Journal of Critical Care 41 (2017) 107–111

Contents lists available at ScienceDirect

Journal of Critical Care

journal homepage: www.jccjournal.org

Interventions affecting mortality in critically ill and perioperative patients: A systematic review of contemporary trials

Martina Baiardo Redaelli, MD [a], Giovanni Landoni, MD, Professor [a,b,*], Stefania Di Sanzo, MSc [c], Samuele Frassoni, MSc [a], Chiara Sartini, MD [a], Luca Cabrini, MD [a], Giacomo Monti, MD [a], Mara Scandroglio, MD [a], Alberto Zangrillo, MD, Full Professor [a,b], Rinaldo Bellomo, MD, Full Professor [d]

[a] Department of Anesthesia and Intensive Care, IRCCS San Raffaele Scientific Institute, Milan, Italy
[b] Vita-Salute San Raffaele University, Milan, Italy
[c] Department of Cardiac Surgery, IRCCS San Raffaele Scientific Institute, Milan, Italy
[d] Faculty of Medicine, University of Melbourne, Melbourne, Australia

ARTICLE INFO

Keywords:
RCT
Randomized clinical trials
Trials
Mortality
Critically ill
Perioperative
Review
Anesthesia
Intensive care

ABSTRACT

Purpose: Confounders in randomized controlled trials (RCTs) reporting significant effects on mortality in critically ill patients using non-surgical techniques have not been systematically explored. We aimed to identify factors unrelated to the reported intervention that might have affected the findings and robustness of such trials.
Methods: We searched Pubmed/MEDLINE for all RCTs on any non-surgical interventions reporting an effect on unadjusted mortality in critically ill patients between 1/1/2000 and 1/12/2015. We assessed: the number needed to treat/harm (NNT or NNH), sample size, trial design (blinded/unblinded, single or multinational, single or multicenter (sRCT or mRCT)), intention to treat (ITT) analysis, and countries of origin.
Results: Almost half of RCTs were sRCTs. Median sample size was small, and 1/3 were not analyzed according to ITT principle. Lack of ITT analysis was associated with greater effect size ($p = 0.0028$). Harm was more likely in mRCTs ($p = 0.002$) and/or in blinded RCTs ($p = 0.003$). Blinded RCTs had double sample size ($p = 0.007$) and an increased NNT/NNH ($p = 0.002$). Finally, mRCTs had higher NNT ($p = 0.005$) and NNH ($p = 0.02$), and harm was only detected in studies from Western countries ($p = 0.007$).
Conclusions: These observations imply that major systematic biases exist and affect trial findings irrespective of the intervention being studied.

1. Introduction

Decreasing mortality in critically ill and postoperative patients is a public health goal. Thus, the primary outcome measure of multiple interventional trials [1]. Such patients are at high risk of death [2-6] and represent one of the main areas of health care expenditure in the western world [7]. Accordingly, any study reporting the effect of an intervention on mortality (either an increase or a decrease) has the potential to significantly change clinical practice worldwide, save thousands of lives, and reduce health-care costs [8].

According to Evidence Base Medicine (EBM) principles, randomized controlled trials (RCT) represent the most robust source of evidence to guide practice [9]. However, in the field of critical care and postoperative medicine, no assessment has been made of what confounding factors may affect the findings of such RCT beyond the intervention itself and whether any systematic biases exist, which may affect trial findings.

Accordingly, we systematically identified all contemporary RCTs of non-surgical intervention in critical care and postoperative medicine (all studies published since 2000) and reported in peer reviewed journals, which showed a statistically significant impact on mortality. The aim of our study was to identify whether there were confounding factors unrelated to the interventions, which might have systematically affected trial findings.

