Real-world causal evidence for planned predictive enrichment in critical care trials
A scoping review

Kaas-Hansen, Benjamin Skov; Granholm, Anders; Sivapalan, Praleene; Anthon, Carl Thomas; Schjørring, Olav Lilleholt; Maagaard, Mathias; Kjaer, Maj-Brit Nørregaard; Mølgaard, Jesper; Ellekjaer, Karen Louise; Fagerberg, Steen Kåre; Lange, Theis; Møller, Morten Hylander; Perner, Anders

Published in:
Acta Anaesthesiologica Scandinavica

DOI:
10.1111/aas.14321

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Real-world causal evidence for planned predictive enrichment in critical care trials: A scoping review

Benjamin Skov Kaas-Hansen1,2 | Anders Granholm1 | Praleene Sivapalan1 | Carl Thomas Anthon1 | Olav Lillesholt Schjørring3,4 | Mathias Maagaard5 | Maj-Brit Nørregaard Kjær1 | Jesper Mølgaard6 | Karen Louise Ellekjær1 | Steen Kåre Fagerberg3 | Theis Lange2 | Morten Hylander Møller1 | Anders Perner1

1Department of Intensive Care, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
2Section of Biostatistics, Department of Public Health, University of Copenhagen, Copenhagen, Denmark
3Department of Anaesthesia and Intensive Care, Aalborg University Hospital, Aalborg, Denmark
4Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
5Centre for Anaesthesiological Research, Department of Anaesthesiology, Zealand University Hospital, Køge, Denmark
6Department of Anesthesiology, Centre for Cancer and Organ Dysfunction, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark

Correspondence
Benjamin Skov Kaas-Hansen, Department of Intensive Care, Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen E, Denmark.
Email: benjamin.skov.caas-hansen@regionh.dk; epiben@hey.com

Funding information
Dagmar Marshalls Fond; Grosserer Jakob Ehrenreich og Hustru Grete Ehrenreichs Fond; Sygeforsikringen “danmark”

Abstract
Background: Randomised clinical trials in critical care are prone to inconclusiveness due, in part, to undue optimism about effect sizes and suboptimal accounting for heterogeneous treatment effects. Although causal evidence from rich real-world critical care can help overcome these challenges by informing predictive enrichment, no overview exists.

Methods: We conducted a scoping review, systematically searching 10 general and speciality journals for reports published on or after 1 January 2018, of randomised clinical trials enrolling adult critically ill patients. We collected trial metadata on 22 variables including recruitment period, intervention type and early stopping (including reasons) as well as data on the use of causal evidence from secondary data for planned predictive enrichment.

Results: We screened 9020 records and included 316 unique RCTs with a total of 268,563 randomised participants. One hundred seventy-three (55%) trials tested drug interventions, 101 (32%) management strategies and 42 (13%) devices. The median duration of enrolment was 2.2 (IQR: 1.3–3.4) years, and 83% of trials randomised less than 1000 participants. Thirty-six trials (11%) were restricted to COVID-19 patients. Of the 55 (17%) trials that stopped early, 23 (42%) used predefined rules; futility, slow enrolment and safety concerns were the commonest stopping reasons. None of the included RCTs had used causal evidence from secondary data for planned predictive enrichment.

Conclusion: Work is needed to harness the rich multiverse of critical care data and establish its utility in critical care RCTs. Such work will likely need to leverage methodology from interventional and analytical epidemiology as well as data science.

KEYWORDS
causal inference, enrichment, intensive care, real-world data, real-world evidence
1 | INTRODUCTION

Undue optimism about effect sizes and suboptimal accounting for heterogeneous treatment effects make randomised clinical trials (RCTs) in critical care prone to inconclusiveness. Prudent enrichment is recommended to help alleviate these key challenges by maximising trial efficiency and can be planned (while designing the trial) or adaptive (when the trial is underway). Enrichment can be defined as, “the prospective use of any patient characteristic to select a study population in which detection of a drug effect (if one is in fact present) is more likely than it would be in an unselected population”. Both planned and adaptive enrichment can be predictive and prognostic: whereas prognostic enrichment seeks to enrol participants prone to experience the primary trial outcome, predictive enrichment seeks to enrol “likely responders” (i.e., participants prone to respond more strongly to the intervention under study).

