Mortality in Multicenter Critical Care Trials: An Analysis of Interventions With a Significant Effect*

Giovanni Landoni, MD[1]; Marco Comis, MD[2]; Massimiliano Conte, MD[3]; Gabriele Finco, MD[4]; Marta Mucchetti, MD[1]; Gianluca Paternoster, MD, PhD[5]; Antonio Pisano, MD[6]; Laura Ruggeri, MD[1]; Gabriele Alvaro, MD[7]; Manuela Angelone, MD[6]; Pier C. Bergonzi, MD[1]; Speranza Bocchino, MD[1]; Giovanni Borghi, MD[1]; Tiziana Bove, MD[1]; Giuseppe Buscaglia, MD[8]; Luca Cabrini, MD[1]; Lino Callegher, MD[9]; Fabio Caramelli, MD[10]; Sergio Colombo, MD[1]; Laura Corno, MD[1]; Paolo Del Sarto, MD[11]; Paolo Feltracco, MD[12]; Alessandro Forti, MD[13]; Marco Ganzaroli, MD[2]; Massimiliano Greco, MD[1]; Fabio Guarracino, MD[14]; Rosalba Lembo, MSc[1]; Rosetta Lobreglio, MD[15]; Roberta Meroni, MD[1]; Fabrizio Monaco, MD[1]; Mario Musu, MD[16]; Giovanni Pala, MD[17]; Laura Pasin, MD[1]; Marina Pieri, MD[1]; Stefania Pisarra, MD[18]; Giuseppe Ponticelli, MD[6]; Agostino Roasio, MD[19]; Francesco Santini, MD[20]; Simona Silvetti, MD[1]; Andrea Székely, MD[21]; Massimo Zambon, MD[1]; Maria Chiara Zucchetti, MD[22]; Alberto Zangrillo, MD[1]; Rinaldo Bellomo, MD[23]

*See also p. 1767.

[1]Department of Anesthesia and Intensive Care, IRCCS San Raffaele Scientific Institute, Milan, Italy.

[2]Cardiac and Vascular Department, Mauriziano Hospital, Turin, Italy.

[3]Department of Anesthesia and Intensive Care, Maria Cecilia Hospital - GVM Care & Research, Cotignola (RA), Italy.

[4]Department of Medical Sciences “M. Aresu,” University of Cagliari, Cagliari, Italy.

[5]Cardiovascular Anesthesia and Intensive Care, San Carlo Hospital, Potenza, Italy.

[6]Division of Cardiac Anesthesia and Intensive Care, Azienda Ospedaliera Dei Colli, V Monaldi, Naples, Italy.

[7]A.O. Mater Domini Germaneto, Catanzaro, Italy.

[8]Cardioanesthesia and Intensive Care, IRCCS University Hospital San Martino Ist, Genova, Italy.

[9]Department of Anesthesia and Intensive Care, S. Maria dei Battuti Hospital ULSS 9, Treviso, Italy.

[10]Cardiothoracic and Vascular Anesthesia and Intensive Care, S. Orsola-Malpighi University Hospital, Bologna, Italy.

[11]FTGM—“G. Pasquinucci” Heart Hospital, Massa, Italy.

[12]Department of Pharmacology and Anesthesiology, University Hospital of Padova, Padova, Italy.

[13]Department of Anesthesia and Intensive Care, “S. Maria di Ca’ Foncello,” Treviso, Italy.

[14]Department of Anaesthesia and Critical Care Medicine, University Hospital of Pisa, Pisa, Italy.

[15]Anesthesia and Critical Care Medicine, Città della Salute e della Scienza Hospital, University of Turin, Turin, Italy.

[16]Department of Anesthesia and Intensive Care, University of Cagliari, Cagliari, Italy.

DOI: 10.1097/CCM.0000000000000974

[17]Cardioanesthesia and Intensive Care, Civil Hospital “SS Annunziata,” Sassari, Italy.

[18]Cardiac and Vascular Department, Casa di Cura Villa Verde, Taranto, Italy.

[19]Department of Anesthesia, Intensive Care Medicine, Cardinal Massaia Hospital, Asti, Italy.

[20]Division of Cardiac Surgery, University of Genova Medical School, Genova, Italy.

[21]Department of Anesthesiology and Intensive Care, Semmelweis University, Budapest, Hungary.

[22]Anesthesia and Resuscitation, United Company Hospital Papardo-Piemonte, Messina, Italy.

[23]Department of Intensive Care, Austin Hospital, University of Melbourne, Melbourne, Australia.

This work was performed at Department of Anesthesia and Intensive Care, IRCCS San Raffaele Scientific Institute, Milan, Italy.

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 (http://journals.lww.com/ccmjournal).

Supported, in part, by departmental funds only.

Dr. Zucchetti disclosed consulting, board membership, and receiving grant support. Dr. Bellomo is currently involved in the intense public debate on the use of starch-containing solutions which might be perceived as an “academic” conflict of interest. The remaining authors have disclosed that they do not have any potential conflicts of interest.

For information regarding this article, E-mail: landoni.giovanni@hsr.it

Objectives: We aimed to identify all treatments that affect mortality in adult critically ill patients in multicenter randomized controlled trials. We also evaluated the methodological aspects of these studies, and we surveyed clinicians’ opinion and usual practice for the selected interventions.

Data Sources: MEDLINE/PubMed, Scopus, and Embase were searched. Further articles were suggested for inclusion from experts and cross-check of references.

Study Selection: We selected the articles that fulfilled the following criteria: publication in a peer-reviewed journal; multicenter randomized controlled trial design; dealing with nonsurgical interventions in adult critically ill patients; and statistically significant effect in unadjusted landmark mortality. A consensus conference assessed all interventions and excluded those with lack of reproducibility, lack of generalizability, high probability of type I error, major baseline imbalances between intervention and control groups, major design flaws, contradiction by subsequent larger higher quality trials, modified intention to treat analysis, effect found only after adjustments, and lack of biological plausibility.