2. Methods

2.1. Systematic search and article selection

PubMed/MEDLINE were searched for all RCTs of any non-surgical intervention influencing unadjusted landmark mortality in critically ill and postoperative patients published between January 1st, 2000 and

* Corresponding author at: Cardiothoracic Anesthesia and Intensive Care, Istituto Scientifico San Raffaele, via Olgettina 60, Milano 20132, Italy.
E-mail addresses: baiardoredaelli.martina@hsr.it (M. Baiardo Redaelli), landoni.giovanni@hsr.it (G. Landoni), s.frassoni1@campus.unimib.it (S. Frassoni), cabrini.luca@hsr.it (L. Cabrini), monti.giacomo@hsr.it (G. Monti), scandroglio.mara@hsr.it (M. Scandroglio), zangrillo.alberto@hsr.it (A. Zangrillo), Rinaldo.BELLOMO@austin.org.au (R. Bellomo).

http://dx.doi.org/10.1016/j.jcrc.2017.05.005
0883-9441/

December 1st 2015 (see full PubMed search strategies in Supplementary Appendix). Additional articles were suggested by experts and obtained from a cross-check of references from primary papers.

Articles were then selected for further assessment if they met all the following criteria:

1) Publication in a peer-reviewed English-language journal;
2) Single-center or multicenter trial design;
3) Randomized controlled trial design;
4) Statistically significant reduction or increase in unadjusted landmark mortality;
5) Postoperative or critical care setting;
6) Publication date between January 2000 and November 2015.

Articles were excluded if they fulfilled at least one of these criteria: 1) Used a quasi-randomized or non-randomized methodology; 2) Dealt only with a pediatric population (<18 years); 3) Did not report mortality data.

We considered patients to be critically ill if, at randomization, they had at least one organ failure and/or were receiving intensive care treatment and/or emergency treatment, regardless of where they were treated (intensive care unit, emergency department, or general ward). Assessment of the eligibility of the identified studies was performed by two authors. Differences of opinion were discussed among authors until consensus was reached.

2.2. Data collection

For each RCT, we extracted details of the paper (title, first author, journal name, year of publication, impact factor of the journal in the year of publication), details of the RCT design (the intervention and its comparator, trial setting, blinding, intention to treat analysis, whether enrollment was interrupted after interim analysis, number of patients randomized to each group, number of patients who experienced an outcome in each group), details of the significant mortality outcome (follow-up time, whether mortality was the primary study outcome, type of statistical test used to assess the difference in mortality, and p-value reported).

2.3. Data analysis

We assessed and recorded the size effect of the intervention, the absolute risk reduction or increase, and the number needed to treat or number needed to harm (NNT/NNH) [10].

For RCTs reporting a significant difference in mortality at more than one landmark time, we chose the longest follow-up time. For trials with more than one comparator treatment and where intervention affected mortality compared to more than one control group, we chose the comparison with the smallest p-value. We analyzed the differences in the NNT/NNH, sample size, number of multicenter randomized clinical trials (mRCTs) and single-center RCTs (sRCTs), number of blinded studies, median impact factor, median p-value of the studies, median number of centers and median number of nations, according to impact on mortality (harmful and beneficial studies), trial design (blinded versus unblinded, single nation versus multinational, and sRCTs versus mRCTs), countries (European and non-European, USA and non-USA, non-European, non-USA, non-Australia and New Zealand (ANZ), non-Canada vs. other countries), assessment of mortality as primary or secondary outcome, conflict of interest (none declared versus declared and not declared), setting (intensive care unit (ICU) and non-ICU); intervention type such non-invasive ventilation (NIV) versus all the other interventions.

2.4. Statistical analysis

The Dataset was created using Microsoft Excel 2010 for Windows (Microsoft Corporation, Redmond, WA, USA) and analyzed with the use of Stata software, version 13 (StataCorp). Continuous variables are reported as medians and interquartile range (IQR) and categorical variables as counts and percentage. Comparisons of dichotomous data were performed by Chi-squared test or Fisher's exact test as appropriate. Continuous measurements were compared with the use of the Wilcoxon – Mann Whitney test where appropriate. To adjust for multiple comparisons, a p-value < 0.01 was considered statistically significant.

3. Results

3.1. General study characteristics

The five search strategies initially returned >60 thousand RCTs. After excluding overlaps, our search identified 56,554 potential manuscripts published between January 2000 and December 2015. Of these, 139RCTs met the inclusion criteria (Fig. 1). The references and the PubMed links for all 139 abstracts are available in Supplementary Table 1.

Of the 139 papers identified, 119 (85.6%) reported interventions that decreased mortality, and 20 (14.4%) reported interventions that increased mortality. In addition, 73 studies (52.5%) were mRCTs, while 66 (47.5%) were sRCTs (Table 1). The country of origin for multinational studies was attributed to the affiliation of the corresponding author (25 RCTs 18.0%), but the majority of studies were single-nation in design (114 RCTs, 82.0%). Out of the 31 countries of origin, the three nations with most frequently published RCTs affecting mortality were the USA (eight sRCTs and 15 mRCTs) Spain (12 sRCTs and 11 mRCTs), and France (two sRCTs and nine mRCTs) (Supplementary Table 2).

3.2. Study size and analysis

Overall, the median sample size was 160 patients [79–341], and the overall number of centers involved was 3451 with a median value of 2 [1–10] centers per study. However, when excluding sRCTs, the median number of the centers involved was 9 [3–31].

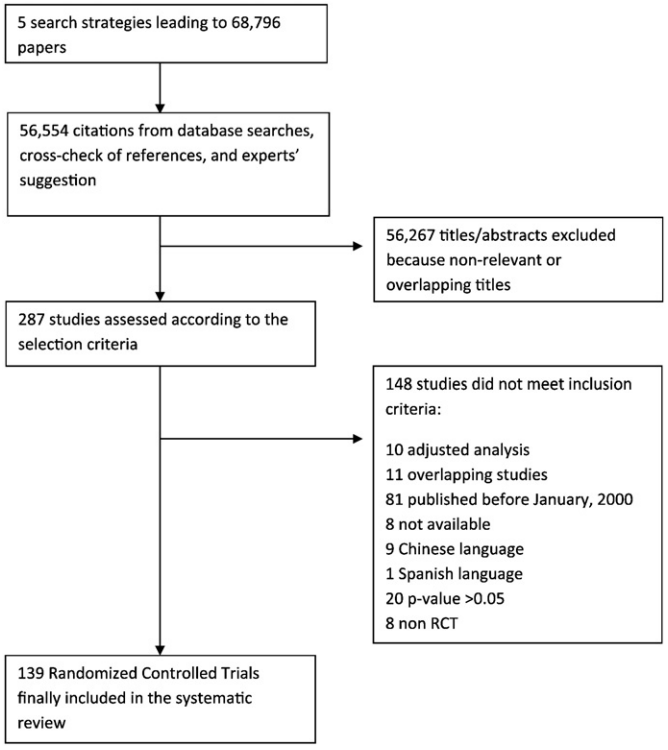

Fig. 1. Flowchart of the systematic review article selection process.

Table 1
Characteristics of single center RCT and multi center RCT.

	Single center RCT	Multi center RCT	p-Value
N of papers (total 139)	66	73	
Studies showing harm, n (%)	3 (5%)	17 (23%)	0.002
N of patients, median [IQR]	97 [59–210]	221 [106–548]	<0.001
p-Value, median [IQR]	0.023 [0.01–0.037]	0.025 [0.014–0.03]	0.4
NNT, median [IQR]	5.0 [3.3–7.1]	6.4 [4.4–10.9]	0.005
NNH, median [IQR]	3.7 [3.3–4.3]	9.5 [7.2–20.2]	0.017
N of centers, median [IQR]	1 [1–1]	3 [9–31]	<0.001
IF, median [IQR]	5 [2.8–12.6]	9.1 [4.2–30.4]	0.004
Blinded trials, n (%)	28 (42%)	38 (52%)	0.2

Abbreviations: N: number; IQR: interquartile range; NNT: number needed to treat; NNH: number needed to harm; IF: impact factor.

Overall, 43 (30.9%) studies did not analyze data according to the ITT principle. Compared to studies that did not perform ITT analysis, the median NNT was statistically significantly higher in studies applying the ITT principle studies (6 [4.5–10.5] vs. 4 [3–7]; p = 0.0028).

3.3. Harmful versus beneficial and blinded versus unblended interventions

Evidence for harmful interventions was mostly observed in mRCTs (85%), while evidence for beneficial interventions was mostly observed in sRCTs (53%; p = 0.002). Moreover, RCTs identifying harmful intervention had a significantly higher median sample size (p = 0.002); involved a greater number of nations and centers (p < 0.0001); and were significantly more likely to have blinding as part of their study design (Table 2). Finally, when comparing blinded and non-blinded RCTs, the NNT/NNH were significantly higher in the blinded studies (7 [5–13] vs 5 [4–7]; p = 0.006) as was median sample size (224 [93–497] vs. 114 [61–216]; p = 0.007) (Table 3).

3.4. Country and culture of origin and impact factor

USA studies were mostly blinded while European studies were published in journals with higher median impact factor (IF; Supplementary Table 3.a).Comparing western culture countries (Europe, Australia, New Zealand, USA, and Canada) to all other countries, we found that mRCTs were more frequent (64% vs. 28%; p = 0.002), median sample size was greater (190 patients [91–485] vs 86.5 patients [IQR 50–121]; p < 0.001), and median number of centers greater (3 [1–16] vs 1 [1–2]; p = 0.002) in such countries (Supplementary Tables 3.b). Furthermore, all interventions shown to be harmful were reported in studies conducted in western RCTs, and no harmful interventions were reported in studies by other countries. Finally, the median NNT/NNH was higher in studies from western countries (6.2 [4.1–10.5] vs 4.3 [3.1–6.2]; p = 0.003) and these studies were published in journals with a higher median IF (9.2 [4.2–30.4] vs. 3 [1.5–6.1]; p < 0.001) (Table 4).

Table 2
Characteristics of studies describing interventions increasing mortality (harmful), or increasing survival (beneficial).

	Harm	Benefit	p-Value
N of papers (total 139)	20	119	
mRCT, n (%)	17 (85%)	56 (47%)	0.002
sRCT, n (%)	3 (15%)	63 (53%)	
N of patients, median [IQR]	338 [197.5–1011]	127 [71–275]	0.002
p-Value, median [IQR]	0.022 [0.02–0.03]	0.025 [0.01–0.04]	0.09
NNT/NNH, median [IQR]	9 [5–18]	6 [4–9]	0.021
N of nations median, [IQR]	2 [1–4.5]	1 [1–1]	<0.001
N of centers, median [IQR]	21 [5–42]	1 [1–6]	<0.001
IF, median [IQR]	36.94 [4.3–51.5]	6.3 [3.1–15.4]	0.001
Blinded trials, n (%)	16 (80%)	50 (42%)	0.003

Abbreviations: sRCT: single-center randomized controlled trial; mRCT: multicenter randomized controlled trial; N: number; IQR: interquartile range; NNT: number needed to treat; NNH: number needed to harm; IF: impact factor.

Table 3
Characteristics of RCTs according to blinded or unblinded trial design.

	Unblinded	Blinded	p-Value
N of papers (total 139)	73	66	
Studies showing harm, n (%)	4 (5.48%)	16 (24.24%)	0.003
mRCT, n (%)	35 (47.95%)	38 (57.57%)	0.2
sRCT, n (%)	38 (33.33%)	28 (42.42%)	
N of patients, median [IQR]	114 [61–216]	224 [93–497]	0.007
p-Value, median [IQR]	0.024 [0.01–0.03]	0.025 [0.01–0.04]	0.4
NNT, median [IQR]	5 [4–7]	7 [5–13]	0.001
NNH, median [IQR]	9 [7–24]	9 [5–18]	0.7
N of nations median, [IQR]	1 [1–1]	1 [1–1]	0.053
N of centers, median [IQR]	1 [1–6]	2 [1–19]	0.057
IF, median [IQR]	6.124 [2.9–15.4]	7.35 [4.14–30.39]	0.08

Abbreviations: sRCT: single-center randomized controlled trial; mRCT: multicenter randomized controlled trial; N: number; IQR: interquartile range; NNT: number needed to treat; NNH: number needed to heel; IF: impact factor.