As for any adaptive trial feature, adaptive enrichment uses data collected during the trial whereas planned enrichment relies on data available in the trial build-up phase, be they from other trials or observational in nature. The critical care setting enjoys a rich multiverse of diverse data making it the natural habitat for machine learning. This richness potentially makes observational critical care data a strong base for planned predictive enrichment in RCTs if combined with modern causal-inference methods from aetiological epidemiology so as to underpin causal (if cautious) interpretations.

Indeed, multiple methods exist for estimating causal effects in observational data, including propensity scores, sufficient covariate sets derived from directed acyclic graphs (DAGs), causal discovery, and emulated target trials with clone-censoring. Despite their existence and firm methodological bearings, with some specifically targeting heterogeneous treatment effects—little is known about their use in RCTs in the critical care setting.

Thus, we carried out the scoping review reported herein, to gauge the extent and nature of the use of causal evidence from secondary data for planned predictive enrichment of RCTs in critical care.

2 | MATERIALS AND METHODS

This scoping review was undertaken in the context of the Intensive Care Platform Trial (INCEPT; www.incept.dk) research programme and reported as per the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR), see supplement.

Editorial Comment

In this scoping review, evidence is explored for how analysis of secondary data can contribute to understanding of heterogeneous treatment effects in critical care treatment trials. The authors describe characteristics of critical care trials that will be important in developing more robust analysis methods to assess causal relationships where many factors can be relevant.

Specifically, we set out to answer the following five questions:

1. How many recent critical care RCTs used causal evidence from secondary data for planned predictive enrichment?
2. What causal-inference methodologies did these RCTs use?
3. What data were used?
4. What data models were used?
5. To what extent were publicly available data used?

The protocol was published prior to data extraction and analysis.

2.1 | Search strategy

We used and updated the search and data extraction of a recent scoping review on patient-important outcomes in critically-ill patients (henceforth, the previous review) in RCTs in critical care whose results were published online or in-print on or after 1 January 2018 in 10 general and specialised journals. The start date of the query was arbitrary but chosen to yield a sufficiently large convenience sample of well-conducted and contemporary RCTs. The supplement contains the full search string including the Cochrane Collaboration’s highly sensitive filter for randomised clinical trials.

On 6 February 2023, we used the search string in the supplement to search PubMed from 5 May 2022 onward (the date of the final search in the previous review). We created a separate Covidence instance for title/abstract and full-text screening of this updated search with the same options as in the previous review, except for the ‘Wrong outcomes’ option for full-text exclusion. All articles were assessed independently in duplicate, resolving disagreements by consensus and involving the last author (AP) when needed. Principal reports of relevant RCTs identified through secondary studies were included if they had not already been in the previous review.

2.2 | Eligibility

In this scoping review, we included RCTs enrolling mainly patients receiving critical care-specific treatment in intensive care units (ICUs) or similar high-dependency units (e.g., mechanical
ventilation, continuous use of vasopressors/inotropes, continuous renal replacement therapy, extracorporeal membrane oxygenation). Conventional individually randomised trials as well as quasi- and cluster-randomised trials were eligible. We based the eligibility of RCTs enrolling non-ICU patients on their descriptive summary statistics to establish whether the majority (>50%) of patients could reasonably be assumed critically ill, as done previously.30

We excluded trials with patient-level cross-over designs as well as trials whose interventions primarily happened outside the ICU, such as perioperative interventions in patients subsequently admitted to the ICU or after discharge from the ICU, even if participants were randomised while in the ICU.31

2.3 | Data source, extraction and synthesis

When extracting data for a given trial, we used up to five sources:

1. The principal report.
2. The statistical analysis plan (when published or appended to the principal report).
3. The protocol (when published or appended to the principal report).
4. Record(s) in clinical trial registers such as ClinicalTrials.gov when available (mainly to obtain trial phases and study periods when unavailable in the principal reports).
5. The response from the corresponding author regarding the use of causal evidence from secondary data for planned predictive enrichment (corresponding authors of all 316 trials were contacted as detailed below). Each corresponding author was contacted via email; a second enquiry was sent to those who had not responded within 2 weeks.