Data Extraction: For all selected studies, we recorded the intervention and its comparator, the setting, the sample size, whether enrollment was completed or interrupted, the presence of blinding, the effect size, and the duration of follow-up.

Data Synthesis: We found 15 interventions that affected mortality in 24 multicenter randomized controlled trials. Median sample size was small (199 patients) as was median centers number (10). Blinded trials enrolled significantly more patients and involved more centers. Multicenter randomized controlled trials showing harm also involved significantly more centers and more patients ($p = 0.016$ and $p = 0.04$, respectively). Five hundred fifty-five clinicians from 61 countries showed variable agreement on perceived validity of such interventions.

Conclusions: We identified 15 treatments that decreased/increased mortality in critically ill patients in 24 multicenter randomized controlled trials. However, design affected trial size and larger trials were more likely to show harm. Finally, clinicians view of such trials and their translation into practice varied. (Crit Care Med 2015; 43:1559–1568)

Key Words: consensus conference; critically ill patients; intensive care unit; multicenter randomized controlled trials; noninvasive ventilation; nonsurgical interventions; treatments to increase and decrease mortality

Critically ill patients have high mortality rates (1) and account for a large part of hospital expenditure in the Western world (2). Any intervention leading to a reduction in mortality in such patients may save thousands of lives per year worldwide.

Over the last 50 years, the *New England Journal of Medicine* (NEJM) has published at least 12 multicenter randomized controlled trials (mRCTs) (3–14) performed in critically ill patients, which showed a statistically significant difference in unadjusted landmark mortality between treatment and control groups and were not later contradicted by larger or higher quality studies. Over the same period, a further 12 mRCTs (15–26) in the same population were also published in nine other journals. However, despite the presence of such a substantial seemingly robust body of evidence, no studies have assessed whether there are differences in trial features between those reporting benefit versus harm and whether clinicians use the findings of such trials in clinical practice.

We identified all mRCTs reporting an effect on mortality in critically ill patients, we assessed their internal and external validity in a consensus conference, and finally we surveyed more than 500 physicians from 61 countries on how such evidence might be currently translated into practice worldwide. Lastly, we evaluate methodological aspects of the selected trials.

MATERIALS AND METHODS

Systematic Search and Initial Article Selection

MEDLINE/PubMed, Scopus, and Embase were searched by four investigators with no publication time limits to identify all mRCTs of any intervention influencing unadjusted landmark mortality in critically ill patients (see full MEDLINE/PubMed search strategy, updated to June 20, 2013, in the Supplementary Appendix, Supplemental Digital Content 1, http://links.lww.com/CCM/B243).

Further articles were suggested from experts and cross-check of references.

Articles were selected only when fulfilling all the following criteria: 1) publication in a peer-reviewed journal; 2) mRCT design; 3) dealing with nonsurgical interventions (drug/technique/bundle of care) in adult critically ill patients; 4) statistically significant reduction or increase in unadjusted landmark mortality.

We considered all those critically ill patients with acute failure of at least one organ and/or need for intensive care treatment and/or emergency treatment, regardless of where they were treated: intensive care ward, emergency department, or general ward. All trials involving more than one hospital were considered multicentric.

Difference in mortality was considered statistically significant when present at a specific time point (landmark mortality) with simple statistical tests and without adjustment for baseline characteristics.

We excluded all studies that: 1) used a quasi-randomized or nonrandomized methodology; 2) dealt with surgical interventions; 3) involved pediatric population; 4) dealt only with the perioperative period; 5) were performed out of hospital; 6) showed a mortality effect only in a population subgroup or showed a mortality effect only after adjusted analysis; or 7) had low (< 50%) agreement levels among surveyed clinicians.

Consensus Conference Meeting and Final Article Selection

On June 20, 2013, a core group of experts participated in a face-to-face consensus conference to assess and evaluate methodological robustness of all interventions identified; several studies were excluded on methodological grounds because of lack of reproducibility or generalizability, high probability of type I error, major baseline imbalances between intervention and control groups, major design flaws, contradiction by subsequent larger trials, modified intention-to-treat analysis, effect found only after adjustments, and lack of biological plausibility. Speci cally, biological plausibility represented the relationship of the study with previous information, such as pathophysiological rationale and estimated size effect, knowledge and investigations in the eld, reproducibility referred to the presence of con rmation in subsequent larger trials of the same intervention, and generalizability represented the external validity of such ndings outside the unique trial settings (27). Trials characterized by small sample size or a low rate of observed events were considered at high risk of type I error. Furthermore, in some studies, the patients in the control group were treated outside current standards of care. These studies were also removed. These evaluations were qualitative and based on an unanimous decision of the consensus group. Table S1 (Supplemental Digital Content 1, http://links.lww.com/CCM/B243) reports the mRCTs excluded and the reason for exclusion.

The International Web-Based Survey

Through an interactive web questionnaire at http://www.democracybasedmedicine.org, active for 3 months, from June 28, 2013, to September 28, 2013, we asked clinicians whether they agreed or disagreed with the validity of each intervention and whether they used or avoided each intervention in clinical practice. The authors included the option "don't know" and "not available" in the questionnaire to allow respondents to state that they had no opinion on a particular issue or do not have the possibility to use a particular drug. The brief questionnaire was rst given to a restricted number of clinicians to test its clarity. No concerns arose and answers were consistent with a correct interpretation of the questions.

E-mail addresses were those of corresponding authors of articles published in the last 10 years on peer-reviewed journals dealing with intensive care, anesthesiology, emergency medicine, cardiac surgery, and cardiology.

Since methodological research suggests that there is no difference in response rate depending on the inclusion or exclusion of the "don't know" option (if < 40%), we reported only the "yes" and "no" frequencies (28).

We excluded interventions with an agreement rate of less than 50% (Table S2, Supplemental Digital Content 1, http://links.lww.com/CCM/B243).

Throughout the process, all participants were asked to disclose any potential con icts of interest.

