Langenbeck's Archives of Surgery (2022) 407:3793–3802
https://doi.org/10.1007/s00423-022-02649-8

RESEARCH

Methodological and ethical quality of surgical trials from 2016 to 2020

Eloise Papet[1] · Grégoire Moutel[2,3] · Jean Pinson[1] · Matthieu Monge[1] · Edouard Roussel[1] · Tom Teniere[4] · Jean-Jacques Tuech[1] · Valérie Bridoux[1]

Received: 1 May 2022 / Accepted: 4 August 2022 / Published online: 27 August 2022

Abstract

Purpose Randomized controlled trials (RCTs) are the gold standard tool used to evaluate therapeutic interventions. The published results showed that progress still needs to be made not only from a methodological point of view but also from an ethical point of view. The aim of this study was to evaluate the methodological and ethical qualities of randomized clinical trials in surgery over the last few years.

Methods All of the articles chosen for review reported on randomized controlled surgical trials and were published in 10 international journals between 2016 and 2020. Eligible studies were identified, selected, and then evaluated based on a broad set of predetermined criteria. Methodological quality was evaluated using the Jadad scale, and ethical quality was evaluated using the Berdeu score.

Results The methodological quality score (Jadad scale) ranged from 5 to 13, with a mean of 10.0 ± 1.54. The methodological quality was insufficient (score ≤ 9) for 162 trials (31.2%). The ethical quality score ranged from 0.25 to 1, with a mean of 0.8 ± 0.11. Fifty-two articles (10%) did not state that informed consent was requested from the participants, and 21 articles (4%) did not report either research ethics committee or institutional committee protocol approval.

Conclusion The randomized clinical surgical trials analyzed showed that they had satisfactory methods in only 70% of the cases and that they had respected the fundamental ethical principles in 90% of the cases. However, less than 8% of the studies reported planned interim analysis, prospectively defined stopping rules, and independent monitoring committee.

Keywords Randomized controlled trial · Surgery · Ethics · Methodology · Research design

✉ Eloise Papet
Eloise.papet@chu-rouen.fr

Grégoire Moutel
gregoire.moutel@gmail.com

Jean Pinson
jean.pinson@chu-rouen.fr

Matthieu Monge
matthieu.monge@chu-rouen.fr

Edouard Roussel
edouard.roussel@chu-rouen.fr

Tom Teniere
tom.teniere@chu-rouen.fr

Jean-Jacques Tuech
jean-jacques.tuech@chu-rouen.fr

Valérie Bridoux
Valerie.bridoux@chu-rouen.fr

[1] Department of Digestive Surgery, Rouen University Hospital, Rouen, France

[2] UNICAEN, Inserm U1086, ANTICIPE, Normandie Universite, Caen, France

[3] Espace Régional de Réflexion Éthique, CHU Caen, Caen, France

[4] Department of Vascular Surgery, Rouen University Hospital, Rouen, France

Introduction

Surgeons were criticized in the past for their lack of rigor and methodological knowledge in clinical research [1]. Evidence-based surgery has been slow to become a science among those in the medical field [2], whereas evidence-based medicine has proven its effectiveness.

The methodological quality of randomized controlled clinical trials in surgery has been assessed in the past [3–6]. The results showed that progress still needed to be made not only from a methodological point of view but also from an ethical point of view [7], since these two criteria determine the quality of an article. It is the future of our patients who is implied by the elaboration of our articles. Bridoux et al. reported an insufficient methodological quality in 37.4% of the randomized controlled trials in gastrointestinal surgery in 2006–2007 [7], and the mean of the ethical score was low (0.36). It seemed interesting to carry out a new evaluation 10 years later. This study provides an overview of research practices in surgery in recent years. Moreover, the study of methodological and ethical quality in surgery, including all specialties, has never been carried out.

The aim of our study was to assess the methodological and ethical quality of phase III randomized controlled trials (RCTs) in surgery between 2016 and 2020.

We also determined whether specific study characteristics, such as the number of patients and the journal impact factor, were associated with reporting quality.

Materials and methods

Search strategy to identify studies

Ten journals were chosen for the review, consisting of nine English surgical journals (The American Journal of Surgery, the Annals of Surgery, the British Journal of Surgery, the Journal of the American College of Surgeons, Surgery, Surgical Endoscopy, the European Journal of Surgical Oncology, and the World Journal of Surgery) and one English surgical journal (The Journal of the American Medical Association Surgery). To identify eligible articles, we searched the MEDLINE database via PubMed using the limits of article type (randomized controlled trials) with journal names. Additionally, two investigators (VB and EP) independently performed a manual search of each published issue of the 10 aforementioned journals between January 2016 and December 2020. The abstracts were reviewed in duplicate, and the articles considered to be eligible RCTs by either reviewer were further evaluated based on their full text.

