

# Alternative Randomized Trial Designs in Surgery

## A Systematic Review

Simone Augustinus, MD,\*† Iris W.J.M. van Goor, MD,‡§  
 Johannes Berkhof, PhD, MSc,|| Lois A. Daamen, MD, PhD, MSc,‡¶  
 Bas Groot Koerkamp, MD, PhD, MSc,# Tara M. Mackay, MD, PhD, MSc,\*†  
 I.Q. Molenaar, MD, PhD,‡ Hjalmar C. van Santvoort, MD, PhD,‡  
 Helena M. Verkooijen, MD, PhD, MSc,# Peter M. van de Ven, PhD, MSc,||\*\*  
 and Marc G. Besselink, MD, PhD, MSc\*†

**Introduction:** Randomized controlled trials (RCTs) yield the highest level of evidence but are notoriously difficult to perform in surgery. Surgical RCTs may be hampered by slow accrual, the surgical learning curve, and lack of financial support. Alternative RCT designs such as stepped-wedge randomized controlled trials (SW-RCTs), registry-based randomized controlled trials (RB-RCTs), and trials-within-cohorts (TwICs) may overcome several of these difficulties. This review provides an overview of alternative RCT designs used in surgical research.

**Methods:** We systematically searched PubMed, EMBASE, and Cochrane Central for surgical SW-RCTs, RB-RCTs, and TwICs. A surgical RCT was defined as a randomized trial that studied interventions in patients undergoing general surgery, regardless of the affiliation of the corresponding author. Exponential regression analysis was performed to assess time trends.

**Results:** Overall, 41 surgical RCTs using alternative designs were identified, including 17 published final RCT reports and 24 published protocols of ongoing RCTs. These included 25 SW-RCTs (61%), 13 RB-RCTs (32%), and 3 TwICs (7%). Most of these RCTs were performed in Europe (63%) and within gastrointestinal/oncological surgery (41%). The total number of RCTs using alternative designs

exponentially increased over the last 7 years ( $P < 0.01$ ), with 95% ( $n = 39/41$ ) of the total number published within this time frame. The most reported reasons for using alternative RCT designs were avoidance of contamination for SW-RCTs and generalizability of the trial population for RB-RCTs and TwICs.

**Conclusions:** Alternative RCT designs are increasingly used in surgical research, mostly in Europe and within gastrointestinal/oncological surgery. When adequately used, these alternative designs may overcome several difficulties associated with surgical RCTs.

**Keywords:** randomized controlled trials, registry-based, stepped-wedge, surgical research, trials-within-cohorts

(*Ann Surg* 2022;276:753–760)

Randomized controlled trials (RCTs) provide the highest level of evidence in clinical practice.<sup>1</sup> However, surgical RCTs are notoriously difficult to perform, mainly due to poor recruitment, patients dropout in the control arm, and high costs.<sup>2–4</sup> These problems are quite common, resulting in 1 in 5 surgical RCTs being discontinued early and 1 in 3 completed surgical RCTs remaining unpublished.<sup>5</sup> For surgical research specifically, additional problems occur due to surgical learning curves, poor generalizability of the trial population, problems with blinding, and difficulties with randomization in life-threatening situations.<sup>6</sup> Thereby, surgical RCTs appear to have moderate impact on daily surgical practice with only 47% of surgeons adhering to recommendations of specific RCTs in clinical practice.<sup>7</sup> In recent years, several alternative RCT designs have been introduced such as stepped-wedge randomized controlled trials (SW-RCTs), registry-based randomized controlled trials (RB-RCTs), and cohort-multiple RCTs, also called trials-within-cohorts (TwICs), which address several of these problems (Table 1).<sup>8–10</sup>

In SW-RCTs, the intervention is sequentially rolled-out and clusters, such as hospitals or hospital wards, switch from standard care to the intervention at different times in a randomized order. Patient inclusion continues throughout the study period so that each cluster contributes to both the control and intervention groups.<sup>11</sup> In RB-RCTs, an existing prospective patient registry is used. Randomization takes place according to the conventional RCT design and data are collected in the existing registry.<sup>12</sup> In TwICs, an existing prospective cohort is used in which outcome measures are collected. Eligible patients are identified from the cohort and consequently randomized for a new intervention or to continue standard care. Only patients randomized to the new intervention are asked for a (second)

From the \*Department of Surgery, Amsterdam UMC, location University of Amsterdam, The Netherlands; †Cancer Center Amsterdam, Amsterdam, The Netherlands; ‡Department of Surgery, Regional Academic Cancer Center Utrecht, University Medical Center Utrecht, Cancer Center & St. Antonius Hospital Nieuwegein, Utrecht University, Utrecht, The Netherlands; §Department of Radiation Oncology, Regional Academic Cancer Center Utrecht, University Medical Center Utrecht, Cancer Center, Utrecht University, Utrecht, The Netherlands; ||Department of Epidemiology and Data Science, Amsterdam UMC, location Vrije Universiteit van Amsterdam, Amsterdam, The Netherlands; #Department of Surgery, Erasmus Medical Center, Rotterdam, The Netherlands; ¶Division of Imaging, University Medical Center Utrecht Cancer Center, Utrecht University, Utrecht, The Netherlands; and \*\*Department of Data Science and Biostatistics, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands.

✉m.g.besselink@amsterdamUMC.nl.

P.M.v.d.V. and M.G.B. shared senior authorship.

The authors report no conflicts of interest.

Supplemental Digital Content is 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 website, [www.annalsofsurgery.com](http://www.annalsofsurgery.com).