We built, hosted and used our own data extraction form and database to enable dynamic standardised data entry (full source code available at https://doi.org/10.5281/zenodo.8068824; screenshots included in the supplement). Data extraction had two phases, in both of which all data were extracted independently and in duplicate, resolving disagreements by consensus and involving AP when needed (BSKH extracted data on all included trials; so did AG, CTA, JM, KLE, MM, MNK, OLS, PS, SKF and TL collectively). In the pilot phase, BSKH and AG extracted data from 10 randomly selected trials,35–42 after which the extraction form was optimised and full extraction from the remaining 306 trials began.

Categorical variables are summarised in table form as counts and percentages with one significant digit; continuous variables as medians (inter-quartile ranges, IQRs) and graphically with empirical cumulative distribution plots. Falling outside the scope of this review type, we forwent risk-of-bias and certainty-of-evidence assessment,28 as well as null-hypothesis significance tests.

The final data set, supporting documentation and analytic code are available at https://doi.org/10.5281/zenodo.8068824. Trials are denoted “[trial acronym/first author’s last name] [year of publication of principal report]” throughout.

2.4 | Deviations from the protocol and post-hoc design decisions

There were three deviations. First, after having extracted data from approximately 80 trials, examples started to emerge of predictive enrichment although not based on secondary-data causal evidence. This gave rise to a new variable Any predictive enrichment. Because the data extracted for this variable proved difficult to discern and would require more elaborate data extraction to yield meaningful insights (see the third paragraph under Results), we stopped extracting data, at which point we had extracted from 246 trials. Second, we used number of randomised participants and not number of enrolled participants as this was the operationalisation we had in mind and aligns with the approach in the previous review.31 Third, inter-rater agreement was forgone as all data sources were charted twice, by two different extractors, and all disagreements were resolved.

We made three post hoc design decisions. First, phase of trial was considered to apply to drug trials only. Second, stopping of adaptive trials without pre-specified sample sizes was considered early stopping; so was reaching sample sizes that had been changed after trial initiation.33,44 Third, the BOX and REMAP-CAP trials were included three45–47 and six48–53 times, respectively, because they yielded the same numbers of principal reports with distinct foci.

3 | RESULTS

We screened 9020 records yielded by the searches and included 316 (Figure 1) unique trials, reported in the 315 publications listed in the supplement, with a total of 268,563 randomised participants (Figure 1). We included all 167 RCTs from the previous review.31 Further, the 101 studies excluded on the grounds of not including so-called patient-important outcomes were still from RCTs eligible for this review. Thus, upon screening these studies anew, we identified 89 principal trial reports and secondary studies of eight additional trials whose principal reports were included, for a total of 264 trials carried over from the previous review.

3.1 | Trial characteristics

Table 1 and Figure 2 summarise the trial metadata. The coverage of the years from 2018 onward was even; 173 (55%) trials tested drug interventions while 101 (32%) tested management strategies and 42 (13%) devices. The median recruitment span was 2.2 (IQR: 1.3–3.4) years, the median number of randomised participants was 253 (IQR: 110–710; 83% of trials randomised less than 1000 participants), and the median number of sites was 10 (IQR: 2–26). 122 (71%) drug trials were in phases 3 and/or 4 (wholly or in part), and 18 (10%) spanned two phases. 36 (11%) trials were restricted to COVID-19 patients. Of the 55 (17%) trials that stopped early, 23 (42%) used predefined rules; futility, slow enrolment and safety concerns were the most common stopping reasons.
3.2 | Secondary-data causal evidence for predictive enrichment

The published sources (points 1. through 4. above) for none of the 316 included trials suggested the use of causal evidence from secondary data for planned predictive enrichment. This was confirmed following contact with the corresponding authors (40% responded). For a single trial, ARISE 2014, observational data (including routine data) were used to identify a population likely to benefit from both interventions under study; none of these observational studies, however, employed causal-inference methods.

Thirty-eight trials (15% of 246 trials, see ‘Deviations from protocol’ above) used some kind of predictive enrichment, and despite some clear-cut examples, most were not. REMAP-CAP, for example, excluded anticoagulated patients from domains studying anticoagulation and immunocompromised patients from the steroid domain.

3.3 | Enrolment over time

Figure 3 shows the number of included trials actively enrolling by month from August 1997, when LaSRS 2006 started enrolment, until November 2022, when Postle 2022 stopped enrolment. At the peak in September and October 2017, 121 (38%) of the included trials were enrolling participants. Figure 3 also suggests a satisfactory coverage of the search for contemporary trials in that 89% of the included trials started enrolment on or after January 2012.