A study was defined as an RCT if a control group was present and if the assignment of the participants to the intervention groups was described by one of the following words: random, randomly, randomized, or randomization. The studies included for review were randomized, controlled, phase III surgical trials published between January 1, 2016, and December 31, 2020. All RCTs that compared different treatments were eligible, regardless of the intervention type. No restrictions were made on the trial location, number of patients enrolled in the trial, treatment modalities, or trial sponsor. The exclusion criteria were as follows: (a) trials published as a letter, an abstract, or a short article; (b) randomized phase II trials; (c) non-experimental (observational) studies; (d) trials that referred to previous publications as the source for a detailed description of the trial methods; and (e) animal-based trials.

Characteristics assessed

Two reviewers (VB and EP), who were not involved in any of the identified studies independently, analyzed the identified RCTs. Any disagreement was resolved through discussion between the two reviewers, and consensus with a third researcher was sought (JJT). The standardized protocol was based on a checklist (available from the authors) that included the country of origin, whether the study was industry-funded (yes or no), the number of randomized patients (< 50, from 51 to 100, from 101 to 200, and > 200), the number of centers (< 2, from 2 to 10, and > 10), and the journal impact factor (< 3 to 3, from 3 to 5, from 5 to 10, and > 10).

The selected articles were evaluated for methodological and ethical quality. Methodological quality was evaluated with the Jadad scale [8], which has been used extensively in other clinical areas because of its efficiency [6]. The maximum possible score was 13 points using an 11-item instrument two items allow for an extra point in case of a positive answer.

Items directly related to the control of bias using the Jadad scale are as follows:

1. Was the study designed as randomized?
2. Was the study designed as double blind?*
3. Was there a description of the withdrawals and dropouts?

Other markers not directly related to the control of bias:

4. Were the objectives of the study defined?
5. Were the outcome measures clearly defined?
6. Was there a clear description of the inclusion and exclusion criteria?
7. Was the sample size justified (for example, with power calculation)?

8. Was there a clear description of the interventions?
9. Was there at least one control (comparison) group?
10. Was the method used to assess adverse effects described?
11. Were the statistical analysis methods described?

Items were scored by the following method:

- A score of 1 point was given for each "yes" and 0 points were given for each "no." There were no intermediary marks.
- One additional point was given if, for question 1, the method used to generate the sequence of randomization was described and appropriate (table of random numbers, computer generated, etc.) and/or if, for question 2, the method of double blinding was described and appropriate (identical placebo, active placebo, dummy, etc.).
- One point was deducted if, for question 1, the method used to generate the sequence of randomization was described and inappropriate (patients were allocated alternately, or according to the date of birth, hospital number, etc.) and/or if, for question 2, the study was described as double blind but the method of blinding was inappropriate (for example, comparison of a tablet versus an injection with no double dummy).
- For item 2: en accord avec Probst et al. le double aveugle a été considéré comme patient et surgeon blinding.

A score of more than 9 points was considered to be a good result, and a score equal to or less than 9 points was considered to be a poor result.

Ethical quality was assessed using the Berdeu score [9]. The Berdeu 10-item scale consists of 10 items, and each item can receive one point. The score is obtained by dividing the sum of the individual scores by the maximum possible score, expressed as a decimal number ranging from 0 to 1.

The presence or absence of intermediate analyses was judged as "yes" or "no." If no information was given in the article, it was judged as "unknown." The same reasoning applied to the prospectively defined stopping rules.

The item "placebo ethical justification" was not consider relevant in the surgical trial to calculate the Berdeu scale.

The "instructions to authors" section in each journal was used to determine whether the journals endorsed the Consolidated Standards of Reporting Trials (CONSORT) statement.

Statistical analysis

The statistical analysis was performed using R version 4.0.2 (2020 the R Foundation for Statistical Computing). A Pearson correlation test was performed.

For continuous variables, a *t*-test or ANOVA (depending on the number of levels of the group) was performed according to the assumption of homogeneity of variance was tested using Breusch-Pagan test of heteroskedasticity. If not, the Kruskal–Wallis rank sum test was realized. Either way, the chi-square test of independence was performed for categorical variables. Statistical analysis was achieved using RStudio 1.4.1717 (R Foundation for Statistical).

Ethical aspects

The study was not approved by a research ethics committee, nor did we request informed consent from the authors of the articles because the research did not involve an experimental design using humans.

No financial support was received for this study.

Results

Between January 2016 and December 2020, 655 articles were identified in the 10 journals included for analysis. After reading the abstracts, 520 randomized controlled surgical trials were finally retained (Fig. 1).

