



# Time to stop randomized and large pragmatic trials for intensive care medicine syndromes: the case of sepsis and acute respiratory distress syndrome

Armand R. J. Girbes<sup>1</sup>, Harm-Jan de Groot<sup>1,2</sup>

<sup>1</sup>Department of Intensive Care Medicine, <sup>2</sup>Department of Anesthesiology, Amsterdam University Medical Centers, location VUmc, Amsterdam, The Netherlands

*Contributions:* (I) Conception and design: All authors; (II) Administrative support: All authors; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: HJ de Groot; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

*Correspondence to:* Armand R. J. Girbes, MD, PhD. Department of Intensive Care Medicine, Amsterdam University Medical Centers, location VUmc, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands. Email: a.girbes@planet.nl; arj.girbes@amsterdamumc.nl.

**Abstract:** In this paper we discuss the limitations of large randomized controlled trials with mortality endpoints in patients with critical illness associated diagnoses such as sepsis. When patients with the same syndrome diagnosis do not share the pathways that lead to death (the attributable risk), any therapy can only lead to small effects in these populations. Using Monte Carlo simulations, we show how the syndrome-attributable risks of critical illness-associated diagnoses are likely overestimated using common statistical methods. This overestimation of syndrome-attributable risks leads to a corresponding overestimation of attainable treatment effects and an underestimation of required sample sizes. We demonstrate that larger and more ‘pragmatic’ randomized trials are not the solution because they decrease therapeutic and diagnostic precision, the therapeutic effect size and the probability of finding a beneficial effect. Finally, we argue that the most logical solution is a renewed focus on mechanistic research into the complexities of critical illness syndromes.

**Keywords:** Acute respiratory distress syndrome (ARDS); research methods; clinical trials; Monte Carlo simulation

Submitted Sep 12, 2019. Accepted for publication Oct 11, 2019.

doi: 10.21037/jtd.2019.10.36

View this article at: <http://dx.doi.org/10.21037/jtd.2019.10.36>

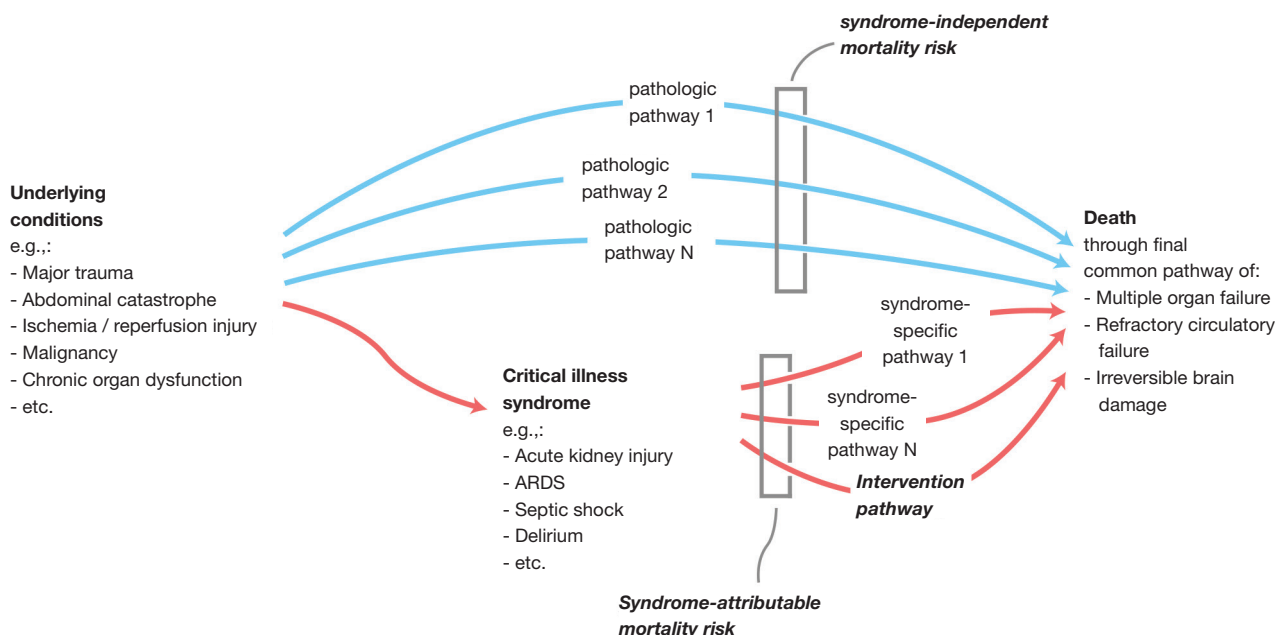
## Introduction

Clinical terminology describing syndromes, such as sepsis, acute kidney injury (AKI) and acute respiratory distress syndrome (ARDS) is extremely useful in everyday practice. A diagnosis of sepsis facilitates early recognition, concise communication and appropriate treatment of serious infections, just as a diagnosis of ARDS can help galvanize a medical team to initiate lung protective ventilator strategies. But the clinical usefulness of these syndromes may not necessarily extend to other aspects of medicine, such as randomized intervention research.