3.5. Type of intervention

More than 50 different non-surgical interventions were identified. The most studied intervention was non-invasive mechanical ventilation (NIV), with 18 RCTs (Supplemental Refs. 12, 22, 35, 39, 44, 66, 68, 71, 76, 82, 86, 99, 106, 109, 113, 114, 118, 120) showing benefit and one showing harm. Other interventions supported by numerous RCTs were antibiotic therapy (13 RCTs) (Supplemental Refs. 18, 43, 45, 69, 81, 83, 89–91, 100, 119, 122, 123), immune-modulating nutrients (11 RCTs) (Supplemental Refs. 5, 19, 21, 33, 48, 58, 73, 87, 93, 107, 137), and renal replacement therapy (nine RCTs) (Supplemental Refs. 46, 60, 67, 80, 88, 92, 94, 101, 129). In contrast, 38 interventions were assessed by only one or two RCTs. The full list of interventions is reported in the supplementary material (Supplementary Table 4).

When comparing the 18 NIV studies to the other studies, mean sample size appeared smaller (p = 0.02) and all but one study lacked blinding (Supplementary Table 5).

3.6. Journals

The New England Journal of Medicine (NEJM), Critical Care Medicine, The Lancet, and The Journal of the American Medical Association (JAMA) published most of these studies with 24, 15, 13, and 10 papers, respectively (Supplementary Table 6).Of the three journals with the highest IF (NEJM, The Lancet, JAMA),the NEJM published the highest percentage of studies showing harm (37.5%), while JAMA published more sRCTs (40%) and the Lancet reported studies with the highest median NNT (7 [45–21.12]) (p ≤ 0.03 for all comparisons) (Supplementary Table 7).

Table 4
Characteristics of studies dividing papers according to geographic origin (Europe, USA, ANZ, Canada vs. all the other countries).

	Other countries	Europe USA ANZ Canada	p-Value
N of papers (total 139)	32	107	
mRCT, n (%)	9 (28%)	64 (60%)	0.002
Studies showing harm, n (%)	0 (0%)	20 (19%)	0.007
N of patients, median [IQR]	86.5 [50–121]	190 [91–485]	<0.001
p-Value, median [IQR]	0.025 [0.01–0.04]	0.023 [0.01–0.03]	0.9
NNT, median [IQR]	4.3 [3.1–6.2]	6.2 [4.1–10.5]	0.003
NNH, median [IQR]	Non-available	9.1 [5.1–18.2]	Non-available
N of centers, median [IQR]	1 [1–2]	3 [1–16]	<0.001
IF, median [IQR]	3 [1.5–6.1]	9.2 [4.2–30.4]	<0.001
Blinded trials, n (%)	11(34%)	55 (51%)	0.09

Abbreviations: sRCT: single-center randomized controlled trial; mRCT: multicenter randomized controlled trial; N: number; IQR: interquartile range; NNT: number needed to treat; NNH: number needed to heel; IF: impact factor.

3.7. Secular trends

The median number of papers published each year during these 15 years remained stable (Supplementary Fig. 1) at 9 [5.75–11] per year. The median IF for the journals that published these RCTs was 6.6 [3.15–28.90] and this trend also remained stable (Supplementary Fig. 2). Out of the 139 critical care studies, 101 studies were performed in the ICU (72.7%), and 32 (23.0%) in the perioperative setting. Of these, 12 were in cardiac surgery patients. No statistically significant difference was observed in NNT/NNH or trial design, or sample size, when comparing ICU and non-ICU studies (Supplementary Table 8).