Among the 520 surgical articles, the most represented specialty was gastrointestinal surgery, with 326 RCTs published (62.6%). The percentages for other disciplines are detailed in Fig. 1.

Two hundred fifty-two (48.4%) of the RCTs involved a surgical technique, one hundred and twenty-eight (24.6%) involved a medication, and eighty-two (15.8%) evaluated a medical device. The remaining RCTs could not be classified into one of these 3 categories (e.g., RCT assessing pedagogy) (Fig. 2).

Demographics and trial design characteristics

The studies were conducted in a variety of countries and continents: 223 (42.9%) in European countries, 150 (28.8%) in Asia, 131 (25.2%) in American countries with 91 trials in the USA (17.5%), 10 in Australia/New Zealand (1.9%), and 6 in Africa (1.1%).

One hundred and eighty-three (35.2%) of the RCTs were multicentric, with 30 international studies. Among the multicentric studies, 126 involved less than 10 centers (24.2%), and 57 involved more than 10 centers (11%).

The number of patients enrolled in the trials ranged from 10 to 10.010 patients, with a total of 137.511 patients. Of these trials, 60 had ≤ 50 patients, 142 had 51–100 patients, 152 had 101–200 patients, and 166 had more than 200 patients.

The journal impact factor (IF) was < 3 for 70 trials (13.5%) and > 3 for 450 trials.