Statistical Analysis

For all selected studies, we recorded and analyzed as variables: 1) the intervention and its comparator; 2) the effect on survival; 3) the setting of the trial; 4) the sample size (number of centers and number of patients); 5) whether enrollment was completed or interrupted after interim analysis; 6) the presence of blinding; and 7) the duration of follow-up.

For each of the selected trials, size effect was assessed. From the data provided in the articles, we calculated relative risk reduction or increase (RRR/RRI), absolute risk reduction or increase (ARR/ARI), and number needed to treat (NNT) or number needed to harm (NNH).

Descriptive statistics were used to examine study variables. Values are expressed as medians with interquartile range (IQR). The difference between two groups was calculated with the Mann-Whitney U test, and when more than two groups were involved, Kruskal-Wallis test was used. To calculate the association between study variables (in this case RRR, ARI, increased mortality, RRI, NNT, and NNH), the chi-square test (in case of dichotomous variables) and/or Spearman correlation test are used. Statistical signi cance was assumed for p value less than 0.05.

The results of the web vote are expressed as percentage of positive votes. Null votes were excluded. We reported both the percentage of agreement with selected literature and use/avoidance in clinical practice.

Statistical analysis was performed using STATA 13 software (StataCorp, College Station, TX).

RESULTS

We identi ed 15 treatments that in uenced (decreased or increased) unadjusted landmark mortality in critically ill patients as documented by 24 mRCTs (3–26), 12 of which published by the NEJM (3–14) (Fig. S1, Supplemental Digital Content 2, http://links.lww.com/CCM/B244; Fig. S2, Supplemental Digital Content 3, http://links.lww.com/CCM/B245).

Interventions That Decreased Mortality

At the time of analysis, seven treatments decreased mortality: 1) noninvasive ventilation (NIV) for speci c population with acute respiratory failure (5, 18–24); 2) mild hypothermia after cardiac arrest (4); 3) prone positioning (6) and 4) low tidal volume ventilation (7, 8, 25) in acute respiratory distress syndrome (ARDS); 5) tranexamic acid in patients with or at high risk of traumatic hemorrhagic shock (26); 6) daily interruption of sedatives in critically ill patients (17); and 7) albumin administration in cirrhotic patients with spontaneous bacterial peritonitis (3) (Tables 1 and 2). Only two of these studies were blinded.

Noninvasive mechanical ventilation (NIV) was the treatment supported by the greatest number of mRCTs, with eight mRCTs showing a statistically signi cant survival improvement in patients affected by acute respiratory failure in a variety of contexts, such as acute exacerbation of chronic obstructive pulmonary disease (COPD) and respiratory acidosis (5, 19, 22, 24), hypoxemic respiratory failure (20), and weaning from invasive mechanical ventilation (18, 21, 23).

However, such evidence in favor of NIV was dependent on an effect in the speci c population of COPD patients (six out of eight showing bene t). These mRCTs enrolled a median of 98 patients (IQR, 48–120) and involved a median of three centers. Only two trials involved more than 10 centers (IQR, 3–7) and only one enrolled more than 199 patients. A total of 916 patients were enrolled. Only another intervention was supported by more than one mRCT (low tidal volume mechanical ventilation with or without high positive end-expiratory pressure in ARDS) (7, 8, 25). These three trials, however, were all interrupted after ad interim analysis because of increased survival in treatment group.

TABLE 1. Multicenter Randomized Controlled Trials of Nonsurgical Intervention Reporting a Significant Reduction in Mortality: Population, Intervention, and Comparator

Treatment	Population	Intervention	Comparator
Albumin in hepatorenal syndrome (3)	Patients with cirrhosis and spontaneous bacterial peritonitis	Cefotaxime and albumin IV	Cefotaxime IV
Daily interruption of sedatives (17)	Mechanically ventilated patients	Daily spontaneous awakening trial + spontaneous breathing trial	Daily spontaneous breathing trial
Mild hypothermia (4)	Patients with return of spontaneous circulation after witnessed cardiac arrest	Therapeutic hypothermia (32–34°C)	Normothermia
Noninvasive ventilation (5)	Decompensated COPD exacerbation	NIV	Standard treatment
Noninvasive ventilation (18)	Intubated COPD patients, after a failed weaning trial	NIV after accelerated weaning and extubation	Invasive PSV and standard weaning
Noninvasive ventilation (19)	COPD exacerbation	NIV + medical therapy	Oxygen therapy + medical therapy
Noninvasive ventilation (20)	Severe hypoxemic ARF	NIV	High-concentration oxygen therapy
Noninvasive ventilation (21)	Patients at high risk for postextubation respiratory failure	NIV immediately after extubation for 24 hr	Oxygen therapy after extubation
Noninvasive ventilation (22)	Intubated COPD patients	NIV after accelerated weaning and extubation	Invasive synchronized invasive mechanical ventilation + PSV and standard weaning
Noninvasive ventilation (23)	Intubated COPD patients	NIV after extubation	Oxygen therapy after extubation
Noninvasive ventilation (24)	Very old (> 75 yr) COPD patients with ARF	NIV	Standard medical therapy
Prone position (6)	Severe ARDS	Prone position for 16 consecutive hours + standard treatment	Standard treatment
Protective ventilation (7)	Severe ARDS	High PEEP, low tidal volume	Low PEEP, tidal volume 12 mL/kg
Protective ventilation (8)	Severe ARDS	Low tidal volume, plateau pressure < 30 cm H_2O	Tidal volume 12 mL/kg, plateau pressure < 50 cm H_2O
Protective ventilation (25)	Severe ARDS	Low tidal volume, PEEP = lower inflection point + 2 cm H_2O	Tidal volume 9–11 mL/kg, PEEP > 5 cm H_2O
Tranexamic acid (26)	Trauma patients with or at risk of significant hemorrhage	Tranexamic acid	Placebo

COPD = chronic obstructive pulmonary disease, NIV = noninvasive ventilation, PSV = pressure support ventilation, ARF = acute respiratory failure, ARDS = acute respiratory distress syndrome, PEEP = positive end-expiratory pressure.