This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2022 The Author(s). Published by Wolters Kluwer Health, Inc. ISSN: 0003-4932/22/27605-0753

DOI: 10.1097/SLA.0000000000005620

**TABLE 1.** Overview of RCT Designs

	Informed Consent	Randomization	Data Collection
Classic RCT	Before enrollment in the trial (all patients)	Individual randomization and informed consent of all patients	Through (electronic) case report forms specific for the trial
SW-RCT	Depending on the intervention, individual informed consent can be waived due to clusters being randomized	Clusters switch from standard care to intervention at time points. The order in which clusters switch from standard care to intervention is randomized	According to conventional RCT design, or an existing registry or prospective cohort is used in which patient data are routinely registered
RB-RCT	According to conventional RCT design	According to conventional RCT design	An existing registry is used in which patient data are routinely registered
TwICs	Broad informed consent for data collection and TwiCs before enrollment in cohort. A second informed consent when a patient is randomized to the intervention arm	Patients meeting the inclusion criteria are identified within the cohort and consequently randomized	An existing registry or prospective cohort is used in which patient outcome data are routinely collected

informed consent. Patients randomized to the comparator arm are not informed and continue standard care.<sup>8</sup>

A systematic review of the experience and use of these alternative RCT designs in surgical research is lacking. Furthermore, the value and suitability of innovative trial designs in surgical research are unclear. The aim of this systematic review is to provide an overview of the experience and use of alternative RCT designs in surgical research, including reported motivations and limitations.

## METHODS

This systematic review was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines.<sup>13</sup>

### Search Strategy and Selection Process

The literature search strategy was developed with an experienced clinical librarian to identify eligible trials in PubMed, EMBASE, and Cochrane Central from inception. The search included various synonyms for the words “stepped-wedge,” “registry-based,” “trial-within-cohort,” and “surgical procedures” (Supplement 1, <http://links.lww.com/SLA/E66>). Inclusion criteria comprised: (1) final RCT reports and published trial protocols of ongoing RCTs; (2) either SW-RCTs, RB-RCTs, or TwiCs; (3) surgical intervention trials. Exclusion criteria comprised: (1) unpublished trial protocols; (2) conference abstracts; (3) secondary publications of previously published trials; (4) commentaries/letters to the editor; (5) publications in languages other than English.

Two reviewers (S.A. and I.W.J.M.v.G.) independently screened all abstracts for relevance. Subsequently, full-text versions of all relevant studies were reviewed and a final selection was made. Disagreements were resolved by reaching a consensus by a third reviewer (M.G.B. and P.M.v.d.V.). The methodological quality was evaluated by 2 independent reviewers (S.A. and I.W.J.M.v.G.) using the Cochrane Risk of Bias 2 tool (RoB-2) for RCTs. Methodological quality was only assessed for final RCT reports, not for protocols of ongoing RCTs, as not all domains could be evaluated for protocols.<sup>14</sup> Adherence to reporting guidelines specific for the design was reported for SW-RCTs and defined as specific reference to the CONSORT extension for either cluster randomized trials or stepped-wedge designs.<sup>15,16</sup> This was only assessed for SW-RCTs published after 2010, the year in which this extension was published. Adherence to reporting guidelines was not assessed for RB-RCTs and TwiCs as a CONSORT extension for these trials was only published in 2021.<sup>17</sup>

### Definitions and Extraction of Data

The definitions used and study data extracted can be found in Supplement 2 (<http://links.lww.com/SLA/E67>).

### Statistical Analysis

Descriptive analyses were used to summarize study characteristics per design. Characteristics were summarized as frequencies with proportions for binary or categorical variables, or as mean with SD or median with interquartile range or range for continuous variables as appropriate. The number of innovative RCTs per type of design per year was calculated and depicted graphically. Exponential regression analysis was performed to assess the change in the number of published studies over the years 2015 to 2021. Reported motivations for use and limitations were ordered according to the number of times reported, separate for each type of design.

## RESULTS

The literature search was performed on February 3, 2022, and identified 4431 articles. After title and abstract screening, full-text screening of 159 articles was performed. In total, 41 surgical RCTs met the inclusion criteria (Fig. 1).

### Overview

Among the 41 included surgical RCTs, 17 were final RCT reports, and 24 were published protocols of ongoing RCTs (Supplement 3, <http://links.lww.com/SLA/E68>). Overall, 25 trials were SW-RCTs (61.0%), 13 were RB-RCTs (31.7%), and 3 were TwiCs (7.3%). Seven out of 25 SW-RCTs (28.0%) and all 3 TwiCs (100%) used data from existing registries for their data collection. Most trials were initiated in Europe (n=26, 63.4%), followed by the United States (n=10, 24.4%), Canada (n=3, 7.3%), and New Zealand (n=2, 4.9%). For individual countries, most trials were initiated in The Netherlands (n=12, 29.3% in all countries and 48% in Europe). Trials were published between 1999 and 2021, of which the vast majority (95.1%) were published since 2015. Since 2015, the volume of innovative trial designs increased exponentially with time (Fig. 2,  $R^2=0.80$ ,  $F_{1,6}=24.7$ ,  $P<0.01$ ). Most trials were performed in the field of gastrointestinal/oncological surgery (41.5%). For SW-RCTs, the majority of trials investigated non-therapeutic interventions (88.0%), whereas for RB-RCTs (76.9%) and TwiCs (100%) the majority investigated therapeutic interventions. Median duration of recruitment was 28 months (interquartile range: 19.2–45 months) for published studies. Of the unpublished trials, current recruitment status was known for 22 of 24 trials. Most of these studies had finished recruiting patients (n=14, 63.6%), 6 studies (26.1%) were currently recruiting