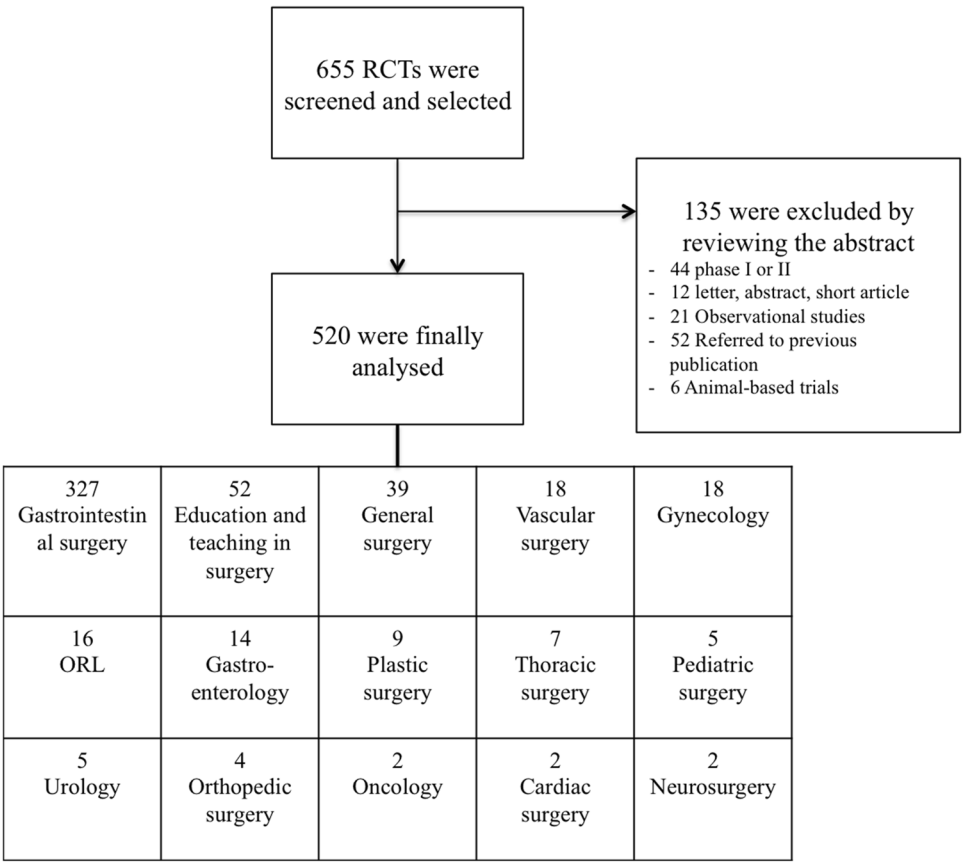

Fig. 1 Flow diagram of the article selection process

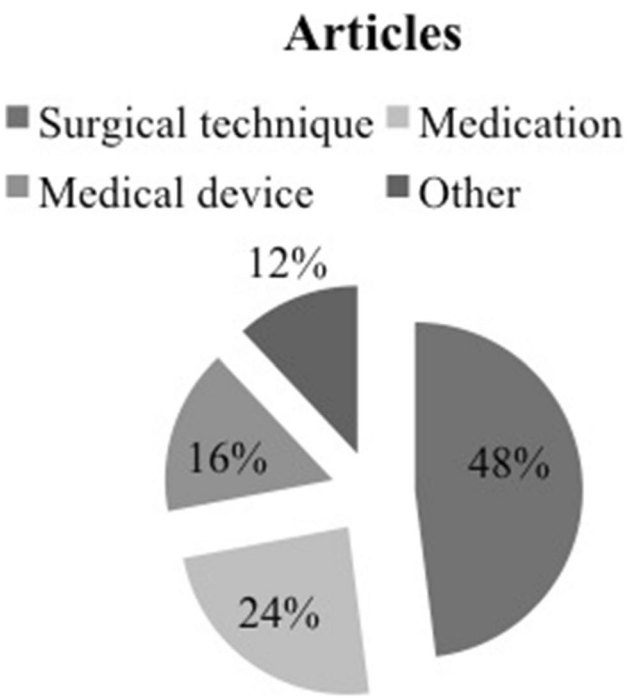

Fig. 2 Distribution of RCTs

Forty-three of the RCTs were published by a cooperative group.

Three hundred and sixty-eight (70.8%) trials disclosed their funding source:

A total of 181 (34.8%) RCTs were non-industry-sponsored trials, 79 (15.2%) were industry-sponsored trials, 54 (10.4%) reported mixed funding (industry and nonindustry), and 54 (10.4%) reported no funding.

Methodological quality

The methodological quality score (Jadad scale) ranged from 5 to 13, with a mean of 10.0 ± 1.54. The methodological quality was insufficient (scores ≤ 9) for 162 trials (31.2%), as shown in Table 1. One hundred and twenty-three trials obtained the maximum possible score. Only 106 articles were described as double-blind and scored an extra 2 points; these articles were eligible for the maximum possible score of 13, and thirty-one achieved this score.

Ethical quality

The ethical quality score ranged from 0.25 to 1, with a mean of 0.8 ± 0.11. The item-by-item frequency of endorsement (Berdeu scale) is summarized in Table 2.

Table 1 Jadad score

Score	1	2	3	4	5	6	7	8	9	10	11	12	13
No. of trials	0	0	0	0	2	9	24	45	82	164	123	40	31

Table 2 Ethical quality

Item	*N* (%)
1. Obtained informed consent from the patients	
Yes	468 (90%)
No	52 (10%)
2. Approval by an REC	
Yes	499 (95.9%)
No	21 (4.1%)
3. Risk–benefit ratio evaluation	
Yes	520 (100%)
No	0
4. Respect for the principle of a priori equivalence	
Yes	520 (100%)
No	0
5. Refusal of consent	
Yes	292 (56.2%)
No	228 (43.8%)
6. Placebo ethical justification*	
Yes	-
No	-
7. Fairness of participant selection (inclusion/exclusion criteria)	
Yes	490 (94.2%)
No	30 (5.8%)
8. Planned interim analysis	
Yes	39 (7.5%)
No	11 (2.1%)
Unknown	470 (90.4%)
9. Prospectively defined stopping rules	
Yes	36 (6.9%)
No	14 (2.7%)
Unknown	470 (90.4%)
10. Independent monitoring committee	
Yes	32 (6.2%)
No	488 (93.8%)

Statistical analysis

There was a low correlation between the methodological and ethical scores ($R = 0.373$, IC95% CI = [0.297, 0.445], p value < 0.001).

We analyzed the correlation between the methodological score and the impact factor, and we found a very low correlation ($R = 0.181$, 95% CI = [0.096, 0.263], p value < 0.001). However, we did not highlight a significant correlation between the ethical score and the impact factor ($R = 0.062$, 95% CI = [− 0.024, 0.148], p value 0.1).

Table 3 Jadad score and the number of patients ($p < 0.0001$)

	$50 \leq N$	$50 < N < 100$	$100 < N < 200$	$200 \leq N$
Mean	9.05	9.86	10.30	10.24
Std. deviation	1.77	1.66	1.32	1.37

Table 4 Berdeu score and the number of patients ($p = 0.0005$)

	$50 \leq N$	$50 < N < 100$	$100 < N < 200$	$200 \leq N$
Mean	0.757	0.794	0.817	0.817
Std. deviation	0.018	0.01	0.007	0.008

There was a significant difference in the Jadad scores between the unicentric studies and the multicentric studies ($p = 0.0008$, mean Jadad score, unicentric = 9.846 ± 1.594 and mean Jadad score, multicentric = 10.32 ± 1.394.) but not in the Berdeu scores.

There was a significant difference in the Jadad score with an increasing number of patients, as well as for the Berdeu score. Increasing the number of patients increased the methodological and ethical quality (Tables 3 and 4).