3.8. Multinational vs. single nation studies

Both the sample size and the NNT were statistically significantly higher ($p < 0.001$) in the 25 (18%) multinational RCTs (606 [310–2314]; median NNT 13 [9–28]), when compared to the 114 (82%) single-nation studies (median sample size 118.5 [64–224] median NNT 5 [4–8]). Furthermore, multinational RCTs were published in journals with much higher median IF (29.1 versus 6.2, $p < 0.001$). Finally, 48% of the multinational RCTs found harm compared with only 7.0% of single-nation RCT finding harm ($p = 0.001$) (Table 5).

3.9. Combined effects and additional trial characteristics

Trials with both multicenter and blinded design randomized three times as many patients, and required twice the NNT to detect an effect (Table 6).

Mortality was the primary endpoint of the studies for 67 RCTs. However, studies with mortality as primary outcome and studies with mortality as a secondary outcome carried not significant differences ($p = 0.18$) (Supplementary Table 9). Length of follow-up varied greatly (Supplementary Table 10.a and b) as did tests used to evaluate the p-value for mortality. However, no biases were detected in these two aspects. The overall median NNT/NNH for mortality was 6 [4–10], and a detailed distribution of this value is available in Supplementary Fig. 1.a and b. The overall median p-value for mortality significance was 0.025 [0.01–0.03]). The p-value distribution, divided by intervals of 0.005 is reported in Supplementary Plot 2.

Conflicts of interest are described in Supplementary Table 11. No statistically significant difference was noted between studies with no conflict of interest, compared to studies where conflict was declared or not specified.

Overall, 26 studies were interrupted early prior to completion of the full study recruitment. No statistically significant difference was noted in the NNT for the studies interrupted after interim analysis, when compared to the ones which completed recruitment (median NNT 5 [3–10] vs. median NNT 6 [4–10], respectively; $p = 0.6$). Finally, no differences were detected according to statistical tests used to detect significance (Supplementary Table 12.a and b).

Table 5
Characteristics of single-nation and multinational studies.

	Single nation	Multinational	p-Value
Number of papers (total 139)	114	25	
Studies showing harm, n (%)	8 (7%)	12 (48%)	<0.001
mRCT, n (%)	48 (42.1%)	25 (100%)	<0.001
sRCT, n (%)	66 (57.9%)	0	
Number of patients, median [IQR]	118.5 [64–224]	606 [310–2314]	<0.001
p-Value, median [IQR]	0.02 [0.01–0.04]	0.03 [0.01–0.3]	0.8
NNT, median [IQR]	5.1 [3.7–7.7]	14.3 [9.2–28.8]	<0.001
NNH, median [IQR]	4.6 [3.5–14.7]	9.6 [7.9–18.9]	0.07
Number of centers, median [IQR]	1 [1–4]	27 [11–124]	<0.001
IF, median [IQR]	6.2 [3.2–12.6]	29.1 [11.3–51.7]	<0.001
Blinded trials, n (%)	50 (43.9%)	16 (64%)	0.06

Abbreviations: sRCT: single-center randomized controlled trial; mRCT: multicenter randomized controlled trial; N: number; IQR: interquartile range; NNT: number needed to treat; NNH: number needed to heel; IF: impact factor.

Table 6
Comparison of unblinded single center RCTs vs. blinded multi center RCT.

	sRCT unblinded	mRCT blinded	p-Value
N of papers (total 76)	38 (50%)	38 (50%)	
N of patients, median [IQR]	95.5 [52–206]	314.5 [162–1218]	<0.001
p-Value, median [IQR]	0.0232 [0.007–0.035]	0.0275 [0.014–0.0317]	0.3
NNT, median [IQR]	4.3 [3.2–6.4]	8.4 [5.7–14.8]	<0.001
NNH, median [IQR]	ND	9.6 [7.2–20.2]	ND
N of nations median, [IQR]	1 [1–1]	1 [1–4]	<0.001
N of centers, median [IQR]	1 [1–1]	14 [5–49]	<0.001
IF, median [IQR]	4.5 [2.4–15.4]	12.6 [6.3–33.6]	0.001

Abbreviations: sRCT: single-center randomized controlled trial; mRCT: multicenter randomized controlled trial; N: number; IQR: interquartile range; NNT: number needed to treat; NNH: number needed to heel; IF: impact factor.