It has been reported that up to 95% of all randomized critical care trials fail to demonstrate a positive and

reproducible mortality effect (1). This abysmal success rate has ethical implications for the recruitment of patients and has important consequences for the allocation of resources. Researchers before us have pointed at the heterogeneity of the critical care syndromes as a major cause of negative trials. But so far, structured analyses of heterogeneity as a cause of ‘negative’ trials have been limited to the role of differential treatment effects along the spectrum of illness severity (2).

Here, we describe why sepsis and other clinical syndromes are unsuitable for randomized controlled trials aiming to detect a mortality reduction—or, conversely, why mortality is a flawed primary endpoint for trials



**Figure 1** Schematic representation of syndrome-attributable risk. Critical illness syndromes are associated with high mortality risks, but the underlying conditions that cause these syndromes are themselves important root causes of mortality. ARDS, acute respiratory distress syndrome.

with syndrome-defined populations. Syndromes such as sepsis, AKI, ARDS or delirium are only reductive models for inherently multifactorial processes and are therefore inappropriate to identify target populations for randomized trials aiming to demonstrate a mortality benefit.

Firstly, we demonstrate how the fraction of mortality risk that is specifically attributable to the critical care syndromes is usually overestimated. This has led to an overestimation of the attainable magnitude of treatment effects and a consequent underestimation of required trial sample sizes by several orders of magnitude. Secondly, we show how large ‘pragmatic’ randomized controlled trials are no effective solution because they water down both treatment effects and diagnostic precision.

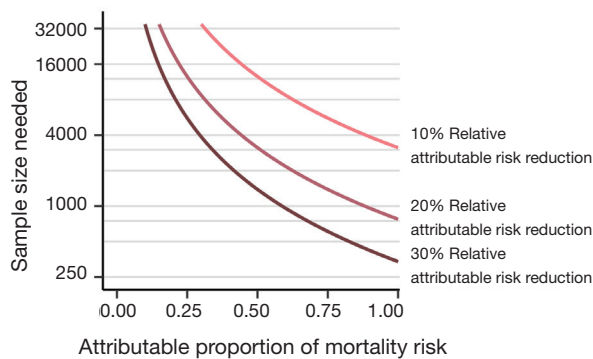
### The importance of syndrome-attributable risk

As an illustration, we can examine a hypothetical critical illness syndrome that is associated with a 35% mortality rate, it could be sepsis but also ARDS. When testing a novel intervention that could potentially reduce the harmful sequelae of this syndrome, it may seem reasonable to power a randomized controlled trial to detect a 20% relative risk reduction. It would take a 1,378-patient trial for 80% power

to demonstrate this effect (see Supplementary file). But this common trial design neglects a fundamental question: which fraction of the total mortality risk is attributable to the syndrome rather than the underlying disease? Observing that the syndrome is associated with a 35% mortality rate is not sufficient. The underlying and independent causes of the syndrome (including systemic infections, abdominal catastrophes, severe pancreatitis, etc.) are at the root of the high mortality rate.

*Figure 1* depicts how mortality depends on the one hand on pathologic pathways that are independent of the syndrome, and on the other hand on the syndrome-attributable mortality risk. The underlying conditions that may cause the syndrome are themselves important risk factors for mortality. To put this in a practical example, if we consider in a patient ARDS as the syndrome and colon carcinoma with perforation as the underlying condition: the therapy for ARDS may be outstandingly performed by the best intensivist expert in the world on mechanical ventilation, but if the colon perforation is not well treated the patient will die no matter how well the patient is mechanically ventilated.

Estimating the possible mortality effects of a syndrome-specific intervention (e.g., low tidal volumes in ARDS, low



**Figure 2** Trial sample size requirements as a function of the syndrome-attributable fraction of mortality for a two-group trial at a 35% control-group mortality rate, 5% type-I error rate and 20% type-II error rate. The relationship between required sample sizes and the syndrome-attributable fraction of mortality is displayed for three different hypothesized effect sizes. Vertical axis on the log scale.

driving pressures) requires information about the syndrome-attributable mortality risk and—an estimation—of the possible attributable risk reduction of the intervention. The concept and importance of attributable risk was already described in the 1970s but seems not to have been addressed sufficiently in critical care trials (3). If only a small part of the mortality risk is conferred through syndrome-specific pathways, then syndrome-specific interventions can only lead to infinitesimal mortality effects. At a 35% mortality rate, a syndrome-attributable fraction of mortality around 0.50 to 0.25 would—depending on the expected treatment effect—realistically require 5,700 to more than 23,000 patients (Figure 2). This means that most large randomized trials with critically ill patients—with sample sizes typically between 500 and 2,000—are based on unrealistic power calculations.

