups, we calculated the median percentage and IQR for each group and compared the differences in these

categories for each factor (Figure 5). The overall reporting adherence did not differ between the years 2011–2019 and 2020–2021 based on the publication date of the CONSORT statement for crossover trials (40.5% (IQR: 32.4%, 54.1%) versus 40.5% (IQR: 32.4%, 52.0%), $p = 0.942$). Among them, the item reporting adherence of ‘Funding’ and ‘Methods-Sequence generation types’ improved the most. However, item reporting adherence of ‘Methods-Settings and location’ and ‘Methods-Blinding’ decreased (Supplementary Table 4). Overall adherence significantly increased in 2016–2021 compared to 2011–2015 [35.1% (IQR: 29.7%, 46.0%) versus 48.7% (IQR: 35.1%, 56.8%), $p < 0.001$]. Among them, 11 items of the item reporting adherence improved (Supplementary Table 4). Forest plots also showed that reporting adherence was significantly higher in healthy population subgroups than for patients, in positions of condition in intestinal and hepatic–biliary–pancreatic than in gastroesophageal and others, food interventional registered studies having performed sample size estimation and with pre-specified binary primary end point (Figure 5).

Subgroup analysis of risk of bias

To explore the difference in high risk of bias in different categorical factor subgroups, we calculated the percentage and 95% confidence interval for each group and compared the differences in these categories for each variable (Figure 6). In subgroups of hepatic–biliary–pancreatic and other positions of condition, no registration, no pre-estimation of sample size, no washout period, and open-label studies, there were more trials with a high risk of bias. To investigate the reason for the high risk of bias, we paid attention to the sample size estimation and found that the randomized number and the number for analysis significantly differed in trials without considering dropout rate [$n = 146$, 22 (14, 49) versus 23 (15, 51); exact Wilcoxon signed-rank test $p < 0.001$]. The sample size calculation was not pre-planned in nearly 40% of the articles. In the true sample size, trials with and without sample size estimation significantly differed [42 (20, 80) versus 19 (12, 27), Wilcoxon rank-sum test $p < 0.001$].

Discussion

This study conducted a comprehensive review to evaluate the quality of crossover trials in digestive

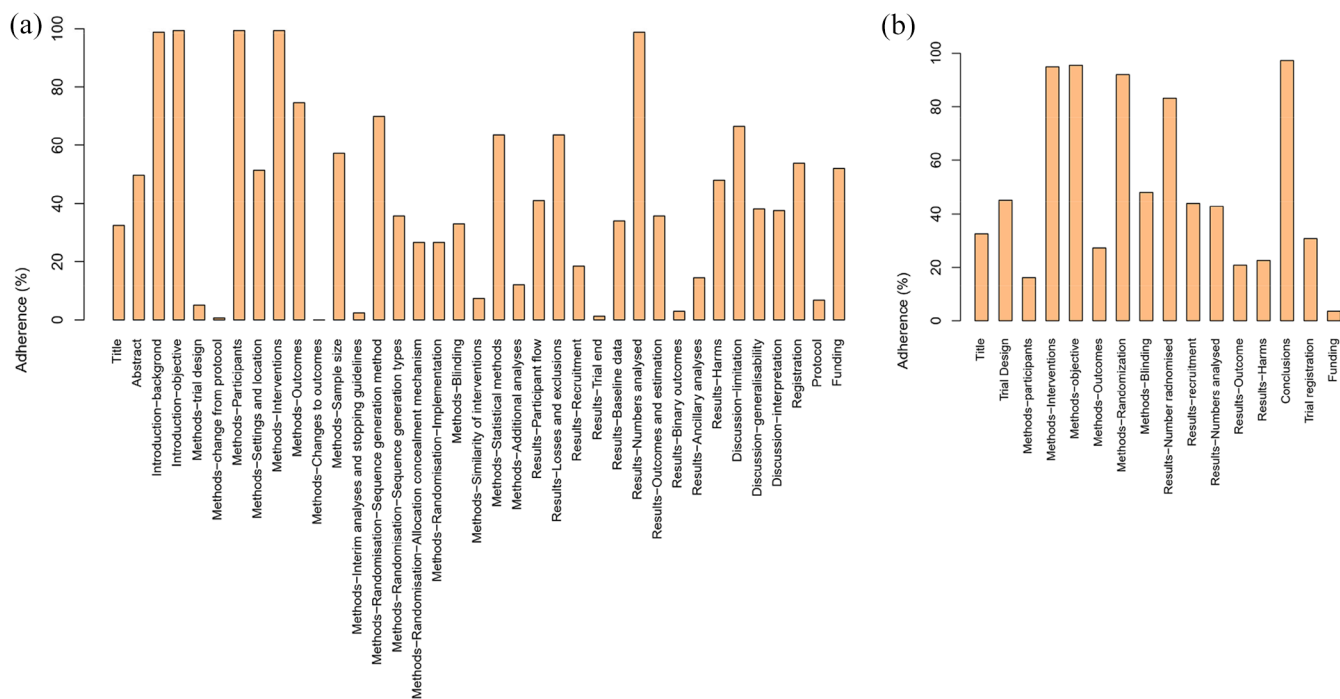


Figure 3. Completeness of reporting individual CONSORT 2010 statement: extension to randomized crossover trials items. (a) Checklist for full text. (b) Checklist for abstract.

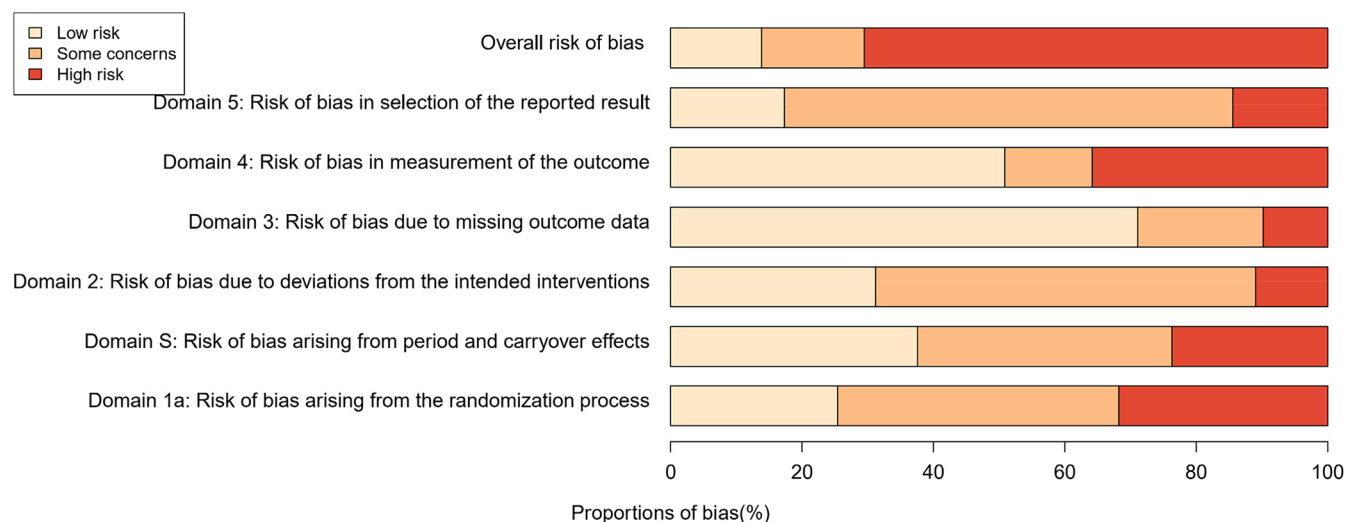


Figure 4. Risk of bias proportions obtained using ROB 2.0.

diseases based on the recently published CONSORT reporting guidance⁷ and the risk of bias assessment for crossover trials.²² It included 173 randomized crossover trials published in digestive journals over the past decade. Our results show that crossover designs are not only widely

used in digestive drug efficacy trials but also endoscopic ultrasound trials and dietary studies of digestive diseases. The overall reporting adherence to the CONSORT statement extension for crossover trials is insufficient, and a large proportion of these trial studies presented a high risk of bias.