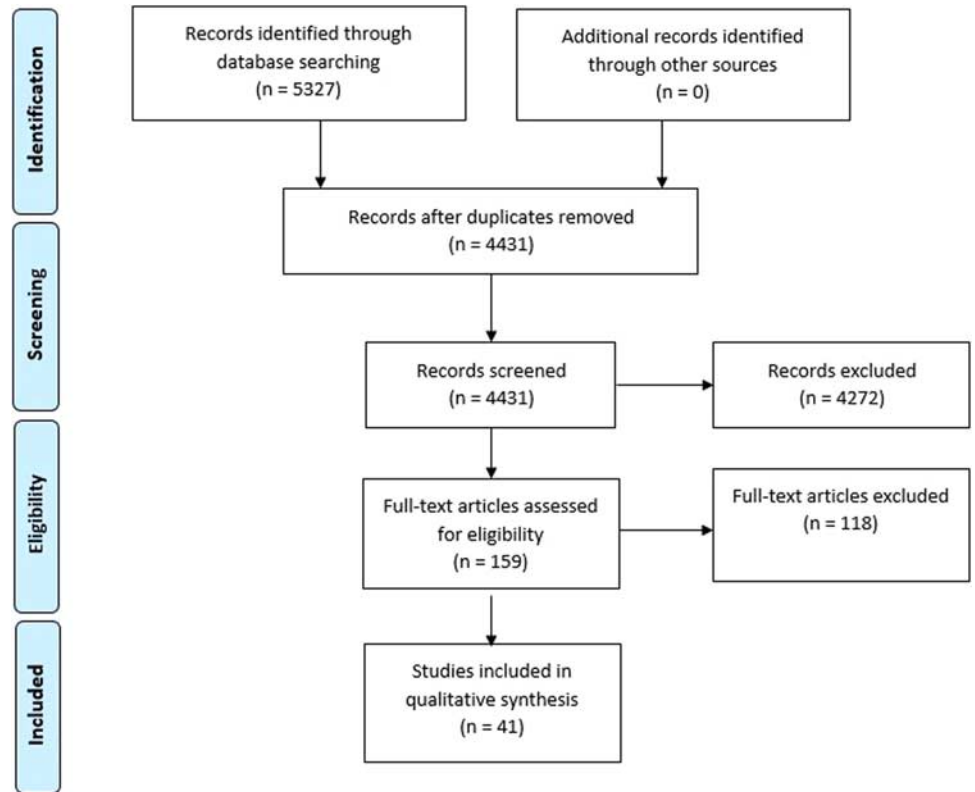


FIGURE 1. PRISMA flow diagram.

patients, and only 2 studies (9.1%) were stopped prematurely, of which 1 (4.2%) was stopped because of slow accrual (Supplement 4, <http://links.lww.com/SLA/E69>).

### Risk of Bias

Risk of bias was considered low for 10 out of 17 final RCT reports (58.8%), although some concerns were reported<sup>18–29</sup> (Table 2). Concerns were mostly related to assessors not being blinded for the intervention, although it was acknowledged that this was unlikely to influence the results. Of the 24 final RCT reports and protocols using the SW-RCT design (published after 2010), only 6 (25.0%) reported the use of a CONSORT extension.

### Reported Motivations

The most reported reason for choosing the SW-RCT design was the minimization of contamination (for all reported motivations, Supplement 5, <http://links.lww.com/SLA/E70>). For example, one trial implemented a national histopathology service to aid the selection of better quality kidneys. If this trial would randomize individual kidneys with or without the option to use the histopathology service, the extra information for only some kidneys would probably “contaminate” (ie, change) acceptance practice of kidneys offered without it.<sup>30</sup>

The most reported reason for using a RB-RCT design, also when this was incorporated in a SW-RCT or TwiCs study, was improved generalizability of the trial (for all reported motivations, Supplement 6, <http://links.lww.com/SLA/E71>). For example, by facilitating the inclusion of a relatively large group of patients and embedding the research question in the clinical practice (ie, the registry).<sup>31,32</sup>

The most reported reason for using TwiCs, besides the advantages of a registry, was avoidance of disappointment bias.

As only patients randomized to the intervention group are made aware of the randomization result, patients will not be disappointed and dropout of the study due to randomization into the control arm.<sup>33</sup>