The latest sepsis criteria were derived from risk patterns in large cohorts, so that sepsis is by definition a risk factor for poor outcomes. A ‘septic’ patient deserves immediate attention and treatment of the underlying cause. But what is the causal contribution of sepsis to the mortality risk? The sepsis syndrome can be seen, for example, in a 70-year-old patient with cardiac disease and a large bowel perforation or in a 45-year-old with chemotherapy-induced aplasia and an infection of unknown origin. It is no surprise that with these different and independently catastrophic underlying pathologies, the causal fraction of risk conferred through the sepsis syndrome could be much lower than is

commonly assumed. Indeed, the sepsis-attributable fraction of mortality was recently estimated to be as low as 0.15 by Shankar-Hari *et al.* (4). But as we will point out, even this modest estimate is likely an overestimation.

### Overestimation of the attributable mortality risk: A Monte Carlo simulation study

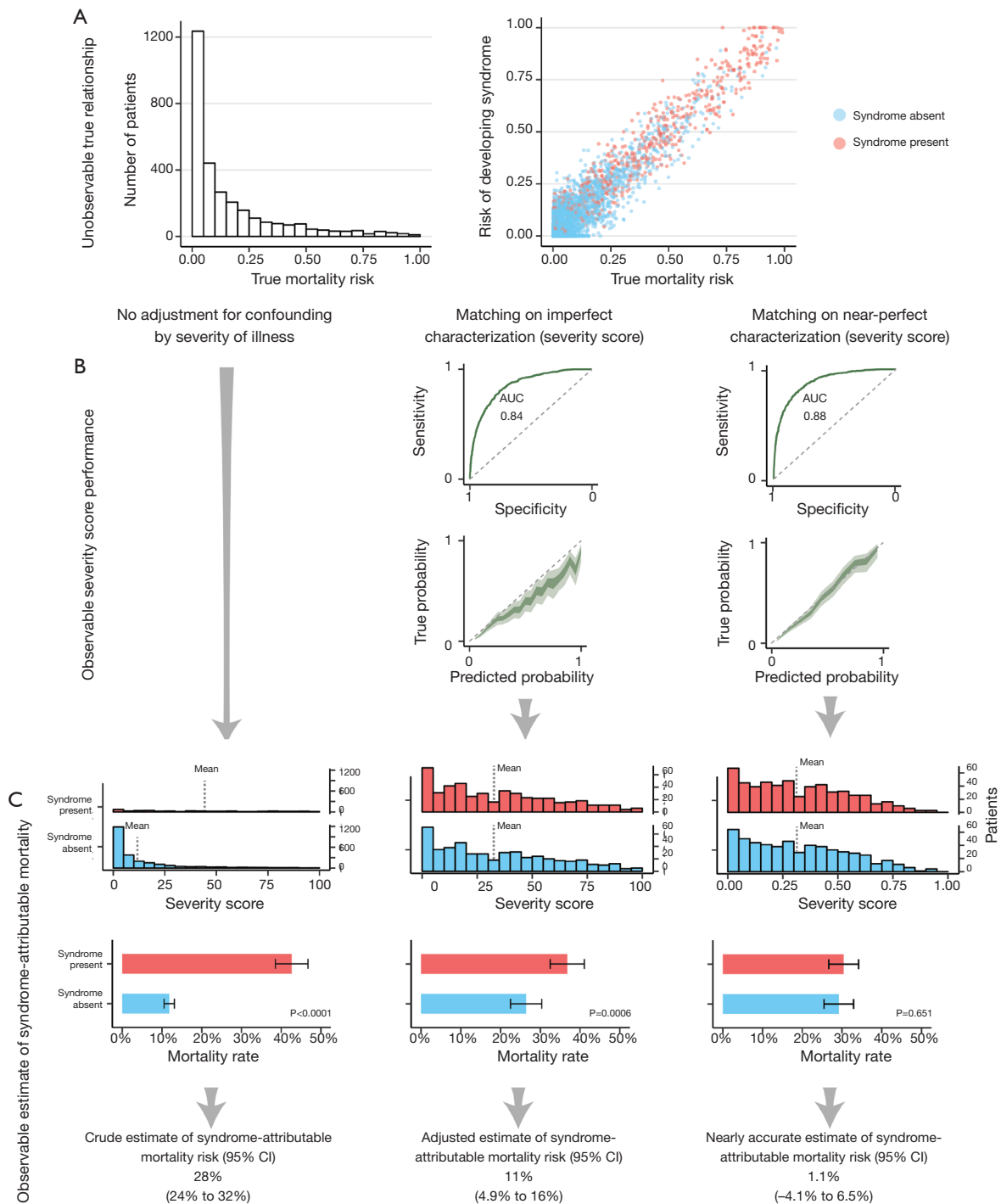
Our impression of attributable mortality can be misleading: syndrome-attributable risks are easily overestimated because critical illness syndromes are strongly associated with the severity of the underlying disease, as can be illustrated with a computer simulated example, which is in detail presented in the technical supplement.