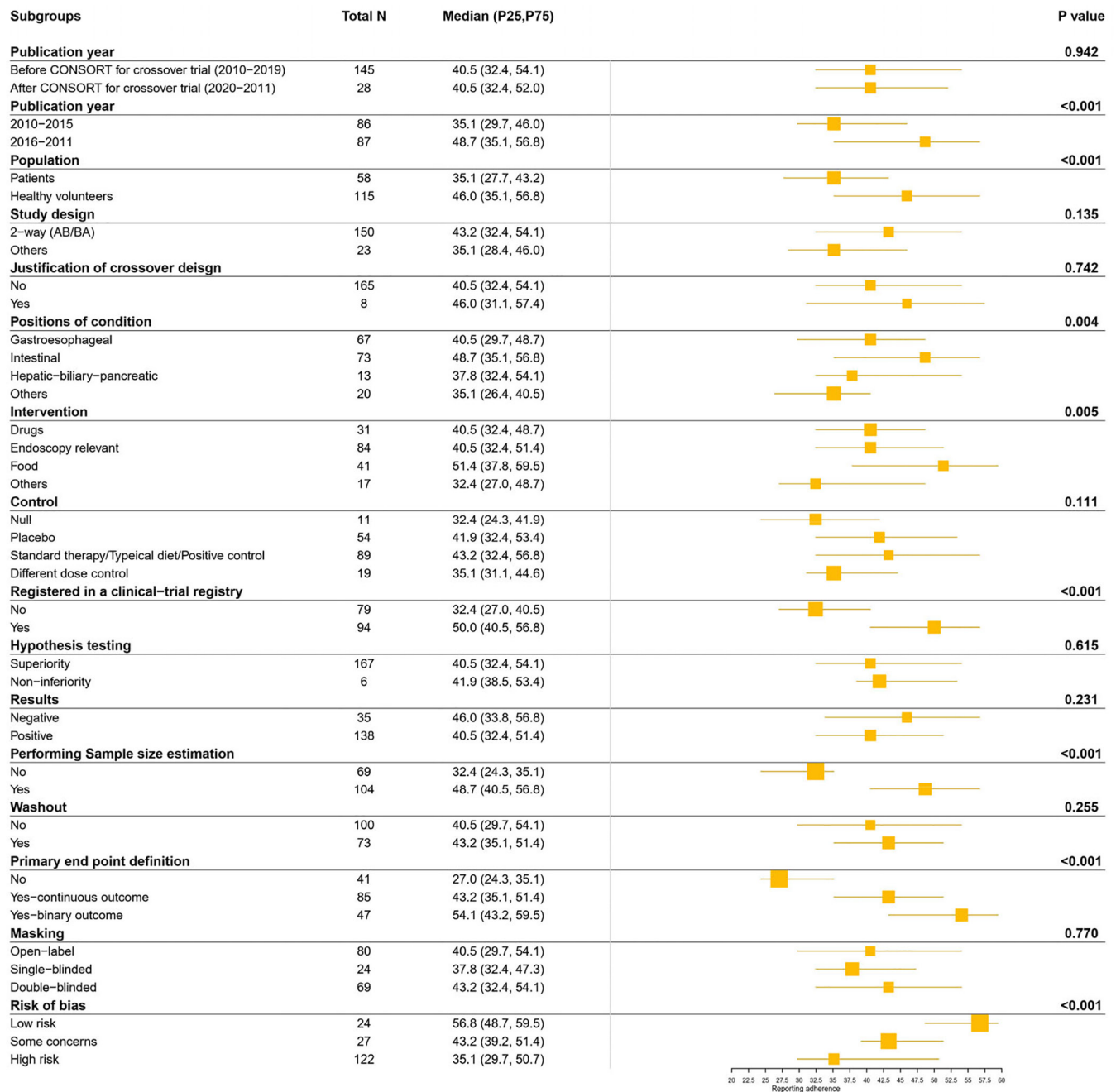


Figure 5. Subgroup analysis of reporting adherence percentage to CONSORT statement extension for crossover trials. The number of studies in each subgroup was displayed in the second column. The reporting adherence percentage was calculated for each study; every subgroup contained several percentages. The medians and interquartile ranges of the reporting adherence percentages were used for data description and presented in the third column. *P* values were calculated using the Wilcoxon test for two independent groups or the Kruskal–Wallis test for three or more groups.

Although the CONSORT statement extension for crossover trials was published 2 years ago, citations to this guidance were lacking in the included trials. As prospective interventional trials, these crossover trials also failed to refer to the

updated CONSORT guidelines for parallel randomized trials that have existed for 10 years.²⁵ These two guidelines share several features, such as items for reporting randomization, blinding, and adverse events. The crossover trials in this