4. Discussion

4.1. Key findings

We performed a systematic analysis of all contemporary trials reporting interventions that significantly affected mortality in critically ill or postoperative patients. We found that almost half of such trials were single center in design, that median sample size was small, and that one third were not analyzed according to the ITT principle, a strategy that inflated effect size by a third. Moreover, we found that mRCTs were more likely to show harm; that studies showing harm were twice as likely to be mRCTs; randomized many more patients; were conducted in more countries; involved 20-times more centers; were twice as likely to be blinded, and increased the NNT three-fold. Coherent with such findings, blinded RCTs were four times more likely to show harm; had double the sample size, and markedly increased the NNT to detect an effect, In contrast, no unblinded sRCT reported harm. Finally, we found that harm was only detected in studies from western culture countries and that studies from such countries randomized twice as many patients in three times as many centers; were more likely to be mRCTs, and reported a significantly greater NNT.

4.2. Relationship to previous studies

Our findings identify several systematic biases, which appear to affect trial findings irrespective of the interventions applied [11]. In particular, analysis according to the ITT principle; multicenter design; blinding of intervention; and geographic location of the trial, all appear to affect trial findings [12-14]. These findings are consistent with previous reports that neglect of the ITT principle leads to significant bias toward a greater likelihood of a beneficial treatment effect [14,15]. The findings that, compared to mRCTs, sRCTs are significantly less likely to detect harm is consistent with previous studies [16]. The germane observation that unblinded studies are also less likely to detect harm is also consistent with previous studies in other patient groups [13,17,18]. In our population, we provide an effect estimate of the impact of blinding at approximately one third, meaning that the NNT/NNH increases by this proportion when blinding is applied. Similarly, mRCTs and/or applying the IIT principle increase the NNT/NNH by about a third. We are the first to report a geographical and/or cultural effect, with no studies outside of western countries showing harm in trials involving critically ill or postoperative patients.

4.3. Implications of study findings

Our observations imply that systematic biases exist in contemporary RCTs of non-surgical interventions applied to critically ill or post-

operative patients, which affect trial findings irrespective of the intervention being studied. Specifically, they imply that, compared with blinded mRCTs which used the ITT principle, unblinded sRCTs and avoidance of the ITT principle all markedly inflate treatment effect. Moreover, they imply that such trials are also much less likely to detect harm. Finally, they imply that even the geographical location of a trial may lead to a systematic bias that makes identification of harm less likely.

4.4. Strengths and limitations

Our study has several strengths. It is the first systematic and detailed assessment of contemporary trials reporting a mortality effect with non-surgical interventions in critically ill or post-operative patients. It assessed such trials for a variety of characteristics and potential sources of bias. By assessing more than a hundred such trials, it was able to detect such biases and provide an estimate of their effect. Its findings are consistent with logical expectations, previous literature, and have robust levels of statistical significance. Finally, the implications for policy makers, clinicians and expert groups charged with the development of guidelines or consensus statements are important.

Our study also carries some limitations. First it was confined to a specific group of patients. However, these unique patients are those with a high risk of short-term mortality, their cost of care is high and yet no such assessment of evidence quality had been previously conducted. Second, combining critically ill patients and post-operative patients into one group is open to criticism. However, these patients are typically taken care of in post-operative high dependency or intensive care units by a group of clinicians whose therapeutic decisions are affected by trials in this sphere. Third, the biases we identify have been described in other fields of trial medicine and have been previously reported, but not in this particular group of patients and not in such detail. By doing so, we highlight the limitations of sRCTs and unblinded RCTs and of RCT outside of western countries, we provide estimates of effect inflation and, finally, raise concern about the validity of using unblindeds RCTs to generate guidelines [9,19,20].