There was no difference in the Jadad and Berdeu scores regardless of whether the study was industry-funded ($p = 0.4481$ for the Jadad score and $p = 0.6151$ for the Berdeu score).

In the subgroup analysis, there was no difference in the Jadad and Berdeu scores regardless of whether the RCTs assessed the surgical technique (Tables 5 and 6).

A detailed analysis of the scores was performed, and a statistically significant difference was found in the presence or absence of double-blinding. The RCTs that evaluated a surgical technique showed less double-blinding than the RCTs that did not evaluate a surgical technique (Table 5). An analysis of the type of blinding was performed, on the 91 articles considered as "double blind" in the methodology: surgeon was blind in 33% of cases, patient in 90% of cases, outcome assessor in 92% of cases, data collector in 72% of cases, and data analyst in 23% of cases.

No other statistically significant differences were found for methodological or ethical quality (Tables 5 and 6).

However, in a descriptive analysis of the RCTs, a statistically significant difference existed in the presence or absence of a placebo; this difference was present in 14.7% of the nontechnical RCTs and in 0.8% of the technical RCTs, with a p value < 0.001.

Table 5 Subgroup analysis: methodological quality in nontechnical and technical RCTs

	Nontechnicals RCTs ($N=266$)	Technical RCTs ($N=252$)	*P* value
Jadad score (mean ± SD)	**10.0 ± 1.59**	**10.0 ± 1.50**	**0.845**
1. Was the study designed as randomized?			
Yes	266 (100%)	252 (100%)	0.538
The sequence of randomization was described and appropriate			0.103
No	87 (32.7%)	65 (25.8%)	
Yes	179 (67.3%)	187 (74.2%)	
2. Was the study designed as double blind?			
No	200 (75.2%)	212 (84.1%)	**0.0159**
Yes	66 (24.8%)	40 (15.9%)	
The double blinding was described and appropriate			**0.0274**
No	212 (79.7%)	220 (87.3%)	
Yes	54 (20.3%)	32 (12.7%)	
3. Was there a description of the withdrawals and dropouts?			0.0938
No	61 (22.9%)	42 (16.7%)	
Yes	205 (77.1%)	210 (83.3%)	
4. Were the objectives of the study defined?			
Yes	266 (100%)	252 (100%)	0.538
5. Were the outcome measures clearly defined?			
No	5 (1.9%)	6 (2.4%)	0.928
Yes	261 (98.1%)	246 (97.6%)	
6. Was there a clear description of the inclusion and exclusion criteria?			0.214
No	32 (12.0%)	21 (8.3%)	
Yes	234 (88.0%)	231 (91.7%)	
7. Was the sample size justified (for example, with power calculation)?			0.913
No	34 (12.8%)	34 (13.5%)	
Yes	232 (87.2%)	218 (86.5%)	
8. Was there a clear description of the interventions?			0.633
No	2 (0.8%)	4 (1.6%)	
Yes	264 (99.2%)	248 (98.4%)	
9. Was there at least one control (comparison) group?			0.538
Yes	266 (100%)	252 (100%)	
10. Was the method used to assess adverse effects described?			0.937
No	155 (58.3%)	145 (57.5%)	
Yes	111 (41.7%)	107 (42.5%)	
11. Were the statistical analysis methods described?			0.479
No	4 (1.5%)	7 (2.8%)	
Yes	262 (98.5%)	244 (96.8%)	
Unknown	0 (0%)	1 (0.4%)	

The "bold" entries are the main and significant results

Discussion

Research allows us to develop and improve our practices, but it must be based on conclusions with a sufficient level of evidence. The methodological quality of a clinical trial defines the conclusions and practices that can be drawn from it, so it is an integral part of evidence-based surgery.

Moreover, a trial cannot be accepted for publication without respecting the fundamental ethical principles that are at the heart of research today. Although we have come a long way since the Nuremberg Code, there is still a need for regular and unbiased evaluations of our practices to get us back on track if needed.

In this study, total of 69.8% of the RCTs that were analyzed were shown to have satisfactory methods, a proportion shown in a high range of previous studies.