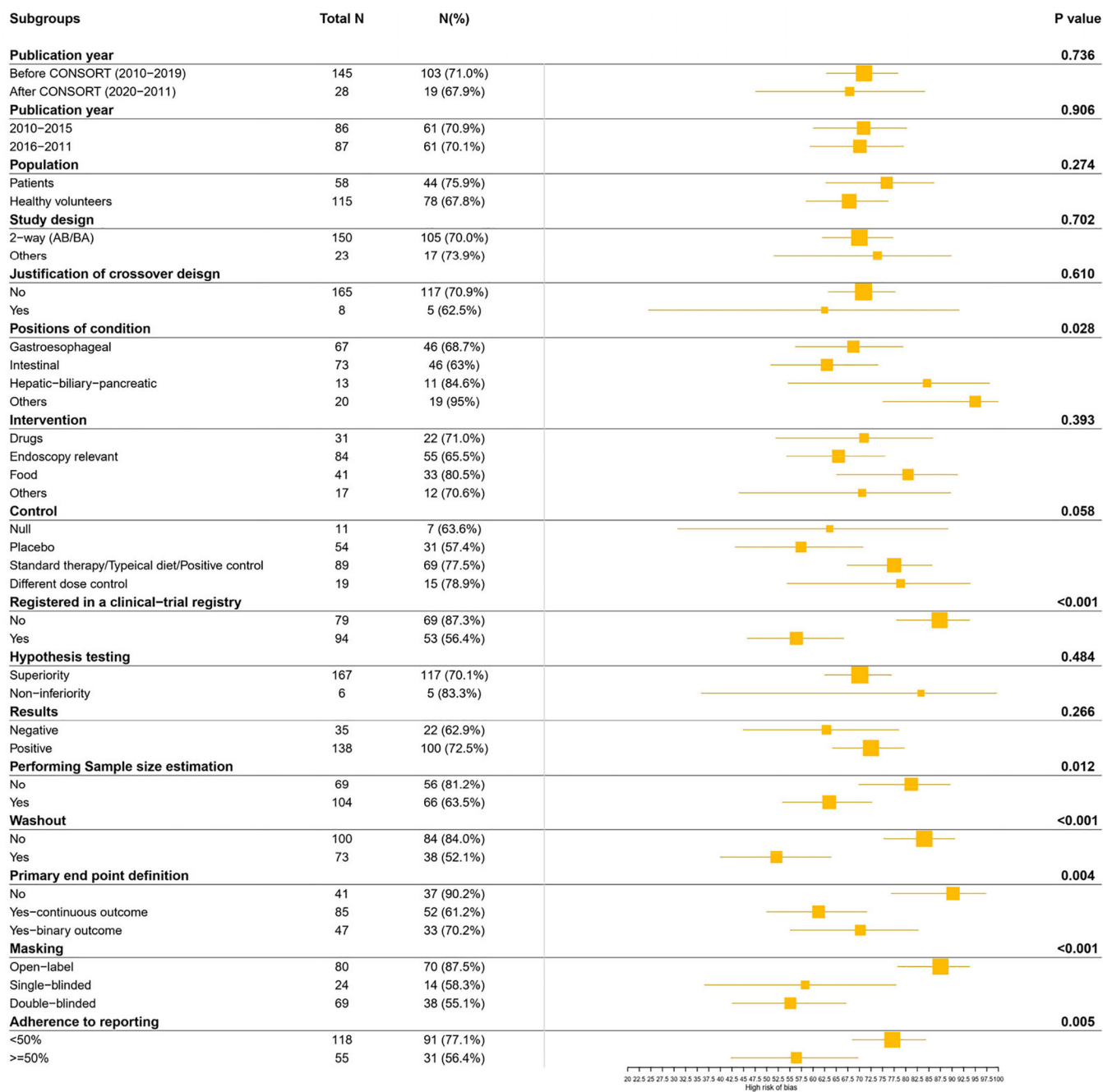


Figure 6. Subgroup analysis of high risk of bias. Total N was the number of articles in the subgroup. N (%) was the number and proportion of the high risk of bias articles. P values were calculated using the Chi-square test or Fisher’s exact test.

study reported fewer washout periods than Mills’ review of 116 crossover trials in various fields 20 years ago.⁸ Consistent with previous findings,^{19,26} our study also displayed limited clarity in reporting certain design details, estimates of treatment effects, and associated variability, and methods to accommodate missing data remain common.

Investigators should comply more closely with the existing CONSORT checklist for crossover trials. Our study found a significant negative correlation between reporting adherence and the risk of bias. However, while reporting adherence has improved in the last 5 years, it has declined in the past 2 years since the CONSORT was published,

suggesting that the quality of reporting is unstable. Specifically, failure to specify whether the trial was registered in advance, approved by ethical committees, reported funding information, and had study protocol reporting affect the judgment of a study's reporting adherence. The worst part of the reporting was the study design details, which could be attributed to the fact that most crossover trials were not designed for formal phase III clinical trials and that a standard trial protocol was often unavailable. Researchers must design and outline the research plan before the trial is registered in a clinical trial registry. In particular, although many studies have adopted a crossover design, they have failed to clearly describe the random sequences and period of the design and reasonably explain the use of the crossover design. In our study, the CONSORT extension provided a reasonable guideline for crossover trials that should account for these specific issues. In other words, reporting adherence is the first important step in improving the quality of randomized crossover trials.