### Reported Limitations

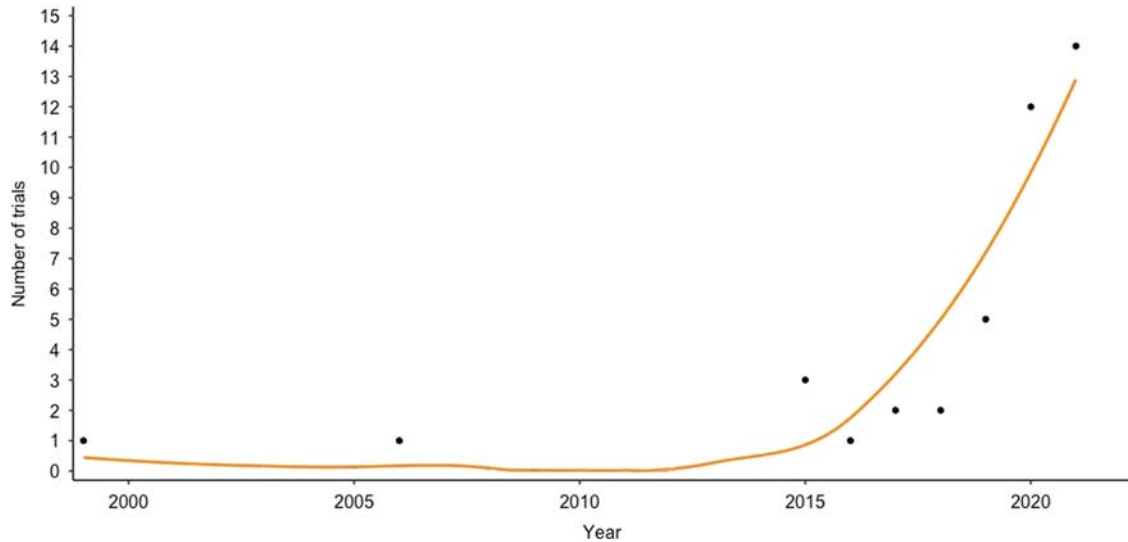
For SW-RCTs, the most common limitation was confounding of treatment effect by time. (for all reported limitations, Supplement 7, <http://links.lww.com/SLA/E72>). For example, 1 SW-RCT implemented an intensified follow-up schedule to detect recurrence after curative colorectal cancer treatment. However, as incidence of recurrence tends to change over time during follow-up, it can never be known whether the observed effects are completely due to the intervention.<sup>34</sup>

For RB-RCTs, the most frequently reported limitation was that the trial was limited to the variables recorded in the registry and the patient population included in the registry (for all reported limitations, Supplement 8, <http://links.lww.com/SLA/E73>). For example, in 1 RB-RCT trial available long-term outcomes were limited to allograft failure and patient death, but outcomes of kidney function, metabolic complications, cardiovascular event, and infections were also of interest but not registered.<sup>35</sup>

For TwiCs, selective patient refusal was mentioned as a limitation. Patients allocated to the intervention arm have to sign a second informed consent but may refuse the intervention, which dilutes the effect estimates in intention-to-treat analyses.<sup>33</sup>

## DISCUSSION

This first systematic review of alternative RCT designs in surgical research identified 41 RCTs, including 17 published and 24 ongoing trials. Of these, the majority were SW-RCTs (61.0%),



**FIGURE 2.** Time trend in published surgical RCTs and RCT protocols with alternative designs.

13 were RB-RCTs (31.7%), and 3 were TwiCs (7.3%). Most trials were initiated in Europe and performed within gastrointestinal/oncological surgery. The vast majority of alternative RCTs were published after 2015 demonstrating that their use is increasing rather rapidly.

The alternative RCT designs were introduced to overcome challenges encountered within classic surgical RCTs. One of the biggest challenges is poor patient accrual, leading to long inclusion times and high costs.<sup>5</sup> The most effective and most widely used design for overcoming this challenge seems to be SW-RCT. In SW-RCTs, whole clusters are being randomized, often without the requirement of individual informed consent.<sup>11</sup> Also, RB-RCTs may improve patient accrual as they allow for broader inclusion criteria, and recruiting from multiple providers and regions. Thereby, centers are generally more willing to participate as costs are minimal when data collection is already ongoing within the registry.<sup>36</sup> In TwiCs, the dropout of patients randomized to the control group is almost nonexistent.<sup>37</sup> Therefore, each of these alternative trial designs potentially increase the accrual rate. This is supported by the relatively short recruitment period of 28.0 months for the published studies included and two third of the ongoing RCTs already having completed recruitment (63.6%). However, these results should be interpreted with caution because of potential selection bias. Authors may be more likely publish protocols of studies expected to finish recruitment. Unpublished protocols of ongoing RCTs are not included in this review.

When deciding between designing a “classical” or an “alternative” RCT design, surgeon-scientists should consider which specific challenges are relevant for their study on a case-by-case basis. Problems of the learning curve and blinding are not solved by the alternative designs. Guidelines should be developed assessing the suitability of the alternative trial designs for each setting, for example, by performing a Delphi study including surgeons and epidemiologists with expertise on the topic. Below we further explain the merits per alternative design with examples (summarized in Table 3).

SW-RCTs are best suited for trials in which there is a strong evidence or belief that the new intervention is beneficial and a decision has already been made to implement this

intervention across all clusters.<sup>41</sup> This scenario mostly applies to nontherapeutic interventions (88.0% of SW-RCTs in this review investigated nontherapeutic interventions) such as quality improvement programs or best practice implementations. Notably, these programs or pathways in themselves may include specific interventions. A textbook example is the evaluation of a national quality improvement program that implements a care pathway for emergency abdominal surgery.<sup>38</sup> The main advantage reported in the present systematic review was minimization of contamination and SW-RCTs are associated with high participation rates as all clusters are automatically exposed to the new intervention. Whereas the main limitation reported was confounding of treatment effect by time. A previous review focusing specifically on statistical methods in SW-RCTs indicates that only 33% of SW-RCTs corrected for time effects. Since the proportion of subjects in SW-RCT that receive the new intervention increases over time and outcomes generally depend on time or confounding factors changing over time, a correction for time is essential in SW-RCTs.<sup>42,43</sup> Within these reviews, additional statistical problems for SW-RCTs are mentioned (which are not reported in the present review), only 75% of SW-RCTs reported a sample size calculation and only 73% adjusted for clustering.<sup>42,43</sup> Therefore, when performing SW-RCTs we advise to involve an experienced statistician to improve (methodological) quality.