Interventions That Increased Mortality

Eight interventions increased mortality: 1) diaspirin cross-linked hemoglobin in traumatic hemorrhagic shock (15); 2) hydroxyethyl starch in septic shock (12); 3) ventilation with high-frequency oscillation (13); 4) IV salbutamol (16) in ARDS; 5) glutamine supplementation (14); 6) growth hormone treatment (10); 7) supranormal systemic oxygen delivery (9); and 8) intensive insulin therapy (11) (**Tables 3** and **4**). Of these studies, five were blinded. See **Tables S3** and **S4** (Supplemental Digital Content 1, http://links.lww.com/CCM/B243) for trial characteristics.

Major Exclusions

Sixteen articles were excluded by the Consensus Conference (details in Tables S1, S3, and S4, Supplemental Digital Content 1, http://links.lww.com/CCM/B243). Sample size

TABLE 2. Multicenter Randomized Controlled Trials of Nonsurgical Intervention Reporting a Significant Reduction in Mortality: Trial Size, Size Effect, Follow-Up, End of Enrollment, and Blinding

Treatment	Centers	Patients	*p*	Absolute Risk Reduction	Relative Risk Reduction	Number Need to Treat to Save One Life	Follow-Up	Stopped at Interim Analysis	Blinding
Albumin in hepatorenal syndrome (3)	7	126	0.01	0.191	0.668	5	Hospital discharge[a]; 90 d[a]	No	Yes
Daily interruption of sedatives (17)	4	336	0.01	0.134	0.232	7	28 d; 1 yr[a]	No	No
Mild hypothermia (4)	9	275	0.02	0.142	0.258	7	Hospital discharge, 6 mo[a]	No	No
Noninvasive ventilation (5)	5	85	0.02	0.193	0.675	5	Hospital discharge[a]	No	No
Noninvasive ventilation (18)	3	50	0.009	0.2	0.714	5	60 d[a]	No	No
Noninvasive ventilation (19)	14	236	0.05	0.101	0.498	10	Hospital discharge[a]	No	No
Noninvasive ventilation (20)	3	105	0.028	0.213	0.548	5	ICU discharge[a]; 90 d[a]	No	No
Noninvasive ventilation (21)	2	162	0.025	0.142	0.871	8	ICU discharge[a]; hospital discharge; 90 d[a]	No	No
Noninvasive ventilation (22)	11	90	0.015	0.12	0.828	7	Hospital discharge[a]	No	No
Noninvasive ventilation (23)	3	106	0.0244	0.197	0.64	5	ICU discharge; hospital discharge; 90 d[a]	No	No
Noninvasive ventilation (24)	3	82	0.014	0.122	0.836	8	Hospital discharge[a]; 6 mo[a]; 1 yr[a]	No	No
Prone position (6)	27	474	< 0.001	0.168	0.512	6	28 d; 90 d[a]	No	No
Protective ventilation (7)	2	53	< 0.001	0.329	0.465	3	ICU discharge[a]; hospital discharge; 28 d[a]	Yes	No
Protective ventilation (8)	10	861	0.007	0.088	0.222	11	Hospital discharge[a]	Yes	No
Protective ventilation (25)	8	103	0.017	0.238	0.441	4	ICU discharge[a]; hospital discharge[a]; 28 d[a]	Yes	No
Tranexamic acid (26)	247	20,211	0.0035	0.015	0.094	68	Hospital discharge[a]	No	Yes

[a]Significant.

was generally small: median number of patients 115 (IQR, 77–180) and median number of centers 7 (IQR, 4–13). Ten of these mRCTs were blinded, and all 16 studies showed an improved survival.

Four more interventions were excluded after the web-based survey because of low agreement (< 50%) of their efficacy among clinicians (Table S2, Supplemental Digital Content 1, http://links.lww.com/CCM/B243). Two interventions may improve survival (antimicrobial therapy in patients with ventilator-associated tracheobronchitis [29] and enteral antioxidant supplementation [30]) and two may increase mortality (NIV in early respiratory failure after extubation [31] and nitric oxide synthase inhibitor (546C88) in septic patients [32]). These studies enrolled a median number of patients of 223 (IQR, 180–367) and involved a median of 25 centers (IQR, 10–59). Two of them (30, 32) were blinded.

TABLE 3. Multicenter Randomized Controlled Trials of Nonsurgical Intervention Reporting a Significant Increase in Mortality: Population, Intervention, and Comparator

Treatment	Population	Intervention	Comparator
Supranormal elevation of systemic oxygen delivery (9)	Patients who failed to reach target hemodynamic values after fluid resuscitation	Dobutamine + standard intensive care	Standard intensive care
Diaspirin cross-linked hemoglobin (15)	Traumatic hemorrhagic shock	10% modified tetrameric hemoglobin solution	Saline infusion
Growth hormone (10)	Patients expected to need intensive care for at least 10 d	Growth hormone	Placebo
Tight glucose control (11)	Patients expected to need intensive care for at least 3 d	Target blood glucose range 81–108 mg/dL	Target blood glucose < 180 mg/dL
IV Salbutamol (16)	ARDS (within 72 hr of onset)	Salbutamol IV	Placebo
Hydroxyethyl starch (12)	Severe sepsis	Fluid resuscitation with 6% hydroxyethyl starch 130/0.42	Fluid resuscitation Ringer's acetate
High-frequency oscillatory ventilation (13)	Moderate and severe ARDS	High-frequency oscillatory ventilation	Low tidal volumes and high positive end-expiratory pressure
Glutamine supplementation (14)	Intubated patients with multiple organ failure	Glutamine supplementation IV and enterally + selenium IV and enterally + zinc, beta carotene, vitamin E, and vitamin C enterally	Placebo IV and enterally

ARDS = acute respiratory distress syndrome.