Conditional on an association between the risk of developing a syndrome and the underlying mortality risk, common statistical methods to adjust for confounding by illness severity (such as regression or propensity matching) will lead to significantly inflated estimates of the syndrome-attributable risk, even when the actual attributable risk of a syndrome is zero.

We generated a simulated cohort of 4,000 patients with a randomly generated heterogenous mortality risk distribution and a median mortality risk of 7% (IQR, 2–23%). The risk of developing a hypothetical critical illness syndrome was generated semi-randomly to correlate with baseline mortality risk (correlation =0.80). Using a random binomial generator, each patient was then assigned a syndrome diagnosis (absent or present, based on the risk of developing a syndrome) and an outcome (dead or alive, based on the true mortality risk). Importantly, the mortality risk didn’t shift depending on the presence of the syndrome, so that the presence of the syndrome didn’t cause an increase in mortality risk: The true attributable risk was zero (Figure 3A).

To simulate how syndrome-attributable risks can be estimated by investigators, we constructed three analysis scenarios (Figure 3B,C). In the first scenario, a hypothetical researcher makes a crude mortality comparison between those with the syndrome and those without. The researcher will observe that the crude mortality rate of patients with the syndrome is 41% versus 13% for those without it, so that the crude syndrome-attributable mortality is severely overestimated as 28% ( $P<0.0001$ ).

In the second scenario, the researcher makes use of a severity of illness score to match each patient with the syndrome to a patient without the syndrome but with a similar severity score. The severity score is not a perfect



**Figure 3** Results of a simulation study demonstrating how syndrome-attributable risk are overestimated by matching for severity of illness. (A) In a simulated cohort of 4,000 patients, the risk of developing a critical illness syndrome is correlated to the mortality risk, but the presence of the syndrome doesn't *cause* an increase in mortality risk (no horizontal shift between those with the syndrome compared to those without). (B) To adjust for confounding by illness severity, a researcher wants to match patients with the syndrome to similar patients without the syndrome on the basis of disease characterization (summarized by severity scores). (C) The unadjusted comparison shows patients with the syndrome are increased risk of dying. After adjustment by a good but imperfect severity score, the syndrome-attributable mortality appears to be 11%, while full characterization of the underlying (syndrome-independent) mortality risk reveals that the syndrome-attributable risk is negligible.

reflection of the true mortality risk, but is well-calibrated and has an area under the receiver operating characteristics (AUROC) curve of 0.84 to discriminate between survival and death. In this matched cohort, the mortality rate of the patients with the syndrome is 37% versus 26% for those without it, so that the adjusted attributable risk is estimated to be 11% ( $P=0.0006$ ), still a significant overestimation.

In the final scenario, the researcher has full information and perfect characterization of the underlying diseases, so that he can make use of a perfect severity score (or any other distance metric) to match each patient with the syndrome to a patient without the syndrome. The theoretically maximum attainable AUROC for the underlying risk distribution is 0.88. Using full information about the underlying illness risks to construct a matched cohort, the mortality rate of the patients with the syndrome is found to be 30% versus 29% for those without it, so that the accurate adjusted attributable risk is estimated to be a negligible ( $P=0.651$ ).

This simulation shows that, conditional on an association between the risk of developing a syndrome and the underlying mortality risk, the syndrome-attributable mortality risk can only be accurately estimated when there is perfect risk information (or disease characterization) of the underlying pathological conditions. Obviously, this is not a realistic scenario and we conclude that syndrome-attributable risks are often overestimated.

The full code for this simulation study in the R language is available in the Supplementary file.