Although this proportion could be overestimated based on the journals from which the RCTs were selected, several

Table 6 Subgroup analysis: ethical quality in nontechnical and technical RCTs

	Nontechnical RCTs ($N=266$)	Technical RCTs ($N=252$)	*P value*
Berdeu score (mean ± SD)	**0.803 ± 0.107**	**0.804 ± 0.108**	**0.895**
1. Obtained informed consent from the patients			0.816
Yes	238 (89.5%)	228 (90.5%)	
No	28 (10.5%)	24 (9.5%)	
2. Approval by an REC			0.444
Yes	253 (95.1%)	244 (96.8%)	
No	13 (4.9%)	8 (3.2%)	
3. Risk–benefit ratio evaluation			0.538
Yes	266 (100%)	252 (100%)	
4. Respect for the principle of a priori equivalence			
Yes	266 (100%)	252 (100%)	
5. Refusal of consent			0.869
Yes	148 (55.6%)	143 (56.7%)	
No	118 (44.4%)	109 (43.3%)	
6. Placebo ethical justification			0.963
Yes	265 (99.6%)	250 (99.2%)	
No	1 (0.4%)	2 (0.8%)	
7. Fairness of participant selection (inclusion/exclusion criteria)			0.45
Yes	253 (95.1%)	235 (93.3%)	
No	13 (4.9%)	16 (6.3%)	
Unknown	0 (0%)	1 (0.4%)	
8. Planned interim analysis			0.867
Yes	19 (7.1%)	20 (7.9%)	
No	5 (1.9%)	6 (2.4%)	
Unknown	242 (91%)	226 (89.7%)	
9. Prospectively defined stopping rules			0.797
Yes	18 (6.8%)	18 (7.1%)	
No	6 (2.3%)	8 (3.2%)	
Unknown	242 (91%)	226 (89.7%)	
10. Independent monitoring committee			1
Yes	16 (6.0%)	16 (6.3%)	
No	250 (94.0%)	236 (93.7%)	

The "bold" entries are the main and significant results

studies evaluated the methodological quality of RCTs in different specialties, but this have never been done for surgical trials in a global way. The proportion of RCTs with good methodological quality varies from 50 to 80% depending on the study [3, 4, 7, 10].

A study that evaluated the quality of RCTs published in the International Urogynecology Journal found that 50.2% of the RCTs had good methodological quality [3].

Another study, published in 2012 by Bridoux et al. [7], reported good methodological quality for 62.6% of the RCTs for gastrointestinal surgery.

Finally, Lai et al. [4] published a study that evaluated the methodological quality of RCTs in ophthalmology with the methods used by Mills et al. (the CONSORT statement modified with 11 items) and reported good methodology quality in 60% of the RCTs.

Moreover, a review of 150 RCTs [10] for endovascular surgery showed an improvement of the methodological quality in the last 10 years. This is most likely due to the increase with respect to the CONSORT statement. These results were in agreement with a study published in 2006, which found that an increase in the use of the CONSORT checklist tends to increase the methodological quality of RCTs [11].

These different studies highlighted heterogeneous results depending on the year and specialties, regardless of whether the CONSORT statement was used. When compared to medical studies, the results were similar [12–14]. Half of the items in the 2010 Overall Quality Score were poorly

reported in at least 40% of the trials evaluating combined chemoradiotherapy for nasopharyngeal carcinoma. Chen et al. concluded that the reporting quality for NPC trials is unsatisfactory and that investigators should be more aware of the CONSORT items.

In a study published by Lu et al. evaluating the quality of reporting in randomized controlled trials conducted in China for the treatment of cancer pain, they reported that only one of the papers mentioned sample size calculation (2.2%).

A systematic review that assessed the benefits and harms of surgical interventions in randomized clinical trials published in the British Journal of Surgery in 2020 by Stubenrouch et al. [15] reported that although the CONSORT statement is supported widely, adherence to this guideline needs to be improved to facilitate evidence-based clinical decision-making.

Although the CONSORT statement has been designed to optimize the reporting of trials, this systematic review found that current publications still show suboptimal reporting of data.

One of the limitations of our study was the assessment of the reports of RCTs and not of the RCTs themselves, and the authors of the original articles were not asked to provide additional information on the methodological and ethical details that were not adequately reported in the articles.

Moreover, the choice to use these 10 journals was to create a homogeneous group of publications and conditions that allowed standardized analysis, but this arbitrary choice may have introduced a bias causing overestimation of the quality of the analyzed RCTs.

In this study, we assessed RCTs comparing surgical and non-surgical interventions. The main difference in the assessment between medicine and surgery is related to the "operating room" technique. This is the reason that we performed a subgroup analysis between the RCTs that assessed the surgical technique (technical RCTs) and those that did not (nontechnical RCTs). There is no difference in the methodology, whereas we might think that the nonsurgical treatments tend to be better analyzed and monitored before the outcomes are reported. The only difference showed is that the presence of double blinding is more important in nontechnical RCTs, which is explained by the difficulty of blindness for surgery. An article published by Probst et al. [16] showed a blinding of the patient of 24.6% which is consistent with our results since it represents 20% in our study. Moreover, this article published recommendations for future use of blinding and recommendations for reporting of blinding in surgical trials, and especially for surgical systematic reviews, it is recommended to describe the blinding measures and their impact on endpoints for every study contributor individually.