Characteristics of the Selected Trials

Overall, only seven trials (29%) were blinded. Blinding was associated with trial size. Blinded trials enrolled more patients (median, 532 [IQR, 126–1,223] vs 106 [90–336]; $p = 0.039$) and involved more centers (median, 26 [IQR, 18–46] vs 5 [IQR, 3–11]; $p = 0.008$) than nonblinded trials. Furthermore,

TABLE 4. Multicenter Randomized Controlled Trials of Nonsurgical Intervention Reporting a Significant Increase in Mortality: Trial Size, Size Effect, Follow-Up, End of Enrollment, and Blinding

Treatment	Centers	Patients	p	Absolute Risk Increase	Relative Risk Increase	Number Needed to Harm	Follow-Up	Stopped at Interim Analysis	Blinding
Supranormal elevation of systemic oxygen delivery (9)	2	100	0.04	0.2	0.667	5	ICU discharge[a]; hospital discharge[a]	No	No
Diaspirin cross-linked hemoglobin (15)	18	112	0.015	0.221	0.902	5	48 hr[a]; 28 d[a]	No	Yes
Growth hormone (10)	18	532	< 0.001	0.221	1.163	5	ICU discharge[a]; 6 mo[a]	No	Yes
Tight glucose control (11)	42	6,104	0.02	0.026	0.104	38	Hospital discharge; 28 d[a]; 90 d[a]	No	No
IV Salbutamol (16)	46	326	0.02	0.109	0.468	9	ICU discharge; hospital discharge; 28 d[a]	Yes	Yes
Hydroxyethyl starch (12)	26	804	0.03	0.075	0.174	13	28 d; 90 d[a]	No	Yes
High-frequency oscillatory ventilation (13)	39	548	0.005	0.117	0.332	9	ICU discharge[a]; hospital discharge[a]; 28 d[a]	Yes	No
Glutamine supplementation (14)	40	1,223	0.05	0.052	0.191	19	Hospital discharge[a]; 28 d; 6 mo[a]	No	Yes

[a]Significant.

nonblinded trials were more likely to show a mortality bene t than blinded trials ($p = 0.011$).

In addition, mRCTs showing an increase in mortality involved more centers (median, 33 [IQR, 18–41] vs 5 [IQR, 3–11]; $p = 0.017$) and enrolled almost ve times more patients (median, 540 [IQR, 219–1,014] vs 116 [IQR, 88–306]; $p = 0.043$) than those showing an improved survival. Five mRCTs were interrupted after interim analysis: three for bene t (7, 8, 25) and two for harm (13, 16).

Overall sample size was small with a median (IQR) of 199 (IQR, 102–536) patients and a median of 10 (IQR, 3–26) centers. As shown in Figure S3 (Supplemental Digital Content 4, http://links.lww.com/CCM/B246), large mRCTs were a minority, although both the number of centers involved and the number of patients enrolled appear to have increased over the time (Fig. 1).

Duration of follow-up varied greatly across the studies, ranging from 48 hours to 1 year but was not related to outcome. Most studies (21 out of 24) investigated medium-term mortality (i.e., in-hospital survival and 28-d survival). Among the studies that showed a decrease in mortality, nine out of 16 had a longer-term (i.e., from 60 d to 1 yr) follow-up. Among the studies with increased mortality, four out of eight had a longer-term follow-up (from 60 d to 6 mo). Two trials (21, 24) showed a statistically signi cant decrease in mortality when measured early (ICU) mortality and 1-month mortality, but no effect after longer follow-up (in-hospital mortality and 1-yr mortality, respectively).

The median ARR for interventions that decreased mortality was 0.12 (IQR, 0.12–0.2), and the median RRR was 0.53 (IQR, 0.35–0.69). The median ARI for interventions that increase mortality was 0.11 (IQR, 0.06–0.21), and the median RRI was 0.4 (IQR, 0.18–0.78).

The median NNT was 7 (IQR, 5–8) and median NNH was 9 (IQR, 5–16). No statistically signi cant correlation was found between effect size and outcome or blinding. We found a statistically signi cant correlation between trial size and effect size (Fig. S3, Supplemental Digital Content 4, http://links.lww.com/CCM/B246). Funding was declared in 21 studies (82%) and came from public sources in most cases (16 studies; 67%). See Table S5 (Supplemental Digital Content 1, http://links.lww.com/CCM/B243) for statistical analysis details.

For sensitivity, we repeated these analyses focusing on all studies identi ed by the systematic analysis (Table S6, Supplemental Digital Content 1, http://links.lww.com/CCM/B243) and we compared the descriptive statistics of the selected articles with that of the excluded ones (Table S7, Supplemental Digital Content 1, http://links.lww.com/CCM/B243). Our ndings were not signi cantly changed. As observed in the selected articles, smaller trials were more likely to show an improvement in survival ($p < 0.01$) and to be unblinded ($p = 0.01$). Trials that showed a positive effect on survival had a smaller NNT (6 vs 10; $p = 0.02$) and a larger ARR (0.177 vs 0.107; $p = 0.04$). This correlation was lost in the selected article. Excluded trials were smaller, but this difference was not statistically signi cant. All trials excluded by the consensus showed an improved survival.

Clinicians' Responses

In total, 555 clinicians from 61 countries responded to our survey at http://www.democracybasedmedicine.org and reported a variable degree of agreement with trial results and use in clinical practice (Tables S8 and S9, Supplemental Digital Content 1, http://links.lww.com/CCM/B243). The more represented countries in the web poll were the United States (11%), Australia (11%), and Italy (11%) (Table S10, Supplemental Digital Content 1, http://links.lww.com/CCM/B243). Eighty percent of the voters identi ed themselves as intensive care specialists. Agreement with literature did not differ according to trial outcome; trials showing decreased mortality had a median agreement rate of 81.3% (SD, 9.3%), whereas those showing increased mortality had an 81.6% median agreement rate (SD, 7.2%). The percentage of use/avoidance was not in uenced by the year of publication (Spearman correlation test, $p = 0.92$; Fig. S5, Supplemental Digital Content 6, http://links.lww.com/CCM/B248). On average, only 71% of those who agreed with the veracity of the effect of the selected interventions declared to routinely use/avoid them in their clinical practice, and the percentage of those who agreed with the scienti c validity of these interventions but did not routinely use/avoid them increased with a decrease in general agreement (Fig. S6, Supplemental Digital Content 7, http://links.lww.com/CCM/B249). NIV showed the highest percentages of both agreement and use in clinical practice.