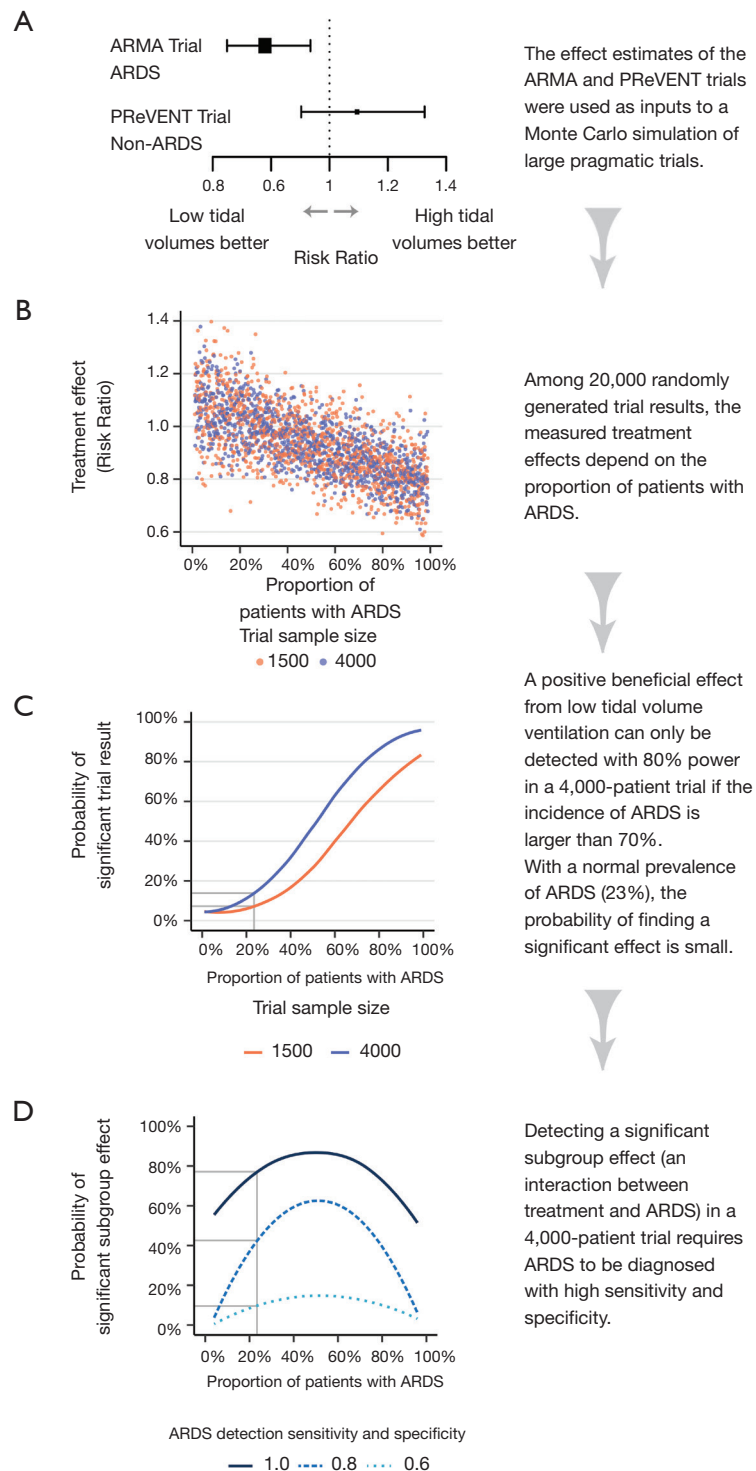
### Large ‘pragmatic’ critical care trials are not the solution

We’ve demonstrated how the attributable mortality rate of syndromes such as sepsis or ARDS is easily overestimated, leading to a corresponding overestimation of attainable treatment effects and an underestimation of required sample sizes. Taking into account the difficulties of large RCTs in terms of patient recruitment and data collection, it may seem tempting to consider large ‘pragmatic’ trials a reasonable solution. These multicenter (often international) trials have broad inclusion criteria and relatively sparse data collection requirements for the participating centers. Aside from the decreased resources needed per recruited patient, pragmatic trials have the additional advantage of good external validity by investigating the ‘real world’ effectiveness of an intervention. But we will show that this is also not the way to go. Recent data on ARDS help us to make the point.

Since the publication of the landmark ARMA trial, low tidal volume ventilation has been unequivocally accepted as the standard of care for patients with ARDS (5), but the history of ARDS could have been different: The 1991 edition of Harrison’s classic medical stated that the goal of ventilation in patients with ARDS was to increase lung volume by applying tidal volumes of 10–15 mL/kg. What if, instead of the ARMA trial, low tidal volume ventilation would have been tested in the same manner as recent large pragmatic trials? For example, what would be the results of a trial that tested the benefit of low tidal volume ventilation in all mechanically ventilated patients using a cluster-randomized design? Could a large pragmatic trial ever demonstrate a beneficial effect of low tidal volume ventilation?

We can answer this thought experiment with the help of the recent Protective Ventilation in Patients Without ARDS (PReVENT) trial, in which 961 patients without ARDS were randomized to mechanical ventilation with tidal volumes of 6 versus 10 mL/kg (6). No meaningful differences were found between the treatment arms for any of the clinical outcomes, including ventilator-free days, length of stay and mortality. We now know the efficacy of low tidal volume ventilation in two complementary populations: In the ARMA trial, care was taken to include only patients with ARDS, whereas in PReVENT trial those patients were strictly excluded. By combining the populations and results of both trials, we can begin to see what a pragmatic trial including all mechanically ventilated patients would look like.