4 | DISCUSSION

In this scoping review, none of the 316 included contemporary critical care RCTs had used causal evidence from secondary data for planned predictive enrichment. Although our interest revolved mainly around the use of observational, routine data (increasingly referred to as real-world data) we employed a wider definition of secondary data to also comprise secondary analyses of trial data (whether they respected the original randomisation or not). Eligibility criteria always serve to narrow down the study population to the target population of interest leaving a somewhat ill-defined distinction between eligibility criteria (e.g., to establish clean exposure) and enrichment. Often, principal reports did not detail whether patients were excluded if chronically treated with a given drug (e.g., anticoagulants) or only in the period leading up to the index ICU admission. Further, some trials used eligibility criteria that could be considered both predictive and prognostic enrichment. One example was EUHPRATES 2018 testing polymyxin B haemoperfusion: enrolment was conditional on an endotoxin activity level of at least 0.6, which is also considered associated with elevated ICU mortality. Another example is Ho et al. testing a retrievable vena cava filter in severely injured patients in whom anticoagulation was contraindicated; these patients were both more likely to benefit from the treatment and to experience the primary
4.1 | Subjunctive mood

Insights from prior RCTs may come in two forms. If, for example, a phase 3 trial is meant to confirm a positive phase 2 or feasibility trial, the case for enrichment is arguably thin: one assumes an overall effect in the full population, and so the concern is one of reproducibility rather than increasing trial efficacy. In this case, only rare diseases would really be an argument for enrichment. Many trials conduct subgroup and/or post hoc analyses, but these are almost always underpowered because they are designed considering the primary endpoint in the eligible population. In neutral and in conclusive trials, there seems to be a great interest in identifying subgroups who might benefit from the intervention and, then, undertake a follow-up trial seeking to verify this. DESIRE, for example, found no statistically significant effect of sedation with dexmedetomidine, against sedation without, in sepsis patients requiring mechanical ventilation, despite promising results of a subgroup analysis. Also, Doig was a phase 2 trial testing the effect on kidney function of a short-term daily intravenous standard amino acid supplement against standard care in critically ill. Although Doig mentions a hypothesis-generating subgroup analysis indicating that higher daily protein intake be associated with lower risk of requiring renal replacement therapy, in patients with elevated risk of renal dysfunction at baseline, these results were not used to enrich the trial population in Doig.

Thus, even though the evidence underlying the eligibility criteria technically arises from prior trials in both scenarios above, the composite outcome (pulmonary embolism or all-cause mortality 90 days after randomisation).
FIGURE 2  Empirical cumulative distributions of four continuous metadata variables. The horizontal axes in the lower-panel sub-plots are log10-scaled. Empirical cumulative density plots use neither smoothing (as density plots do) nor binning (as histograms do), avoid overplotting and allow for reading arbitrary quantiles (e.g., the quartiles shown with dashed lines) from the plot.

FIGURE 3  Number of actively recruiting trials over time (monthly resolution). The secondary horizontal axis at the top shows the cumulative proportion of included trials (when they started recruiting). For example, 121 trials were recruiting in September and October 2017, and 45% of the included trials had started recruiting by 2016 (and some had also stopped).
follow-up trials have distinct goals and rest on disparate foundations. Although we did not systematically gauge this in our review, it is the impression that the results in trials with predictive enrichment from subgroup analyses in previous inconclusive or neutral pilot trials often contradict those very subgroup-analysis results. Consequently, incorporating causal evidence from secondary data could bolster the foundation of such trials.

Likely owing to the scoping nature of this review, a number of trial types emerged for which predictive enrichment is less, if at all, relevant. First, non-inferiority designs seek to make direct comparisons with effect-size estimates in given populations and substantiate that the intervention(s) under study are, at least, not worse than, for example, current standard of care. Similar to confirmatory trials (mentioned above), the case for predictive enrichment is thin as it would defeat the purpose of the trial. Second, predictive enrichment using secondary data in phase 2 trials may, at first, seem awkward as these trials seek to establish efficacy and safety of new treatments. Nevertheless, few medicinal interventions are truly novel, and class effects might exist that would render routine data useful to, at least indirectly, hint at populations of likely beneficiaries. Further, when phase 2 trials test well-known and widely-used drugs (e.g., fentanyl vs. morphine in ANALGESIC and hypertonic versus normal saline in HERACLES), secondary data causal evidence could be relevant.