RB-RCTs are best served for pragmatic trials requiring large numbers of patients representative of a real-world clinical population to show effectiveness on outcomes collected in routine care.<sup>9</sup> As a textbook example, the randomized single-center mass screening trial for abdominal aortic aneurysm, randomized 12,639 patients to either an abdominal ultrasound scan or the control group with outcomes collected from Danish registries.<sup>39</sup> The advantages and limitations of surgical RB-RCTs in this review are in agreement with articles evaluating the RB-RCTs in general.<sup>9,36,44</sup> RB-RCTs primarily overcome the problem of poor generalizability of the trial population, as registries are generally less restrictive than standard trial inclusion criteria. Important limitations mainly concern the lack of detailed data and internal validity. Only in 11.3% of the RB-RCTs, the quality of the registry data is mentioned.<sup>36</sup> Therefore, when performing

**TABLE 2.** Risk of Bias of 17 Published Final Reports of RCTs With Alternative Designs

References	Domain 1a: Randomization Process	Domain 1b: Cluster	Domain 2: Deviations From Intended Interventions	Domain 3: Missing Outcome	Domain 4: Measurement of Outcome	Domain 5: Selection of the Reported Result	Overall: Risk of Bias
Buhre et al <sup>18</sup>	●	●	●	●	●	●	●
Couwenberg et al <sup>13</sup>	●	NA	●	●	●	●	●
Deeken et al <sup>19</sup>	●	●	●	●	●	●	●
Gilbert et al <sup>20</sup>	●	●	●	●	●	●	●
Lashoher et al <sup>21</sup>	●	●	●	●	●	●	●
Masters et al <sup>22</sup>	●	NA	●	●	●	●	●
Møller et al <sup>23</sup>	●	NA	●	●	●	●	●
Peden et al <sup>18</sup>	●	●	●	●	●	●	●
Petro et al <sup>24</sup>	●	NA	●	●	●	●	●
Scharze et al <sup>25</sup>	●	●	●	●	●	●	●
Sier et al <sup>26</sup>	●	●	●	●	●	●	●
Tastaldi et al <sup>27</sup>	●	NA	●	●	●	●	●
Verberne et al <sup>14</sup>	●	●	●	●	●	●	●
Wilmink et al <sup>18</sup>	●	●	●	●	●	●	●
Woodle et al <sup>15</sup>	●	NA	●	●	●	●	●
Zatzick et al <sup>19</sup>	●	●	●	●	●	●	●

Black: low risk of bias; gray: some concerns.  
NA indicates not applicable.

a surgical RB-RCT, the potential lack of quality of the registry data should be carefully weighed against the advantage of efficient data collection, or measures should be taken to improve the quality. Some registries may be able to add trial-specific variables temporarily to the registry.

The TwiCs design is best used when dropout in standard RCTs is expected to high for subjects randomized to the control arm and when multiple new treatments for the same condition are expected to be evaluated (almost) simultaneously.<sup>45</sup> As a textbook example, the MEDOCC-CrEATE study investigates how many stage II colorectal cancer patients with detectable circulating tumor DNA after surgery will accept adjuvant chemotherapy and whether this reduces the risk of recurrence in these patients.<sup>40</sup> Data are collected and patients are identified in the Prospective Dutch Colorectal Cancer Cohort (PLCRC). Main advantages reported in this review are conform literature, namely the increase of inclusion rates by reducing dropout due to disappointment bias and more efficient data collection due to the large number of potential controls.<sup>45,46</sup> In terms of limitations, the number of patients refusing participation in the intervention arm is usually higher than in a conventional RCT due to the fact that additional informed consent is obtained after randomization.<sup>46,47</sup> However, given that recruitment is more easy, this most likely weights up to the patients refusing participation in the intervention arm.<sup>48</sup> Thereby, the percentage of patients accepting the intervention can reflect the acceptance of patients in current clinical practice<sup>46</sup>. However, selective refusal should be taken into account in the sample size calculation to avoid low statistical power.<sup>48</sup> Furthermore, the study design should be explained well to both researchers and patients.

The use of reporting guidelines, including the Consolidated Standards of Reporting Trials (CONSORT) statement,

improves reporting quality and enables adequate assessment of RCTs.<sup>49</sup> In 2010, a CONSORT extension for SW-RCTs has been published, and in 2021 an extension for trials was conducted using cohorts and routinely collected data (CONSORT-ROUTINE).<sup>15,17</sup> In this review, only 25% of SW-RCTs mention the use of the CONSORT extension, and CONSORT-ROUTINE was not yet available for the RB-RCTs and TwiCs. Improved use of these CONSORT extensions can be considered an essential step to increase the quality of studies using alternative trial designs. A next step would be to incorporate these CONSORT extensions into a decision-making and risk assessment tool, such as has been done in the RoB-2 tool for cluster-RCTs.<sup>50</sup> In this risk of bias that has been evaluated, 7 out of 10 final RCT reports indicate some concerns, of which 5 were related to outcome measure. This was due to assessors not being blinded for the intervention, although it was acknowledged that this was unlikely to influence the results. This is comparable with regular RCT designs, and not specific for alternative RCT designs as in a review evaluating all surgical RCTs adequate generation and concealment of allocation seem to be a problem in 47% to 50% of RCTs.<sup>51</sup>

This study has several limitations. First, most included studies were published trial protocols of ongoing RCTs. In these protocols, only the a priori motivations and limitations were reported, and additional motivations and limitations (including performance) may become apparent at a later stage. A related limitation is that risk of bias could not be evaluated for the protocols. Nevertheless, the published protocols of ongoing RCTs do give an overview of the use of the trials designs, and often in the published protocols, methodology is described more elaborate giving us more information about the motivations and limitations of the designs. Second, the RoB-2 tool for includes an