To do so, we ran a Monte Carlo simulation of trials by letting software generate trials with either 1,500 or 4,000 included patients, with a varying proportion of patients with ARDS. Using the results of the ARMA and PReVENT trials, patients with and without ARDS were assigned baseline mortality risks of 40% and 32%, respectively (5,6). Patients were then randomly assigned either treatment or control. Again using the ARMA and PReVENT results, the effect of treatment on patients with ARDS was set as a relative risk of 0.78 (95% CI: 0.65–0.93), and the treatment effect on patients without ARDS was a relative risk of 1.09 (95% CI: 0.90–1.33) (*Figure 4A*) (5,6). Finally, an outcome of death or survival was generated for each patient using a binomial random generator, with the probability of death a function of baseline mortality risk (dependent on ARDS status) multiplied by the treatment effect (depending on ARDS status and assignment to treatment or control). We ran the simulation 10,000 times with 1,500 patients per trial



**Figure 4** Simulation of hypothetical large ‘pragmatic’ trials investigating low tidal volume ventilation. The effects of low tidal volume ventilation in patients with ARDS and without ARDS were based on two large RCTs (A). We simulated the results of large ‘pragmatic’ trials in populations with different incidences of ARDS (B,C) and with changing diagnostic precision (D). ARDS, acute respiratory distress syndrome. ARDS, acute respiratory distress syndrome.

and 10,000 times with 4,000 patients per trial (*Figure 4B*).

*Figure 4C* shows the probability of finding a significant benefit from low tidal volumes in 1,500 and 4,000 patient trial as a function of the proportion of patients with ARDS in those trials. At the worldwide average ARDS prevalence of 23% (7), the probability that a 1,500-patient or 4,000-patient trial would find a benefit from low tidal volume ventilation is 7% or 11%, respectively (*Figure 4C*). Only at an ARDS prevalence above 70% does a 4,000-patients trial have reasonable power (80%) to find a benefit from low tidal volumes.

We can then ask whether low tidal volume ventilation can at least be identified as beneficial in the subgroup of ARDS patients within such a large pragmatic trial. To answer this question, we have to evaluate how accurately ARDS is diagnosed in practice. The ten clinical sites of the ARMA trial who collaborated in the ARDS network were likely much better than averagely trained in the recognition of ARDS. The ARMA trial itself will have increased the focus of the participating clinicians and researchers to the precise diagnosis of ARDS. Real-world data paint a bleaker picture (8). A large study in 459 intensive care units (ICUs) in 50 countries indicated that ARDS is severely underdiagnosed, with clinician-recognition of only 60% of cases (7). Even in the context a hypothetical pragmatic trial with increased checks on patient data the diagnostic accuracy may well be less than perfect, for example due to the large interobserver variability in interpreting the radiographic diagnosis of ARDS (9,10).

*Figure 4D* shows that discovering a significant beneficial effect on the ARDS subset of patients with >80% power requires ARDS to be diagnosed with perfect sensitivity and specificity. With diagnostic sensitivity and specificity of 0.80 the power to discover a significant interaction is only 43% at an ARDS incidence of 23%. With a sensitivity and specificity of 0.60 so many patients with ARDS are confused for no ARDS (and vice versa) that the probability of discovering a beneficial effect remains negligible no matter how high the prevalence.

In summary, this tells us that a large pragmatic trial with unselected ventilated patients would be unlikely to ever demonstrate a beneficial effect of low tidal volumes. Many real-world factors will further decrease the likelihood of discovering a significant effect. For example, we've not considered that even within a target population larger trials tend to demonstrate smaller effects, possibly due to less vigorous protocol adherence (11-13). A recent study found that the limited number of multicenter trials reporting a

significant mortality effect had only moderate sample sizes and few participating centers (median 199 patients and 10 centers) (14).

### How to proceed: a renewed focus on understanding

Over the past half century, there has been an epistemological shift away from physiology- and pathophysiology-based medicine towards knowledge acquisition through clinical trials. In critical care research, this has gone hand in hand with mortality as the gold-standard endpoint. But we must acknowledge that the yield of this research paradigm has been abysmally poor: More than 2,000 randomized controlled trials with sepsis patients have been performed (15), all of which have not resulted in a single beneficial intervention (1,14). Meanwhile, the mechanisms behind organ dysfunction in sepsis remain very much a mystery, as we have only recently discarded the simplistic model of sepsis as a purely unhinged hyperinflammatory reaction to infection (16,17). The cart has been put before the horse by investing heavily in randomized trials without adequately understanding the syndromal pathophysiology. We wonder how far our understanding of sepsis could have come if the resources dedicated to the many randomized trials had been spent on preclinical and clinical mechanistic investigations.

We believe a rigorous rebalancing of research priorities is needed. Only if we properly understand the harmful processes in our complexly ill patients can we hope to distill common pathways. An acute sepsis patient in the hyperinflammatory phase may share more pathophysiological characteristics with a patient with uninfected severe pancreatitis (18) than with another sepsis patient in the immunoparalytic phase (19). Lumping together patients because they fit a consensus definition decreases the proportion of shared pathways and reduces the chance of finding a therapy that benefits all.