Estimating causal effects from real-world data has grown in popularity and several prominent examples have emerged in recent years such as RCT-DUPLICATE and OHDSI LEGEND. Indeed, eliciting causal evidence from real-world critical care data could be a low-hanging fruit for formalising foundational design decisions of RCTs: estimating realistic effect sizes and outcome distributions/event probabilities (crucial to realistic sample-size estimation) and identifying clinically pertinent subgroups of likely responders (important to trial efficiency). Considering the success of reproducing RCT results with emulated trials in other clinical domains, even if such attempts are undertaken after the target RCTs have already been completed and reported, it may be reasonable to expect similar positive results in critical care settings with richer data of much greater temporal and phenotypical granularity. However, to the best of our knowledge, this line of research is yet to be pursued in the domain of critical care.

Our results suggest very limited, if any, uptake of causal-inference methods in secondary data in critical care trials, the reason for which probably has several components. First, the division between aetiological and interventional epidemiology is considerable. Thus, although causal inference is not a new discipline, trialists may remain unexposed to methodological advances in the field, leaving the synergistic potential unexplored and unrealised. Second, enrolment in about half the included trials started before May 2016 (Figure 2), and given that multi-year inception periods are common, the included trials may simply have started too early to fully exploit these advances in causal inference. Third, a potential and (in our view) important reason may be inadequate access to fit-for-purpose data in harmonised formats: publicly available data on non-North American ICU patients have only recently started to emerge, and all datasets use idiosyncratic data models that complicate their joint use for generating causal evidence (as well as developing and calibrating prediction models for prognostic enrichment). Converging on shared data formats would likely facilitate the generation of real-world causal evidence.

4.2 Strengths and limitations

This scoping review has several strengths, including adherence to standards for scoping review; prior publication of the protocol with few, minor deviations (all reported and motivated); a large convenience sample; and satisfactory coverage of contemporary critical care trials (with almost 90% of the trials starting enrolment in the past decade).

Our review has three main limitations. First, we restricted the search to four general and six specialised journals. Thus, although our search strategy may have missed some trials, we believe most well-conducted, contemporary RCTs in critical care were included as these journals are major outlets of critical care RCTs. This notwithstanding, our findings may not apply to critical care RCTs published in smaller or niche journals. Second, we focused on planned predictive enrichment and did not extract information on the use of prognostic enrichment. Although we found no examples of causal evidence from secondary data for planned predictive enrichment, we believe this was a prudent choice: these two enrichment schemes are disparate and require different methodological considerations, and focusing on predictive enrichment better underpin a concerted discussion and identification of signposts for interesting avenues to explore at this nexus between interventional and analytical epidemiology. We did not extract data on it, but the impression after having extracted data from all included trials is that prognostic enrichment is relatively common and might enjoy a stronger evidence base than predictive enrichment; after all, it will usually rely on some kind of scoring mechanism such as prediction models of mortality risk. Further, a subsequent scoping review (with a broader scope that includes prognostic enrichment) could reuse some of the legwork done for this review. Third, the relatively recent advances in modern causal inference and their availability in user-friendly software may simply have precluded uptake in critical care trials launched early enough to be eligible for this review.

5 Conclusion

In this scoping review, none of the 316 included contemporary critical care RCTs had used causal evidence from secondary data for planned predictive enrichment. Methodological work is needed to harness the rich multiverse of critical care data and establish its utility for critical care RCTs. Such work will likely need to leverage methodological advances in interventional and analytical epidemiology as well as data science.

Author Contributions

Conceptualisation: Benjamin Skov Kaas-Hansen. Study design: all authors. Data extraction: Anders Granholm, Benjamin Skov
The outlined study was conducted as part of the Intensive Care Platform Trial (INCEPT; www.incept.dk) research programme, which has received funding from Sygeforsikringen ‘danmark’, Grosserer Jakob Ehrenreich og Hustru Grete Ehrenreichs Fond and Dagmar Marshalls Fond. All authors declare no personal conflicts of interest.

The data and analytical code that support the findings of this study are openly available on Zenodo at https://doi.org/10.5281/zenodo.8068824.

DATA AVAILABILITY STATEMENT

The data and analytical code that support the findings of this study are openly available on Zenodo at https://doi.org/10.5281/zenodo.8068824.

REFERENCES


SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.