After reviewing the various articles in the literature on the subject, the multitude of scales used to assess methodological quality, and therefore the difficulty of drawing conclusions, is highlighted. The CONSORT assessment, although supported by most journals, still tends to be missing in RCTs. In 2010, the question "Do we practice what we preach [17]?" was asked. The answer was that in one-third of the surgical units, what is published is not practiced. Is it then because of a lack of confidence in the results?

Another limitation of the study is the utilization of the Jadad scale.

Three methods are used to assess quality: individual markers, checklists, and scales. Scales have the theoretical advantage over the other methods in that they provide quantitative estimates of quality that can be replicated easily. The main disadvantage of quality scales is that there is a dearth of evidence supporting either the inclusion or exclusion of items or the numerical scores attached to each of those items. Calculating a summary score inevitably involves assigning "weights" to different items in the scale, and it is difficult to justify the weights assigned.

The total ethical score (Berdeu scale) calculated for the 520 studied RCTs was 0.8 ± 0.11 and demonstrates real improvement. This score is higher than the score obtained by Berdeu et al. [18] for clinical trials involving elderly patients (0.334 ± 0.118), the score obtained by Tuech et al. [18] for oncology trials (0.42 ± 0.133), and the score published in 2012 [7] for gastrointestinal surgery trials (0.36 ± 0.08).

Of the 520 clinical trials investigated in this study, 52 (10%) did not state that informed consent was requested from the participants, and 21 (4%) did not report receiving research ethics committee or institutional committee protocol approval. As explained above, it is difficult to know whether consent was obtained since we only reviewed the reports of RCTs.

The prior assessment of the individual and collective benefit/risk ratio was respected for all the articles studied, a score that has never been described. In addition, the ethical rationale for using or not using a placebo was not described for 3 clinical trials (0.6%).

This score demonstrates a considerable improvement in practices from an ethical point of view. Respect for the individual is now at the center of research activity.

In RCTs, an independent monitoring committee reviews the data, performs an interim analysis, and decides whether to close the study on the basis of predetermined early stopping rules that relate to toxicity or outcomes. These predetermined rules must be defined before the trial and put in place to protect the participating patients [19]. If excess harm is observed or if a statistically significant benefit is observed, the study is stopped early, and the patients are informed of the results. If the treatment is ongoing, the patient is typically offered the opportunity to receive the regimen that is perceived to be superior [20].

In our study, we found 42 (8%) RCTs with prespecified interim analyses and 32 (6.1%) RCTs with a reported data safety and monitoring board (DSMB).

A study published in 2016 by Stegert et al. [21] analyzed 894 RCT protocols. They reported 32.3% of the RCTs had prespecified interim analyses and 28.7% had a DSMB.

This higher incidence of interim analysis can be explained by the fact that the selected RCTs were all approved by research ethics boards, with further analyses of their protocols. Furthermore, in this study, we did not know the proportion of surgical and nonsurgical trials. The hypothesis that surgical trials are less compliant with the development of interim analysis cannot be asserted.

Conclusion

The randomized clinical surgical trials that were analyzed were shown to have satisfactory methods in only 70% of cases and have respect fundamentals ethical principles in 90% of cases. However, less than 8% of the studies reported planned interim analysis, prospectively defined stopping rules, and independent monitoring committee.

We believe that efforts must be made by the surgical community in order to improve both methodological and ethical quality of their trials.

Acknowledgements The authors acknowledge the dedicated efforts of Pierre Blavier for his statistical analysis.