In addition, we found that the proportion of studies with a high risk of bias was greater than 50% in both $\geq 50\%$ and $< 50\%$ adherence to reporting groups, meaning that the risk of bias was high in general. Therefore, high-quality research requires not only sufficient reporting but also other aspects. Our subgroup analysis showed that research quality was distinct in different condition positions and populations. The risk of bias varied according to whether the trial was registered, the sample size estimated, and the primary end point defined. Taking the sample size estimation as an example, although the randomized number in crossover trials was relatively small, the sample size used in studies with a sample size plan was larger than in that with no pre-estimation, meaning that many trials were underpowered and increased the risk of bias.

Therefore, not only should reporting adherence be improved, but the risk of bias should also be decreased. In a previous study, Gewandter and Ding *et al.* proposed recommendations for evaluating bias in crossover trials from several aspects, such as statistical methods, blinding status, and randomization. In our study, we applied the recent ROB 2.0 for randomized crossover trials to the included studies and found a high proportion of trials with a risk of bias. Compared with the two previous checklists produced by Gewandter

et al. and Ding *et al.*, ROB 2.0 was more formal and complete in terms of evaluating the risk of bias for each outcome. Risk of bias was demonstrated more accurately and in more detail in our study, thus comprehensively explaining the risk of bias in current crossover trials in the field of digestive diseases.

This study had some limitations. First, we restricted the journal impact factor to greater than 3.0. However, we reviewed all crossover trials in pre-specified digestive journals of the past 10 years from three literature databases. Our review included several articles published in high-quality journals. To a large extent, our included articles could represent the average or higher quality and quantity level of the current state of research. Furthermore, we performed correlation analysis and found that the journal impact factor was significantly correlated with reporting adherence (Spearman correlation coefficient = 0.252, $p < 0.001$) and risk of bias (Spearman correlation coefficient = -0.248, $p < 0.001$), though the absolute values of correlation coefficients were not large enough. Second, we may have missed some trials that did not report the second period of the crossover study if they failed to report the original crossover design in their full text. However, deleting many subjects results in inadequate statistical power for the analysis. This indicates a poor study quality and does not change our conclusion. Finally, studies published following the CONSORT statement were fewer than those published before. Specifically, these two groups did not statistically differ in reporting adherence (40.5% *versus* 40.5%, respectively) and in the percentage of studies with a high risk of bias (70.0% *versus* 67.9%, respectively). Our power analysis showed that the data would have a power of above 80% to detect at least an 8% increase in reporting adherence for post-CONSORT studies compared with the pre-CONSORT group. In addition, we did not focus on comparing the trials' other characteristics between the two groups because of the small sample size. We obtained overall insufficient reporting and a high risk of bias from all the included crossover trials.

In conclusion, the study showed general reporting deficiencies and a high risk of bias in digestive disease crossover trials by reviewing 173 randomized crossover studies published in specialized journals in the past decade. Adherence to

reporting was not found to be significantly increased over the years, especially after the publication of the CONSORT extension. In the future, the reporting guidance will have to be followed; however, improving research quality and reducing the risk of bias by improving reporting and other procedures, including specifying primary outcomes, planning sample size, and pre-registration before the trial is started, is more important.

Author contributions

Qian Zhou: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

Zhi-hang Chen: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Validation; Visualization; Writing – original draft; Writing – review & editing.

Jin-xin Zhang: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Validation; Visualization; Writing – original draft; Writing – review & editing.

Sui Peng: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing.

Conflict of interest statement

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Ethical approval

This article does not contain any studies with human participants or animals performed by the authors.

ORCID iD

Qian Zhou  <https://orcid.org/0000-0003-3954-5449>

Supplemental material

Supplemental material for this article is available online.