Finally, declarations of any con icts of interests assessed for each intervention ranged from 0% to 1.26% per intervention, and the exclusion of these participants did not affect the overall results.

Figure 1. Trend of number of patients and number of centers over time. The diameter of the balloons represents the number of patients enrolled in each trial.

DISCUSSION

Key Findings

We identi ed all nonsurgical interventions for which there is mRCT evidence of an effect on unadjusted landmark mortality in adult critically ill patients. Such mRCTs have small sample size (median patient number below 200 and median center number of 10). Only seven trials were blinded and five were interrupted after interim analyses. Notably, unblinded trials did not study medications, but the use of specific devices (such as NIV or high-frequency oscillatory ventilation) or therapeutic strategies (daily interruption of sedatives, prone position, mild hypothermia after cardiac arrest, protective ventilation, tight glucose control, and supranormal oxygen delivery). In these cases, blinding is very difficult or even impossible. In keeping with this, we found an association between blinding and trial size. Blinded trials enrolled significantly more patients and involved more centers. Furthermore, mRCTs reporting an increase in mortality involved more centers and enrolled more patients than those showing decreased mortality. Finally, there was a clear correlation between the effect size and trial size. Among treatments showing decreased mortality, NIV was supported by the greatest number of mRCTs, but such robustness was essentially dependent on its effect in COPD patients (six trials). Protective ventilation was the only other treatment supported by more than one mRCT. Finally, surveying more than 500 clinicians in 61 countries showed a variable degree of agreement for both scientific validity and the clinical use of such interventions. NIV showed the highest percentages in both agreement and use in clinical practice.

Previous Literature and Methodology

Given the great heterogeneity of critically ill patients, the lack of robust surrogate outcomes, and their high mortality rates, mortality is generally considered the most important primary outcome in ICU RCT (33). However, interventions reported to influence mortality in those patients are relatively few, small or single center in design, and at high risk of type I error. As such, they should only be considered hypothesis generating (34, 35). For these reasons, we focused our attention only on mRCTs as they represent the highest grade of evidence and have a higher degree of external validity and the lowest risk of type I or type II error (34, 35). However, even multicenter investigations in ICU setting often fail to demonstrate effects on mortality or demonstrate an exaggerated effect that is contradicted by subsequent trials. Negative trials may result from true lack of effect or patient heterogeneity, logistic and organizational difficulties (34), limited power, unidentified confounders, or variability in clinician behavior in the complex and peculiar ICU environment (36). The risk of type I error, on the other hand, is generally due to small sample size or paucity of observed events. In a critical care context, investigators might have difficulties to enroll a large number of patients, even with a multicenter design.

Accordingly, interventions showing a significant effect on mortality in critically ill patients in mRCTs are few. They become even fewer after a detailed assessment of quality and adequacy. In agreement with our findings, in 2008, Ospina-Tascón et al (33) assessed all ICU adult RCTs of more than 50 patients with mortality as the primary outcome. These investigators found that only 10 studies reported a beneficial effect and that seven reported harm. Fifty-five studies reported no effect. Furthermore, in 2010, Aberegg et al (37) used high-impact journals to assess RCTs in ICU over a 10-year period and compared the predicted effect with the reported effect. In 38 trials, they found that the mean predicted effect was 10.1% and that the mean actual reported effect was only 1.4%.

Recently, Mueller et al (38) challenged the ethical and scientific validity of stopping RCT early because of apparent benefit. When the observed event rate is low, an unlikely high effect size is needed to reach statistical significance. This can lead to an overestimate of the therapeutic effect of a treatment as well as to a decreased ability to detect serious side effects. Notably, only three trials were interrupted ad interim for benefit, all of them investigating efficacy of protective ventilation strategies in ARDS. Even if the number of events accrued before discontinuation was small in two out of three trials, the reproducibility of the results confirmed their reliability.

Implications for Clinical Practice

Our findings have implications for trials in intensive care. They suggest the need to increase size, centers number, and efforts to blind interventions or to at least blind adjudication when blinding is not possible. Finally, they suggest the need to assessing unadjusted landmark mortality at a time that is remote from the intervention applied in ICU. These steps may increase clinician confidence in the robustness of the results and their translation into practice. The observation that trials showing an increase in mortality appear of greater quality reinforces concerns about the robustness of "positive" findings as does the lack of confirmatory mRCTs after "positive" investigations.

Our findings also have implications for clinicians who are charged with translating evidence into practice. By showing that trials that report harm are of greater quality, they suggest the need to perhaps both consider translating their findings into practice with greater confidence and simultaneously view trials that show benefit with greater caution.

Strengths and Limitations

This study has several strengths. For the first time, to our knowledge, we reviewed all mRCTs reporting an effect on mortality for interventions in critically ill patients. We found that such trials are generally small, raising concerns about the risk of a type I error, and that studies showing an increase in mortality were larger in size, implying that type I errors may be more likely for trials that show an improved survival and that studies that proved an increase in mortality may, therefore, be statistically and perhaps clinically more robust. We also found that blinded trials involved more centers and more patients, suggesting that their statistical robustness adds to their ability to decrease selection bias. Finally, this is the first time in literature that self-reported practice on these interventions has been collected. We found variable degrees of agreement about the use of those findings when clinicians were surveyed, suggesting that translation of evidence into practice remains a complex process even when evidence comes from mRCTs and the outcome is landmark unadjusted mortality. As a matter of fact, some apparently well-established interventions, such as protective ventilation and prone positioning in ARDS, or tranexamic acid in major bleeding, were used by a surprisingly low rate of responders.

