


REVIEW

Open Access



Prognosticating the outcome of intensive care in older patients—a narrative review

Michael Beil¹, Rui Moreno^{2,3,4}, Jakub Fronczek⁵, Yuri Kogan⁶, Rui Paulo Jorge Moreno⁷, Hans Flaatten⁸, Bertrand Guidet⁹, Dylan de Lange¹⁰, Susannah Leaver¹¹, Akiva Nachshon¹², Peter Vernon van Heerden¹², Leo Joskowicz¹³, Sigal Sviri¹, Christian Jung^{14*}  and Wojciech Szczeklik⁵

Abstract

Prognosis determines major decisions regarding treatment for critically ill patients. Statistical models have been developed to predict the probability of survival and other outcomes of intensive care. Although they were trained on the characteristics of large patient cohorts, they often do not represent very old patients (age ≥ 80 years) appropriately. Moreover, the heterogeneity within this particular group impairs the utility of statistical predictions for informing decision-making in very old individuals. In addition to these methodological problems, the diversity of cultural attitudes, available resources as well as variations of legal and professional norms limit the generalisability of prediction models, especially in patients with complex multi-morbidity and pre-existing functional impairments. Thus, current approaches to prognosticating outcomes in very old patients are imperfect and can generate substantial uncertainty about optimal trajectories of critical care in the individual. This article presents the state of the art and new approaches to predicting outcomes of intensive care for these patients. Special emphasis has been given to the integration of predictions into the decision-making for individual patients. This requires quantification of prognostic uncertainty and a careful alignment of decisions with the preferences of patients, who might prioritise functional outcomes over survival. Since the performance of outcome predictions for the individual patient may improve over time, time-limited trials in intensive care may be an appropriate way to increase the confidence in decisions about life-sustaining treatment.

Keywords Intensive care, Critical care, Prediction, Very old patients

*Correspondence:

Christian Jung

christian.jung@med.uni-duesseldorf.de

¹ Department of Medical Intensive Care, Hadassah Medical Center and Faculty of Medicine, Hebrew University of Jerusalem, Jerusalem, Israel

² Unidade Local de Saúde São José, Hospital de São José, Lisbon, Portugal

³ Centro Clínico Académico de Lisboa, Lisbon, Portugal

⁴ Faculdade de Ciências da Saúde, Universidade da Beira Interior, Covilhã, Portugal

⁵ Center for Intensive Care and Perioperative Medicine, Jagiellonian University Medical College, Krakow, Poland

⁶ Institute for Medical Biomathematics, Bene Ataroth, Israel

⁷ Imperial College Business School, London, UK

⁸ Department of Research and Development, Haukeland University Hospital, Bergen, Norway

⁹ INSERM, