EDITORIAL

Focus on clinical trial interpretation

Morten Hylander Møller^{1,2*}, Lennie P. G. Derde^{3,4} and Rob Mac Sweeney⁵

© 2020 Springer-Verlag GmbH Germany, part of Springer Nature

Deceased Professor Douglas Altman once said, “We need less research, better research and research done for the right reasons” [1]. We believe that critical appraisal of research is important for trustworthy research interpretation and an important skill for physicians, including intensivists. In this focus editorial, we outline key areas in clinical trial interpretation, as recently discussed in relevant scientific journals.

In a recently published meta-epidemiological study of 604 randomised clinical trials (RCTs) published between 1977 and 2018 from 53 Cochrane systematic reviews in critical care, less than 7% of the RCTs had overall low risk of bias, i.e. had high methodological quality, and the proportion did not increase over time [2]. Also, the median sample size of the trials did not seem to increase over time. As trials with lower methodological quality and higher risk of bias are associated with inflated effect estimates (Fig. 1) [3], we need to critically appraise and interpret critical care trials considering their methodological quality.

One important methodological characteristic to consider is clinical heterogeneity. It has been argued that critically ill patients constitute the most heterogeneous population in the hospital and that no two patients are the same [4]. Clustering of critically ill patients into sub-phenotypes based on clinical parameters reflecting disease mechanisms and/or treatment responses, rather than admission diagnoses has therefore been proposed [5]. In a recently published commentary, it has even been suggested that we should abandon large pragmatic trials in sepsis and acute respiratory distress syndrome (ARDS) research because of heterogeneous populations, insensitive outcomes and underpowered trials [6]. The

authors propose that we return to basic and translational research to improve the understanding of the mechanisms of illness. While we agree that we need better understanding of the mechanisms of illness, e.g. by basic and translational research, we should also continue to do large high-quality pragmatic trials. If we continue to use interventions in the ICU where the balance between benefits and harms is unknown, it is of utmost importance to assess their real-world effectiveness in unselected broad patient groups. Not doing this has the imminent risk that we continue to use ineffective, and even harmful, interventions in the ICU [7]. The growing use of adaptive and platform trials [8], as well as registry-based data collection, to improve trial power, efficiency and cost-effectiveness, places a burden on contemporary clinicians to understand the strengths and weaknesses of these increasingly frequent designs. We need both exploratory research and pragmatic trials to provide high-quality care and improve the outcome of patients in the ICU.

Clustering of critically ill patients has been suggested as a basis for assessment of heterogeneity treatment effects (HTE). HTE analyses have recently been done in patients with ARDS included in the ART trial [9] and in ICU patients at risk of gastrointestinal bleeding included in the SUP-ICU trial [10]. Using HTE analyses, both studies identified sub-phenotypes of patients who had higher and lower probability of the outcomes of interest, respectively. Predefined HTE analyses in trials could perhaps supplement (or replace) conventional subgroup analyses in the future.

Use of artificial intelligence and machine learning to diagnose and risk-stratify critically ill patients has been another subject of interest in recent years [11]. In a systematic review and meta-analysis, use of different machine learning models to diagnose sepsis was assessed [12]. The authors identified 130 different models which predicted sepsis onset ahead of time accurately; however, systematic reporting and clinical implementation studies

*Correspondence: mortenhylander@gmail.com

¹ Department of Intensive Care 4131, Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen, Denmark
Full author information is available at the end of the article

would be needed before machine learning can be used routinely in daily clinical practice.

In an interesting series of studies, the most prolific authors in ICM were characterised. In one paper, the perception of authorship and conflicts of interest among 70 prolific authors were assessed [13]. Surprisingly, 1 out of 3 responders declared to not have met all the ICMJE criteria in some of the papers that they had co-authored, and 1 out of 4 declared that they had been influenced by the industry in one or more of their publications. In another paper, a principal component analysis found that the number of publications by a prolific author was positively correlated with the H-index and the number of publications in high-impact journals, whereas working as a journal editor was negatively correlated with the number of publications [14]. Lastly, it was observed that the most popular topics among prolific critical care authors were: cardiac arrest, trauma, respiratory diseases, sepsis and ARDS [15].

In a highly interesting and thoughtful commentary in *Nature*, 3 scientists and 800 signatories called for a rise against statistical significance [16]. In a convenience sample of 791 papers in 5 journals, the authors found that around half mistakenly assumed that lack of statistical significance (a P value of less than 0.05) means that there was no effect. The authors propose that we should stop using the terms “no difference” and “no association” just because a P value is larger than a threshold such as 0.05 or, equivalently, because a confidence interval includes zero. Furthermore, they suggest we stop using P values in the conventional, dichotomous way, i.e. to decide whether a result refutes or supports a scientific hypothesis, and instead put more emphasis into clinical significance (relevance) and absolute differences and consider

using the term “compatibility intervals” instead of “confidence intervals”, as the confidence in trial findings is often low [16].

In summary, there is growing evidence suggesting that the scientific methods used in research play an important part and that they, to a larger extent, need to be critically appraised when interpreting research findings. We encourage the critical care community to take an active part in this important task.

Author details

¹ Department of Intensive Care 4131, Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen, Denmark. ² Centre for Research in Intensive Care, Copenhagen, Denmark. ³ Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands. ⁴ Intensive Care Center, University Medical Center Utrecht, Utrecht, The Netherlands. ⁵ Regional Intensive Care Unit, Royal Victoria Hospital, Belfast, Northern Ireland.

Compliance with ethical standards

Conflicts of interest

MHM is on the Editorial Board for ICM and declares no other competing interests. LPGD and RMS declare no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 6 February 2020 Accepted: 3 March 2020

Published online: 12 March 2020

References

- Altman DG (1994) The scandal of poor medical research. *BMJ* 308:283–284
- Anthon CT, Granholm A, Perner A, Laake JH, Moller MH (2019) Overall bias and sample sizes were unchanged in ICU trials over time: a meta-epidemiological study. *J Clin Epidemiol* 113:189–199
- Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J, Als-Nielsen B, Balk EM, Gluud C, Gluud LL, Ioannidis JP, Schulz KF, Beynon R, Welton NJ, Wood L, Moher D, Deeks JJ, Sterne JA (2012) Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 157:429–438
- Castela Forte J, Perner A, van der Horst ICC (2019) The use of clustering algorithms in critical care research to unravel patient heterogeneity. *Intensive Care Med* 45:1025–1028
- Seymour CW, Kennedy JN, Wang S, Chang CH, Elliott CF, Xu Z, Berry S, Clermont G, Cooper G, Gomez H, Huang DT, Kellum JA, Mi Q, Opal SM, Talisa V, van der Poll T, Visweswaran S, Vodovotz Y, Weiss JC, Yealy DM, Yende S, Angus DC (2019) Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA* 321:2003–2017
- Girbes AR, de Grooth H-J (2020) 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* 12(Suppl 1):S101–S109
- Landoni G, Comis M, Conte M, Finco G, Mucchetti M, Paternoster G, Pisano A, Ruggeri L, Alvaro G, Angelone M, Bergonzi PC, Bocchino S, Borghi G, Bove T, Buscaglia G, Cabrini L, Callegher L, Caramelli F, Colombo S, Corno L, Del Sarto P, Feltracco P, Forti A, Ganzaroli M, Greco M, Guarracino F, Lembo R, Lobreglio R, Meroni R, Monaco F, Musu M, Pala G, Pasin L, Pieri M, Pisarra S, Ponticelli G, Roasio A, Santini F, Silveti S, Szekely A, Zambon M, Zucchetti MC, Zangrillo A, Bellomo R (2015) Mortality in multicenter critical care trials: an analysis of interventions with a significant effect. *Crit Care Med* 43:1559–1568

8. Coalition APT (2019) Adaptive platform trials: definition, design, conduct and reporting considerations. *Nat Rev Drug Discov* 18:797–807
9. Zampieri FG, Costa EL, Iwashyna TJ, Carvalho CRR, Damiani LP, Taniguchi LU, Amato MBP, Cavalcanti AB, Alveolar Recruitment for Acute Respiratory Distress Syndrome Trial I (2019) Heterogeneous effects of alveolar recruitment in acute respiratory distress syndrome: a machine learning reanalysis of the Alveolar Recruitment for Acute Respiratory Distress Syndrome Trial. *Br J Anaesth* 123:88–95
10. Granholm A, Marker S, Krag M, Zampieri FG, Thorsen-Meyer HC, Kaas-Hansen BS, van der Horst ICC, Lange T, Wetterslev J, Perner A, Moller MH (2020) Heterogeneity of treatment effect of prophylactic pantoprazole in adult ICU patients: a post hoc analysis of the SUP-ICU trial. *Intensive Care Med*. <https://doi.org/10.1007/s00134-019-05903-8>
11. Zampieri FG, Salluh JIF, Azevedo LCP, Kahn JM, Damiani LP, Borges LP, Viana WN, Costa R, Correa TD, Araya DES, Maia MO, Ferez MA, Carvalho AGR, Knibel MF, Melo UO, Santino MS, Lisboa T, Caser EB, Besen B, Bozza FA, Angus DC, Soares M, Investigators OS (2019) ICU staffing feature phenotypes and their relationship with patients' outcomes: an unsupervised machine learning analysis. *Intensive Care Med* 45:1599–1607
12. Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, Swart EL, Girbes ARJ, Thorat P, Ercole A, Hoogendoorn M, Elbers PWG, (2020) Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med* Jan 21 [Epub ahead of print]
13. Hjortrup PB, Robba C, Weiss E, De Jong A, Helms J (2019) Are critical care authors publication dealers? *Intensive Care Med* 45:1667–1669
14. Weiss E, Hjortrup PB, Robba C, De Jong A, Helms J (2019) Highly prolific authors in critical care: which factors influence their scientific output? *Intensive Care Med* 45:1673–1675
15. Robba C, Weiss E, Hjortrup PB, De Jong A, Helms J (2019) Who are these highly prolific authors in critical care? *Intensive Care Med* 45:1670–1672
16. Amrhein V, Greenland S, McShane B (2019) Scientists rise up against statistical significance. *Nature* 567:305–307