References

1. Peery AF, Dellon ES, Lund J, *et al.* Burden of gastrointestinal disease in the United States: 2012 update. *Gastroenterology* 2012; 143: 1179–1187. e3.
2. El-Serag HB, Sweet S, Winchester CC, *et al.* Update on the epidemiology of gastro-oesophageal reflux disease: a systematic review. *Gut* 2014; 63: 871–880.
3. Kaplan GG. The global burden of IBD: from 2015 to 2025. *Nat Rev Gastroenterol Hepatol* 2015; 12: 720–727.
4. Pimentel M. Evidence-based management of irritable bowel syndrome with diarrhea. *Am J Manag Care* 2018; 24(Suppl. 3): S35–S46.
5. Louis TA, Lavori PW, Bailar JC, *et al.* Crossover and self-controlled designs in clinical research. *N Engl J Med* 1984; 310: 24–31.
6. Sibbald B and Roberts C. Understanding controlled trials. Crossover trials. *BMJ* 1998; 31: 1719.
7. Dwan K, Li T, Altman DG, *et al.* CONSORT 2010 statement: extension to randomized crossover trials. *BMJ* 2019; 366: 14378.
8. Mills EJ, Chan AW, Wu P, *et al.* Design, analysis, and presentation of crossover trials. *Trials* 2009; 10: 27.
9. Katz PO, Xue S and Castell DO. Control of intragastric pH with omeprazole 20 mg, omeprazole 40 mg and lansoprazole 30 mg. *Aliment Pharmacol Ther* 2001; 15: 647–652.
10. Yoshiji H, Karino Y, Suzuki Y, *et al.* Cross-over study in hyperammonemia patients for efficacy, safety, and acceptability of a new lactulose preparation (SK-1202) compared to approved drug. *Hepatol Res* 2018; 48: 1178–1183.
11. Shariff MK, Bird-Lieberman EL, O'Donovan M, *et al.* Randomized crossover study comparing efficacy of transnasal endoscopy with that of standard endoscopy to detect Barrett's esophagus. *Gastrointest Endosc* 2012; 75: 954–961.
12. Adenis A, Hecquet B and Bonnetterre J. Statistical analysis of crossover studies. *Gastroenterology* 1992; 103: 1994–1995.

13. Sedgwick P. What is a crossover trial? *BMJ* 2014; 348: g3191.
14. Cummings P. Carryover bias in crossover trials. *Arch Pediatr Adolesc Med* 2010; 164: 703–705.
15. Sturdevant SG and Lumley T. Testing for carryover effects after cessation of treatments: a design approach. *BMC Med Res Methodol* 2016; 16: 92.
16. Elbourne DR, Altman DG, Higgins JPT, *et al.* Meta-analyses involving cross-over trials: methodological issues. *Int J Epidemiol* 2002; 31: 140–149.
17. Nolan SJ, Hambleton I and Dwan K. The use and reporting of the cross-over study design in clinical trials and systematic reviews: a systematic assessment. *PLoS ONE* 2016; 11: e0159014.
18. Gewandter JS, McDermott MP, McKeown A, *et al.* Reporting of cross-over clinical trials of analgesic treatments for chronic pain: Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks systematic review and recommendations. *Pain* 2016; 157: 2544–2551.
19. Straube S, Werny B and Friede T. A systematic review identifies shortcomings in the reporting of crossover trials in chronic painful conditions. *J Clin Epidemiol* 2015; 68: 1496–1503.
20. Li T, Yu T, Hawkins BS, *et al.* Design, analysis, and reporting of crossover trials for inclusion in a meta-analysis. *PLoS ONE* 2015; 10: e0133023.
21. Microsoft Academic. Extracted on 4 July 2021, <https://academic.microsoft.com/paper/2955930369/citedby/search?q=CONSORT%202010%20statement%3A%20extension%20to%20randomized%20crossover%20trials&qe=RIId%253D2955930369&f=&orderBy=0>
22. Higgins JPT, Li T, Sterne J, *et al.* (eds) Revised Cochrane risk of bias tool for randomized trials (RoB 2) additional considerations for crossover trials, 2021, <https://www.riskofbias.info/welcome/rob-2-0-tool/rob-2-for-crossover-trials>.
23. Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009; 339: b2535.
24. R Core Team. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, 2021, <https://www.R-project.org/>
25. Schulz KF, Altman DG and Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *J Clin Epidemiol* 2010; 63: 834–840.
26. Ding H, Hu GL, Zheng XY, *et al.* The method quality of cross-over studies involved in Cochrane Systematic Reviews. *PLoS ONE* 2015; 10: e0120519.

Visit SAGE journals online
[journals.sagepub.com/
 home/tag](https://journals.sagepub.com/home/tag)

 SAGE journals