We do not advocate a return to mere pathophysiology-based or 'eminence-based' medicine but rather propose that randomized clinical trials with a mortality endpoint should be a final keystone in the chain of evidence when true equipoise remains after the effects of a therapy are properly understood. Such has also been the history of high versus low tidal volume ventilation, the pros and cons of which were already understood before the ARMA trial (20). What remained was uncertainty about the net balance of effects—for which the ARMA trial provided the answer. A trial, we

note, which appears to have been decidedly unpragmatic with very tight protocol adherence (attained tidal volumes 6.2 *vs.* 11.8 mL/kg on protocolized tidal volumes of 6.0 *vs.* 12 mL/kg) (5) compared to real-world scenarios (7).

In adopting the valuable principles of evidence-based medicine the pendulum has swung too far towards large randomized trials as the preferred method of knowledge mining in critical care. A renewed focus on understanding requires the critical care community and funding institutions to re-incentivize basic and translational mechanistic research.

## Conclusions

Mortality is an insensitive trial endpoint because heterogeneous patients pigeonholed into a critical care diagnosis do not share to a large enough extent the pathways that independently lead to death. Ever larger and larger pragmatic trials are not the solution because an increase in ‘pragmatism’ goes hand in hand with a decreased therapeutic and diagnostic precision, thereby reducing the attributable risk, the therapeutic effect size and the probability of finding a beneficial subgroup effect. The way out of this impasse is to direct more resources to basic and translational research to improve our understanding of the mechanisms of illness.

## Acknowledgments

None.

## Footnote

*Conflicts of Interest:* The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## References

- Laffey JG, Kavanagh BP. Negative trials in critical care: why most research is probably wrong. *Lancet Respir Med* 2018;6:659-60.
- Iwashyna TJ, Burke JF, Sussman JB, et al. Implications of Heterogeneity of Treatment Effect for Reporting and Analysis of Randomized Trials in Critical Care. *Am J Respir Crit Care Med* 2015;192:1045-51.
- Walter SD. The estimation and interpretation of attributable risk in health research. *Biometrics* 1976;32:829-49.
- Shankar-Hari M, Harrison DA, Rowan KM, et al. Estimating attributable fraction of mortality from sepsis to inform clinical trials. *J Crit Care* 2018;45:33-9.
- Acute Respiratory Distress Syndrome Network, Brower RG, Matthay MA, et al. 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;342:1301-8.
- Simonis FD, Serpa Neto A, Binnekade JM, et al. Effect of a Low vs Intermediate Tidal Volume Strategy on Ventilator-Free Days in Intensive Care Unit Patients Without ARDS. *JAMA* 2018;320:1872.
- Bellani G, Laffey JG, Pham T, et al. Epidemiology, Patterns of Care, and Mortality for Patients With Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries. *JAMA* 2016;315:788.
- Laffey JG, Misak C, Kavanagh BP. Acute respiratory distress syndrome. *BMJ* 2017;359:j5055.
- Rubinfeld GD, Caldwell E, Granton J, et al. Interobserver variability in applying a radiographic definition for ARDS. *Chest* 1999;116:1347-53.
- Peng JM, Qian CY, Yu XY, et al. Does training improve diagnostic accuracy and inter-rater agreement in applying the Berlin radiographic definition of acute respiratory distress syndrome? A multicenter prospective study. *Crit Care* 2017;21:12.
- Dechartres A, Trinquart L, Boutron I, et al. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. *BMJ* 2013;346:f2304.
- Zhang Z, Hong Y, Liu N. Scientific evidence underlying the recommendations of critical care clinical practice guidelines: a lack of high level evidence. *Intensive Care Med* 2018;44:1189-91.
- Walkey AJ, Goligher EC, Del Sorbo L, et al. Low Tidal Volume versus Non-Volume-Limited Strategies for Patients with Acute Respiratory Distress Syndrome. A Systematic Review and Meta-Analysis. *Ann Am Thorac Soc* 2017;14:S271-9.
- 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.
- Embase search for Major Focus ‘Sepsis’ and study type



- 'Randomized controlled trial'. Available online: <https://www.embase.com> (accessed 10 September 2019).
16. Angus DC, van der Poll T. Severe Sepsis and Septic Shock. *N Engl J Med* 2013;369:840-51.
  17. McConnell KW, Coopersmith CM. Pathophysiology of septic shock: From bench to bedside. *Presse Med* 2016;45:e93-8.
  18. Wilson PG, Manji M, Neoptolemos JP. Acute pancreatitis as a model of sepsis. *J Antimicrob Chemother* 1998;41:51-63.
  19. Leentjens J, Kox M, van der Hoeven JG, et al. Immunotherapy for the Adjunctive Treatment of Sepsis: From Immunosuppression to Immunostimulation. Time for a Paradigm Change? *Am J Respir Crit Care Med* 2013;187:1287-93.
  20. Hall JB. Respiratory system mechanics in adult respiratory distress syndrome. Stretching our understanding. *Am J Respir Crit Care Med* 1998;158:1-2.