**TABLE 3.** Summary of the 3 Alternative Designs

Trial Design	Settings Where Suitable	Textbook Example	Main Motivation(s)	Main Limitation(s)	Points of Consideration
SW-RCT	Strong believe that the new intervention is beneficial Implement of the intervention across all clusters is desirable	Evaluation of a national quality improvement program that implements a care pathway for emergency abdominal surgery <sup>26</sup>	Minimization of contamination High participation rates as all clusters are automatically exposed to the new intervention	Confounding of treatment effect by time Statistically complex (sample size calculation and adjustment for clusters)	Involve an experienced statistician to improve (methodological) quality
RB-RCT	Pragmatic trails requiring large numbers of patients representative of a real-world clinical population	Screening trial for AAA randomized 12,639 patients to either an abdominal ultrasound scan or the control group with outcomes collected from Danish registries <sup>29</sup>	Increased generalizability of the trial population, as inclusion criteria for registries are generally less restrictive than for standard trial	Lack of detailed data Quality of registry data	Add trial-specific variables temporarily to the registry Optimize data quality
TwICs	Dropout in standard RCTs is expected to high for subjects randomized to the control arm Multiple new treatments for the same conditions are to be evaluated (almost) simultaneously	Trial investigates how many stage II colorectal cancer patients with ctDNA will accept adjuvant chemotherapy and whether this reduces the risk of recurrence. Data are collected and patients are identified in the Prospective Dutch Colorectal Cancer Cohort (PLCRC) <sup>32</sup>	Increase of inclusion rates by avoidance of dropout due to disappointment bias More efficient when large number of potential controls are available	Patients refusing participation in the intervention arm may be high. This can dilute the outcome	Take selective refusal into account in sample size calculation to avoid loss of statistical power Requires good explanation of the design to researchers and patients

Table is based on the motivations and limitations reported in this review combined with the available literature.  
AAA indicates aortic abdominal aneurysm; ctDNA, circulating tumor DNA.

extension for cluster-RCTs, but extensions for SW-RCTs, RB-RCTs, and TwiCs are not yet incorporated. As these extensions were not taken into account in the risk of bias evaluation, specific details regarding these designs could not be assessed. This might give an underestimation of bias of the studies as, for example, the quality of the registries and cohorts are not evaluated. Third, only 3 alternative trial designs are included, but more alternative trial designs are being used, such as trials using a patient preference design.<sup>52</sup> Fourth, publication bias may have occurred because of possible limitations arisen during the trial being underreported. Fifth, the number of alternative RCTs may still be considered limited with 17 published RCTs across 3 designs. For instance, only 1 published TwiCs could be included. With more alternative RCTs becoming available the overall assessment of the specific merits could therefore still change (Supplement 9, <http://links.lww.com/SLA/E74>).

To conclude, the use of alternative trial designs within surgical research is increasing, especially over the last years. If adequately used, these innovative trial designs provide the opportunity to overcome specific difficulties associated with surgical research. However, as these designs also have their limitations, the surgeon should decide on a case-by-case basis which design is best suitable for each specific setting and use the CONSORT extensions.