Our study also suffers from important limitations. It was completed in June 2013. Evidence-based medicine is an evolving process, sometimes rather quickly. Accordingly, the bene e cial effects of hypothermia after cardiac arrest have been recently challenged (39).

Some of the criteria used to select the trials of interest during the Consensus, such as biological plausibility, high risk of type I error, and major baseline imbalances (as well as external validity and internal validity), cannot be currently quanti ed. It is our major concern that these dimensions need to be evaluated in order to assess the reliability of trial results. Yet these issues have only been minimally discussed by the evidence-based medicine movement and quantitative tools do not exist. In the absence of such criteria, only a qualitative assessment could be carried out. We decided that the only way to tackle such issues was via a consensus conference and to accept the unanimous decision of the consensus group that a given study carried such limitations and should be excluded. An outstanding example of the importance of these elements is River's Early Goal Directed Therapy study (40). This trial was characterized by limited biological plausibility (only h6 of intervention, incredible effect size), high risk of type I error, and limited external validity (41), yet held sway across an evidence based-medicine-based ICU world for a decade, until Protocol-Based Care for Early Septic Shock and Australasian Resuscitation In Sepsis Evaluation trials (42) contradicted its results. Furthermore, the characteristics of excluded articles (trial size and effect size) did not differ from those of the selected articles (Tables S4 and S7, Supplemental Digital Content 1, http://links.lww.com/CCM/B243). Thus, even if such studies were not excluded, their inclusion would not materially change our ndings or conclusions.

Our survey obtained data from all clinicians who chose to submit their views via internet. Thus, we have no denominator to indicate what percentage of physicians exposed to the survey chose to respond and we cannot assess the representativeness of the sample. Furthermore, the selection method was not validated. However, the number of patients who reported their views represents the largest and most international survey of intensive care clinician opinion on ICU treatment reported so far. Self-reported preferences and practice do not reliably re ect actual practice, but provide an initial appreciation of opinion on the use of such interventions worldwide.

CONCLUSIONS

We identi ed 15 treatments that either decrease or increase mortality in critically ill patients according to 24 mRCTs, with NIV alone having eight mRCTs in support of a mortality reduction. We found, however, that both sample size and median number center were small.

Furthermore, only seven trials were blinded and ve were interrupted after interim analyses. Blinded trials enrolled signi cantly more patients and involved more centers. Similarly, mRCTs showing an increase in mortality involved more centers and enrolled more patients than those showing improved survival. Furthermore, there was a clear correlation between the effect size and trial size, such that the greater the size of the trial the greater the NNT or NNH and the smaller the ARR or ARI and the RRR or RRI. Finally, when we surveyed more than 500 clinicians in 61 countries, we found a variable degree of agreement on their use and, for some interventions, application by responders was surprisingly low (e.g., only 85% for protective ventilation, 56% for tranexamic acid, and 55% for prone position). Our ndings suggest that size, methodology quality, and number of centers involved need to increase in critical care trials to allow greater con dence in their ndings.

ACKNOWLEDGMENTS

We thank Maieutics Foundation and Aleph srl for sharing the technical expertise for the voting process.

REFERENCES

1. Mayr VD, Dünser MW, Greil V, et al: Causes of death and determinants of outcome in critically ill patients. Crit Care 2006; 10:R154
2. Halpern NA, Pastores SM: Critical care medicine in the United States 2000-2005: An analysis of bed numbers, occupancy rates, payer mix, and costs. Crit Care Med 2010; 38:65–71
3. Sort P, Navasa M, Arroyo V, et al: Effect of intravenous albumin on renal impairment and mortality in patients with cirrhosis and spontaneous bacterial peritonitis. N Engl J Med 1999; 341:403–409
4. Hypothermia after Cardiac Arrest Study Group: Mild therapeutic hypothermia to improve the neurologic outcome after cardiac arrest. N Engl J Med 2002; 346:549–556
5. Brochard L, Mancebo J, Wysocki M, et al: Noninvasive ventilation for acute exacerbations of chronic obstructive pulmonary disease. N Engl J Med 1995; 333:817–822
6. Guérin C, Reignier J, Richard JC, et al; PROSEVA Study Group: Prone positioning in severe acute respiratory distress syndrome. N Engl J Med 2013; 368:2159–2168
7. Amato MB, Barbas CS, Medeiros DM, et al: Effect of a protective-ventilation strategy on mortality in the acute respiratory distress syndrome. N Engl J Med 1998; 338:347–354
8. The Acute Respiratory Distress Syndrome Network: Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. N Engl J Med 2000; 342:1301–1308
9. Hayes MA, Timmins AC, Yau EH, et al: Elevation of systemic oxygen delivery in the treatment of critically ill patients. N Engl J Med 1994; 330:1717–1722
10. Takala J, Ruokonen E, Webster NR, et al: Increased mortality associated with growth hormone treatment in critically ill adults. N Engl J Med 1999; 341:785–792
11. Finfer S, Chittock DR, Su SY, et al; NICE-SUGAR Study Investigators: Intensive versus conventional glucose control in critically ill patients. N Engl J Med 2009; 360:1283–1297
12. Perner A, Haase N, Guttormsen AB, et al; 6S Trial Group; Scandinavian Critical Care Trials Group: Hydroxyethyl starch 130/0.42 versus Ringer's acetate in severe sepsis. N Engl J Med 2012; 367:124–134
13. Ferguson ND, Cook DJ, Guyatt GH, et al; OSCILLATE Trial Investigators; Canadian Critical Care Trials Group: High-frequency oscillation in early acute respiratory distress syndrome. N Engl J Med 2013; 368:795–805
14. Heyland D, Muscedere J, Wischmeyer PE, et al; Canadian Critical Care Trials Group: A randomized trial of glutamine and antioxidants in critically ill patients. N Engl J Med 2013; 368:1489–1497
15. Sloan EP, Koenigsberg M, Gens D, et al: Diaspirin cross-linked hemoglobin (DCLHb) in the treatment of severe traumatic

hemorrhagic shock: A randomized controlled ef cacy trial. JAMA 1999; 282:1857–1864