Authors' contributions Study conception and design: E. Papet, G. Moutel, V. Bridoux, J. J. Tuech. Acquisition of data: E. Papet, J. Pinson, M. Monge, E. Roussel, T. Teniere. Analysis, or interpretation of data: E. Papet, J. Pinson, M. Monge, E. Roussel, T. Teniere. Drafting of the manuscript: E. Papet. Critical revision of the manuscript: E. Papet, G. Moutel, V. Bridoux, J. J. Tuech, T. Teniere. Administrative, technical, or material support: G. Moutel, V. Bridoux, J. J. Tuech, J. Pinson, M. Monge, E. Roussel, T. Teniere.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Hüttner FJ, Doerr-Harim C, Probst P, Tenckhoff S, Knebel P, Diener MK et al (2014) Study methods in evidence-based surgery: methodological impediments and suggested approaches for the creation and transfer of knowledge in surgery. Eur Surg Res 53(1–4):86–94
2. Maheshwari G, Maheshwari N et al (2012) Evidence based surgery: how difficult is the implication in routine practice? Oman Med J 27(1):3
3. Kim KS et al (2018) Quality of randomized controlled trials published in the International Urogynecology Journal 2007–2016. Int Urogynecol J 29(7):1011–1017. https://doi.org/10.1007/s00192-017-3465-6
4. Lai TYY, Wong VWY, Lam RF, Cheng ACO, Lam DSC, Leung GM et al (2007) Quality of reporting of key methodological items of randomized controlled trials in clinical ophthalmic journals. Ophthalmic Epidemiol 14(6):390–398. https://doi.org/10.1080/09286580701344399
5. Ahmed Ali U et al (2013) Trends in worldwide volume and methodological quality of surgical randomized controlled trials. Ann Surg 258(2):199–207. https://doi.org/10.1097/SLA.0b013e31829c7795
6. Zeng X et al (2015) The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review: Methodological quality assessment tools. J Evid-Based Med 8(1):2–10. https://doi.org/10.1111/jebm.12141
7. Bridoux V et al (2012) Methodological and ethical quality of randomized controlled clinical trials in gastrointestinal surgery. J Gastrointest Surg 16(9):1758–1767. https://doi.org/10.1007/s11605-012-1952-0
8. Jadad AR et al (1996) Assessing the quality of reports of randomized clinical trials: Is blinding necessary? Control Clin Trials 17(1):1–12. https://doi.org/10.1016/0197-2456(95)00134-4
9. Berdeu D, Alric R, Hervé C, Fourcade J et al (1993) Qualité éthique et qualité méthodologique des essais contrôlés randomisés: analyse à partir des études publiées de 1989 à 1992 dans quatre revues internationales. La Revue de Médecine Interne 14(10):1070. https://doi.org/10.1016/S0248-8663(05)80185-7
10. Hajibandeh S, Hajibandeh S, Antoniou GA, Green PA, Maden M, Torella F et al (2015) Reporting and methodological quality of randomised controlled trials in vascular and endovascular surgery. Eur J Vasc Endovasc Surg 50(5):664–670. https://doi.org/10.1016/j.ejvs.2015.06.114
11. Plint AC et al (2006) Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Med J Aust 185(5):263–267. https://doi.org/10.5694/j.1326-5377.2006.tb00557.x
12. Alam M, Rauf M, Ali S, Nodzenski M, Minkis K et al (2014) A systematic review of reporting in randomized controlled trials in dermatologic surgery: Jadad scores, power analysis, and sample size determination. Dermatol Surg 40(12):1299–1305. https://doi.org/10.1097/DSS.0000000000000166
13. Lu L, Zeng J, Chen Y et al (2011) Quality of reporting in randomized controlled trials conducted in China on the treatment of cancer pain. Expert Rev Anticancer Ther 11(6):871–877. https://doi.org/10.1586/era.10.236
14. Chen Y-P et al (2017) Reporting quality of randomized, controlled trials evaluating combined chemoradiotherapy in nasopharyngeal carcinoma. Int J Radiat Oncol *Biol* Phys 98(1):170–176. https://doi.org/10.1016/j.ijrobp.2017.01.214
15. Stubenrouch FE, Cohen ES, Bossuyt PMM, Koelemay MJW, Vetvander PCR, Ubbink DT et al (2020) Systematic review of reporting benefits and harms of surgical interventions in randomized clinical trials. BJS Open 4(2):171–181. https://doi.org/10.1002/bjs5.50240
16. Probst P et al (2019) Evidence-based recommendations for blinding in surgical trials. Langenbecks Arch Surg 404(3):273–284. https://doi.org/10.1007/s00423-019-01761-6
17. Kain N, Kotur K, Gupta S et al (2011) Evidence based surgery — do we practice what we preach? Surgeon 9(4):187–190. https://doi.org/10.1016/j.surge.2010.09.002
18. Tuech JJ (2005) Methodological quality and reporting of ethical requirements in phase III cancer trials. J Med Ethics 31(5):251–255. https://doi.org/10.1136/jme.2003.007435
19. Henschel AD, Rothenberger LG, Boos J et al (2010) Randomized clinical trials in children — ethical and methodological issues. CPD 16(22):2407–2415. https://doi.org/10.2174/138161210791959854

20. Cannistra SA (2004) The ethics of early stopping rules: who is protecting whom? JCO 22(9):1542–1545. https://doi.org/10.1200/JCO.2004.02.150
21. Stegert M et al (2016) An analysis of protocols and publications suggested that most discontinuations of clinical trials were not based on preplanned interim analyses or stopping rules. J Clin Epidemiol 69:152–160. https://doi.org/10.1016/j.jclinepi.2015.05.023