## REFERENCES

- Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328:1490.
- McDonald AM, Knight RC, Campbell MK, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. *Trials*. 2006;7:1–8.
- Lindström D, Sundberg-Petersson I, Adami J, et al. Disappointment and drop-out rate after being allocated to control group in a smoking cessation trial. *Contemp Clin Trials*. 2010;31:22–26.
- Gollhofer SM, Wiskemann J, Schmidt ME, et al. Factors influencing participation in a randomized controlled resistance exercise intervention study in breast cancer patients during radiotherapy. *BMC Cancer*. 2015;15:1–9.
- Chapman SJ, Shelton B, Mahmood H, et al. Discontinuation and non-publication of surgical randomised controlled trials: observational study. *BMJ*. 2014;349:g6870.
- McCulloch P, Taylor I, Sasako M, et al. Randomised trials in surgery: problems and possible solutions. *BMJ*. 2002;324:1448–1451.
- Oberkofler CE, Hamming JF, Staiger RD, et al. Procedural surgical RCTs in daily practice: do surgeons adopt or is it just a waste of time? *Ann Surg*. 2019;270:727–734.
- Relton C, Togerson D, O’Cathain A, et al. Rethinking pragmatic randomised controlled trials: introducing the “cohort multiple randomised controlled trial” design. *BMJ*. 2010;340:c1066.
- Lauer MS, D’Agostino RB. The randomized registry trial—the next disruptive technology in clinical research? *N Engl J Med*. 2013;369:1579–1581.
- Hemming K, Taljaard M. Reflection on modern methods: when is a stepped-wedge cluster randomized trial a good study design choice? *Int J Epidemiol*. 2020;49:1043–1052.
- Hemming K, Haines TP, Chilton PJ, et al. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*. 2015;350:1–7.
- Li G, Sajobi TT, Menon BK, et al. Registry-based randomized controlled trials—what are the advantages, challenges, and areas for future research? *J Clin Epidemiol*. 2016;80:16–24.
- Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6:e1000097.
- Sterne J, Savovic J, Page M. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:14898.
- Hemming K, Taljaard M, McKenzie JE, et al. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. *BMJ*. 2018;363:k1614.
- Campbell MK, Piaggio G, Elbourne DR, et al. Consort 2010 statement: extension to cluster randomised trials. *BMJ*. 2012;345:e5661.
- Kwakkenbos L, Imran M, McCall SJ, et al. CONSORT extension for the reporting of randomised controlled trials conducted using cohorts and routinely collected data (CONSORT-ROUTINE): checklist with explanation and elaboration. *BMJ*. 2021;373:m857.
- Buhre WFFA, Boer C, Boer DdeK, et al. Routine postsurgical anesthesia visit to improve 30-day morbidity and mortality: a multicenter, stepped-wedge cluster randomized interventional study (the TRACE Study). *Ann Surg*. 2021.
- Deeken F, Sánchez A, Rapp MA, et al. Outcomes of a delirium prevention program in older persons after elective surgery: a stepped-wedge cluster randomized clinical trial. *JAMA Surg*. 2022;157:E216370.
- Gilbert T, Bernard L, Alexandre M, et al. Impact of a geriatric intervention to improve screening and management of undernutrition in older patients undergoing surgery for colorectal cancer: results of the anc stepped-wedge trial. *Nutrients*. 2021;13:2347.
- Lashoer A, Schneider EB, Juillard C, et al. Implementation of the World Health Organization Trauma Care Checklist Program in 11 centers across multiple economic strata: effect on care process measures. *World J Surg*. 2017;41:954–962.
- Masters J, Cook J, Achten J, et al. A feasibility study of standard dressings versus negative-pressure wound therapy in the treatment of adult patients having surgical incisions for hip fractures: the WHISH randomized controlled trial. *Bone Joint J*. 2021;103 B:755–761.
- Møller A, Nielsen HB, Wetterslev J, et al. Low vs high hemoglobin trigger for transfusion in vascular surgery: a randomized clinical feasibility trial. *Blood*. 2019;133:2639–2650.
- Petro CC, Zolin S, Krpata D, et al. Patient-reported outcomes of robotic vs laparoscopic ventral hernia repair with intraperitoneal mesh: The PROVE-IT Randomized Clinical Trial. *JAMA Surg*. 2021;156:22–29.
- Schwarze ML, Buffington A, Tucholka JL, et al. Effectiveness of a question prompt list intervention for older patients considering major surgery: a multisite randomized clinical trial. *JAMA Surg*. 2020;155:6–13.
- Sier MF, Oostenbroek RJ, Dijkgraaf MGW, et al. Home visits as part of a new care pathway (iAID) to improve quality of care and quality of life in ostomy patients: a cluster-randomized stepped-wedge trial. *Colorectal Dis*. 2017;19:739–749.
- Tastaldi L, Bencsath K, Alaadeen D, et al. Telescopic dissection versus balloon dissection for laparoscopic totally extraperitoneal inguinal hernia repair (TEP): a registry-based randomized controlled trial. *Hernia*. 2019;23:1105–1113.
- Wilmink TBM, Quick CRG, Hubbard CS, et al. The influence of screening on the incidence of ruptured abdominal aortic aneurysms. *J Vasc Surg*. 1999;30:203–208.
- Zatzick D, Jurkovich G, Heagerty P, et al. Stepped collaborative care targeting posttraumatic stress disorder symptoms and comorbidity for us trauma care systems: a randomized clinical trial. *JAMA Surg*. 2021;156:462–470.
- Ayorinde JOO, Summers DM, Pankhurst L, et al. PreImplantation Trial of Histopathology in renal Allografts (PITHIA): a steppedwedge cluster randomised controlled trial protocol. *BMJ Open*. 2019;9:1–10.
- Collins MG, Fahim MA, Pascoe EM, et al. Study Protocol for Better Evidence for Selecting Transplant Fluids (BEST-Fluids): a pragmatic, registry-based, multi-center, double-blind, randomized controlled trial evaluating the effect of intravenous fluid therapy with Plasma-Lyte 148 versus 0.9% saline on delayed graft function in deceased donor kidney transplantation. *Trials*. 2020;21:1–19.
- Apte SS, Moloo H, Jeong A, et al. Prospective randomised controlled trial using the REthinking Clinical Trials (REACT) platform and National Surgical Quality Improvement Program (NSQIP) to compare no preparation versus preoperative oral antibiotics alone for surgical site infection rates in elective colon surgery: a protocol. *BMJ Open*. 2020;10:e036866.
- Couwenberg AM, Burbach JPM, Berbee M, et al. Efficacy of dose-escalated chemoradiation on complete tumor response in patients with locally advanced rectal cancer (RECTAL-BOOST): a phase 2 randomized controlled trial. *Int J Radiat Oncol Biol Phys*. 2020;108:1008–1018.
- Verberne CJ, Zhan Z, Van Den Heuvel E, et al. Intensified follow-up in colorectal cancer patients using frequent carcino-embryonic antigen (CEA) measurements and CEA-triggered imaging: results of the randomized “CEAwatch” trial. *Eur J Surg Oncol*. 2015;41:1188–1196.
- Woodle ES, Gill JS, Clark S, et al. Early corticosteroid cessation vs long-term corticosteroid therapy in kidney transplant recipients: long-term outcomes of a randomized clinical trial. *JAMA Surg*. 2021;156:307–314.