16. Gao Smith F, Perkins GD, Gates S, et al; BALTI-2 study investigators: Effect of intravenous-2 agonist treatment on clinical outcomes in acute respiratory distress syndrome (BALTI-2): A multicentre, randomised controlled trial. Lancet 2012; 379:229–235
17. Girard TD, Kress JP, Fuchs BD, et al: Ef cacy and safety of a paired sedation and ventilator weaning protocol for mechanically ventilated patients in intensive care (Awakening and Breathing Controlled trial): A randomised controlled trial. Lancet 2008; 371:126–134
18. Nava S, Ambrosino N, Clini E, et al: Noninvasive mechanical ventilation in the weaning of patients with respiratory failure due to chronic obstructive pulmonary disease. A randomized, controlled trial. Ann Intern Med 1998; 128:721–728
19. Plant PK, Owen JL, Elliott MW: Early use of non-invasive ventilation for acute exacerbations of chronic obstructive pulmonary disease on general respiratory wards: A multicentre randomised controlled trial. Lancet 2000; 355:1931–1935
20. Ferrer M, Esquinas A, Leon M, et al: Noninvasive ventilation in severe hypoxemic respiratory failure: A randomized clinical trial. Am J Respir Crit Care Med 2003; 168:1438–1444
21. Ferrer M, Valencia M, Nicolas JM, et al: Early noninvasive ventilation averts extubation failure in patients at risk: A randomized trial. Am J Respir Crit Care Med 2006; 173:164–170
22. Collaborating Research Group for Noninvasive Mechanical Ventilation of Chinese Respiratory Society: Pulmonary infection control window in treatment of severe respiratory failure of chronic obstructive pulmonary diseases: A prospective, randomized controlled, multi-centred study. Chin Med J (Engl) 2005; 118:1589–1594
23. Ferrer M, Sellarés J, Valencia M, et al: Non-invasive ventilation after extubation in hypercapnic patients with chronic respiratory disorders: Randomised controlled trial. Lancet 2009; 374:1082–1088
24. Nava S, Grassi M, Fanfulla F, et al: Non-invasive ventilation in elderly patients with acute hypercapnic respiratory failure: A randomised controlled trial. Age Ageing 2011; 40:444–450
25. Villar J, Kacmarek RM, Pérez-Méndez L, et al: A high positive end-expiratory pressure, low tidal volume ventilatory strategy improves outcome in persistent acute respiratory distress syndrome: A randomized, controlled trial. Crit Care Med 2006; 34:1311–1318
26. Shakur H, Roberts I, Bautista R, et al; CRASH-2 trial collaborators: Effects of tranexamic acid on death, vascular occlusive events, and blood transfusion in trauma patients with signi cant haemorrhage (CRASH-2): A randomised, placebo-controlled trial. Lancet 2010; 376:23–32
27. Bellomo R, Bagshaw SM: Evidence-based medicine: Classifying the evidence from clinical trials—The need to consider other dimensions. Crit Care 2006; 10:232
28. Van Es JC, Lorence DP, Morgan GW, et al: Don't know responses in environmental surveys. J Environ Educ 1996; 27:13–18
29. Nseir S, Favory R, Jozefowicz E, et al; VAT Study Group: Antimicrobial treatment for ventilator-associated tracheobronchitis: A randomized, controlled, multicenter study. Crit Care 2008; 12:R62
30. Crimi E, Liguori A, Condorelli M, et al: The bene cial effects of antioxidant supplementation in enteral feeding in critically ill patients: A prospective, randomized, double-blind, placebo-controlled trial. Anesth Analg 2004; 99:857–863, table of contents
31. Esteban A, Frutos-Vivar F, Ferguson ND, et al: Noninvasive positive-pressure ventilation for respiratory failure after extubation. N Engl J Med 2004; 350:2452–2460
32. López A, Lorente JA, Steingrub J, et al: Multiple-center, randomized, placebo-controlled, double-blind study of the nitric oxide synthase inhibitor 546C88: Effect on survival in patients with septic shock. Crit Care Med 2004; 32:21–30
33. Ospina-Tascón GA, Büchele GL, Vincent JL: Multicenter, randomized, controlled trials evaluating mortality in intensive care: Doomed to fail? Crit Care Med 2008; 36:1311–1322
34. Vincent JL: We should abandon randomized controlled trials in the intensive care unit. Crit Care Med 2010; 38:S534–S538
35. Bellomo R, Warrillow SJ, Reade MC: Why we should be wary of single-center trials. Crit Care Med 2009; 37:3114–3119
36. Deans KJ, Minneci PC, Suffredini AF, et al: Randomization in clinical trials of titrated therapies: Unintended consequences of using xed treatment protocols. Crit Care Med 2007; 35:1509–1516
37. Aberegg SK, Richards DR, O'Brien JM: Delta in ation: A bias in the design of randomized controlled trials in critical care medicine. Crit Care 2010; 14:R77
38. Mueller PS, Montori VM, Bassler D, et al: Ethical issues in stopping randomized trials early because of apparent bene t. Ann Intern Med 2007; 146:878–881
39. Nielsen N, Wetterslev J, Cronberg T, et al; TTM Trial Investigators: Targeted temperature management at 33°C versus 36°C after cardiac arrest. N Engl J Med 2013; 369:2197–2206
40. The ProCESS Investigators: A randomized trial of protocol-based care for early septic shock. N Engl J Med 2014; 370:1683–1693
41. Rivers E, Nguyen B, Havstad S, et al; Early Goal-Directed Therapy Collaborative Group: Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med 2001; 345:1368–1377
42. ARISE Investigators; ANZICS Clinical Trials Group, Peake SL, Delaney A, et al: Goal-directed resuscitation for patients with early septic shock. N Engl J Med 2014; 371:1496–1506