5. Conclusions

We performed a systematic analysis of all contemporary trials reporting interventions that significantly affected mortality in critically ill or perioperative patients. We found that one-third of trials were not analyzed according to the ITT principle, a strategy that inflated effect size by a third; that harm was more likely to be found in mRCTs and/or blinded RCTs; that blinded or mRCTs increased the NNT/NNH by more than a third, and that harm was only detected in studies from western countries. These observations imply that major systematic biases exist in contemporary RCTs of non-surgical interventions applied to critically ill or post-operative patients, which affect trial findings irrespective of the intervention being studied. These findings have important implications for policy makers, clinicians, and expert groups charged with the development of guidelines or consensus statements.

Funding

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

Conflicts of interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jcrc.2017.05.005.

References

[1] Ospina-Tascón GA, Büchele GL, Vincent JL. Multicenter, randomized, controlled trials evaluating mortality in intensive care: doomed to fail? Crit Care Med 2008;36:1311–22.
[2] Afessa B, Keegan MT, Hubmaye RD, et al. Evaluating the performance of an institution using an intensive care unit benchmark. Mayo Clin Proc 2005.
[3] Christensen S, Johansen MB, Christiansen CF, et al. Comparison of Charlson comorbidity index with SAPS and APACHE scores. Clin Epidemiol 2011;3:203–11.
[4] Heyland DK, Konopad E, Noseworthy TW, et al. Is it "worthwhile" to continue treating patients with a prolonged stay (>14 days) in the ICU? An economic evaluation. Chest 1998;114:192–8.
[5] Eddleston JM, White P, Gurthrie E. Survival, morbidity and quality of life after discharge from intensive care. Crit Care Med 2000;28(7):2293–9.
[6] Mayr VD, Dünser MW, Greil V, et al. Causes of death and determinants of outcome in critically ill patients. Crit Care 2006;10:R154.
[7] Halpern NA, Pastores SM. Critical care medicine in the United States, 2000–2005: an analysis of bed numbers, occupancy rates, ayer mix, and costs. Crit Care Med 2010;38:65–71.
[8] Landoni G, Comis M, Conte M, et al. Mortality in multicenter critical care trials: an analysis of interventions with a significant effect. Crit Care Med 2015;43:1559–68.
[9] Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 1992;268:2420–5.
[10] Bender R. Calculating confidence intervals for the number needed to treat. Control Clin Trials 2001;22:102–10.
[11] Ridgeon EE, Young PJ, Bellomo R, et al. The fragility index in multicenter randomized controlled critical care trials. Crit Care Med 2016;44:1278–84.
[12] Yusuf S, Wittes J. Interpreting geographic variation in results of randomized, controlled trials. N Engl J Med 2016;375:2263–71.
[13] Psaty BM, Prentice RL. Minimizing bias in randomized trials: the importance of blinding. JAMA 2010;304:793–4.
[14] Aberegg SK, Richards DR, O'Brien JM. Delta inflation: a bias in the design of randomized controlled trials in critical care medicine. Crit Care 2010;14:R77.
[15] Abraha I, Cherubini A, Cozzolino F, et al. Deviation from intention to treat analysis in randomised trials and treatment effect estimates: meta-epidemiological study. BMJ 2015;350:h2445.
[16] Bellomo R, WarrillowSJ Reade MC. Why we should be wary of single-center trials. Crit Care Med 2009;12:3114–9.
[17] Day SJ, Altman DG. Blinding in clinical trials and other studies. Statistics notes. BMJ 2000;321:504.
[18] Heinrich JG. Benefits of blind analysis techniques. CDF/MEMO/STATISTICS/PUBLIC/6576 version 1. University of Pennsylvania; 2003 1–14.
[19] Rivers E, Nguyen B, Havstad S, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med 2001;345:1368–77.
[20] Van den Berghe G, Wilmer A, Hermans G, et al. Intensive insulin therapy in the medical ICU. N Engl J Med 2006;354:449–61.