36. Mathes T, Buehn S, Prengel P, et al. Registry-based randomized controlled trials merged the strength of randomized controlled trials and observational studies and give rise to more pragmatic trials. *J Clin Epidemiol*. 2018;93:120–127.
37. Richards DA, Ross S, Robens S, et al. The DiReCT study—improving recruitment into clinical trials: a mixed methods study investigating the ethical acceptability, feasibility and recruitment yield of the cohort multiple randomised controlled trials design. *Trials*. 2014;15:398.
38. Peden CJ, Stephens T, Martin G, et al. Effectiveness of a national quality improvement programme to improve survival after emergency abdominal surgery (EPOCH): a stepped-wedge cluster-randomised trial. *Lancet*. 2019;393:2213–2221.
39. Lindholt JS, Juul S, Fasting H, et al. Preliminary ten year results from a randomised single centre mass screening trial for abdominal aortic aneurysm. *Eur J Vasc Endovasc Surg*. 2006;32:608–614.
40. Schraa SJ, Van Rooijen KL, Van Der Kruijssen DEW, et al. Circulating tumor DNA guided adjuvant chemotherapy in stage II colon cancer (MEDOCC-CrEATE): study protocol for a trial within a cohort study. *BMC Cancer*. 2020;20:1–10.
41. Campbell MJ, Hemming K, Taljaard M. The stepped wedge cluster randomised trial: what it is and when it should be used. *Med J Aust*. 2019;210:253.e1–254.e1.
42. Barker D, McElduff P, D'Este C, et al. Stepped wedge cluster randomised trials: a review of the statistical methodology used and available. *BMC Med Res Methodol*. 2016;16:69.
43. Martin J, Taljaard M, Girling A, et al. Systematic review finds major deficiencies in sample size methodology and reporting for stepped-wedge cluster randomised trials. *BMJ Open*. 2016;6:e010166.
44. Zolin SJ, Petro CC, Prabhu AS, et al. Registry-based randomized controlled trials: a new paradigm for surgical research. *J Surg Res*. 2020;255:428–435.
45. Verkooijen HML, Roes K, van Gils CH. Cohort multiple randomized controlled trial: a solution for the evaluation of multiple interventions. *Ned Tijdschr Geneesk*. 2013;157:A5762.
46. Gal R, Monninkhof EM, van Gils CH, et al. The Trials within Cohorts design faced methodological advantages and disadvantages in the exercise oncology setting. *J Clin Epidemiol*. 2019;113:137–146.
47. Pate A, Candlish J, Sperrin M, et al. Cohort multiple randomised controlled trials (cmRCT) design: efficient but biased? A simulation study to evaluate the feasibility of the Cluster cmRCT design. *BMC Med Res Methodol*. 2016;16:109.
48. Reeves D, Howells K, Sidaway M, et al. The cohort multiple randomized controlled trial design was found to be highly susceptible to low statistical power and internal validity biases. *J Clin Epidemiol*. 2018;95:111–119.
49. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332.
50. Lyu XZ, Sun F, Zhan SY. Risk related to bias assessment: (4) revised Cochrane risk of bias tool for cluster-randomized control trials (RoB2.0). *Zhonghua Liu Xing Bing Xue Za Zhi*. 2018;39:240–244.
51. 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.
52. Wasmann KA, Wijsman P, van Dieren S, et al. Partially randomised patient preference trials as an alternative design to randomised controlled trials: systematic review and meta-analyses. *BMJ Open*. 2019;9:e031151.

## DISCUSSANT

**Inne H.M. Borel Rinke (Utrecht, The Netherlands)**

I would like to thank the ESA for the privilege of discussing this very interesting paper. Dr. Augustinus and her colleagues are to be complimented for this overview of new randomized control trial designs. These are much desired because of the problems encountered when performing RCTs, particularly in the field of surgery. The authors described the three alternatives elegantly.

My first question regards the adherence to the trial results. We published a paper on this topic, which was presented in Madrid two years ago, and less than half of the responding

surgeons admitted that they would comply with these results; more than half would not. Do you think that introducing these new trials will encourage surgeons to adopt their results, or change their behaviors?

Second, you've based your review on trial descriptions in more than half of your RCTs (23 out of 41; only 18 involved true RCT results). Perhaps, this is because they are new trial designs, but is that sufficient to base your conclusions on?

Similarly, you showed one motivation and one limitation per trial design, based on what the authors spontaneously reported. However, despite them being good motivations and limitations, there must be more. Perhaps, it would have been better to interview these authors directly, in order to really understand whether there were other reasons for their motivations.

Third, although the trial designs do appear to speed up or facilitate trial completion, they also each have their own specific problems. You have alluded to some concerns regarding a lack of quality criteria. So, the question remains: will the novel designs ultimately provide quality levels like classic RCTs? What should we do to achieve this?

Finally, the step-wedge trials have all focused on non-surgical interventions. Would surgical interventions be suitable for these trial designs as well?

## Response from Simone Augustinus (Amsterdam, The Netherlands)

Thank you very much for the compliments and questions. To answer your first question, which is perhaps the most important one. I think that the step-wedge trial design can best solve this. In this design, which could also involve low-complex surgical interventions, the intervention is often implemented in all clusters (hospitals and patients) during the trial. Hereafter, if the intervention is proven to be effective, everyone will usually continue working with these interventions. In particular, registry-based trials improve the generalizability of trial results. In the survey you mention, the most commonly reported reason for not adhering was the use of a highly selective population. By incorporating trials into registries, the results can be more generalized and that might overcome the issue that people have when implementing them within a clinical practice. This is the same for Trials Within Cohorts (TWCs), where patients are asked to sign a consent form once they are informed of the intervention that they may become eligible for. So, you can really see whether patients want the intervention in clinical practice, as it is only offered to patients who subsequently randomize for the intervention arm. These are some of the ways that the different trial designs can improve adherence.

Regarding the protocols, clearly not everything is incorporated, but generally, the protocols incorporate extensive information on the motivations and limitations. However, I think that now is the right time to present these results, as we see that these trial designs are increasingly used. Therefore, we do need to pay attention to what the observed limitations are and whether we should perform more research to adequately interpret these results.

I think this also answers your other question regarding the next steps to make sure we have all the motivations and limitations. We could have an expert survey, including surgeons and statisticians with expertise on this topic, to bring additional perspectives.