**Cite this article as:** Girbes ARJ, de Groot HJ. Time to stop randomized and large pragmatic trials for intensive care medicine syndromes: the case of sepsis and acute respiratory distress syndrome. *J Thorac Dis* 2020;12(Suppl 1):S101-S109. doi: 10.21037/jtd.2019.10.36

### Sample size calculations

The sample size calculations in the R language (using the *pwr* package by Champely *et al.*) are available at the end of this technical supplement.

Sample size required to demonstrate a 20% relative risk reduction with a 35% control-group rate, 80% power and 5% type-I error rate is 689 per group or 1,379 for a two-group trial.

For an attributable fraction of mortality of 0.25, and a relative reduction of the attributable fraction of 0.20, the absolute effect on a 35% control group mortality rate will be 1.75%:  $0.35 \times 0.25 \times 0.20 = 0.0175$ . Therefore, the expected intervention-group mortality rate would be 33.25%:  $0.35 - (0.35 \times 0.25 \times 0.20) = 0.3325$ . The required sample size to compare a control-group rate of 35% to an intervention-group rate of 33.25% is 23,042 (80% power, 5% type-I error rate).

Similarly, for an attributable fraction of mortality of 0.50 and a relative reduction of the attributable fraction of 0.20, the absolute effect on a 35% control group mortality rate will be 3.5%:  $0.35 \times 0.50 \times 0.20 = 0.035$ . Therefore, the expected intervention-group mortality rate would be 31.5%:  $0.35 - (0.35 \times 0.50 \times 0.20) = 0.315$ . The required sample size to compare a control-group rate of 35% to an intervention-group rate of 31.5% is 5,685 (80% power, 5% type-I error rate).

In example of sepsis with an attributable fraction of 0.15, at an “optimal” 50% control-group mortality rate and a relative reduction of the attributable fraction of 0.20, the absolute effect will be 1.5%:  $0.50 \times 0.15 \times 0.20 = 0.015$ . A reduction from 50% to 48.5% would require a sample size of 34,873 (80% power, 5% type-I error rate).

### Simulated estimation of syndrome-attributable risk, explaining paper Figure 3

To demonstrate how syndrome-attributable risks are overestimated when the syndrome and the underlying illness severity are correlated, we present a computer simulated example wherein the actual syndrome-attributable risk of a hypothetical syndrome is zero.

We generated a simulated cohort of 4,000 patients with a heterogeneous (randomly generated) mortality risk distribution, with a median mortality risk of 7% (IQR, 2–23%) (Figure 3A). The risk of developing a hypothetical critical illness syndrome was generated semi-randomly to correlate with baseline mortality risk ( $r=0.80$ ). Using a

random binomial generator, each patient was then assigned with a syndrome diagnosis (absent or present, based on the risk of developing a syndrome) and an outcome (dead or alive, based on the true mortality risk). Importantly, the mortality risk didn't shift depending on the presence of the syndrome so that the presence of the syndrome didn't cause an increase in mortality risk (the true attributable risk was zero) (Figure 3A).

To simulate how syndrome-attributable risks can be estimated by investigators, we constructed three scenarios (Figure 3B). In the first scenario, a hypothetical researcher makes a crude mortality comparison between those with the syndrome and those without. The researcher will observe that the crude mortality rate of patients with the syndrome is 41% versus 13% for those without it, so that the crude syndrome-attributable mortality is estimated as 28% ( $P<0.0001$ ) (Figure 3C).

In the second scenario, the researcher makes use of a severity of illness score to match each patient with the syndrome to a patient without the syndrome but with a similar severity score. The severity score is not a perfect reflection of the true mortality risk, but is well-calibrated and has an area under the receiver operating characteristics (AUROC) curve of 0.84 to discriminate between survival and death (Figure 3B). In this matched cohort, the mortality rate of the patients with the syndrome is 37% versus 26% for those without it, so that the adjusted attributable risk is estimated to be 11% ( $P=0.0006$ ) (Figure 3C).

In the final scenario, the researcher has full information and perfect characterization of the underlying diseases, so that he can make use of a perfect severity score (or any other distance metric) to match each patient with the syndrome to a patient without the syndrome. The theoretically maximum attainable AUROC for the underlying risk distribution is 0.88 (Figure 3B). Using full information about the underlying illness risks to construct a matched cohort, the mortality rate of the patients with the syndrome is found to be 30% versus 29% for those without it, so that the accurate adjusted attributable risk is estimated to be a negligible 1.1% ( $P=0.0006$ ) (Figure 3C).

This simulation shows that, conditional on an association between the risk of developing a syndrome and the underlying mortality risk, the syndrome-attributable mortality risk can only be accurately estimated when there is perfect risk information (or disease characterization) of the underlying pathological conditions.

The full code for the simulation study in the R language is available on request from h.degroot@amsterdamumc.nl.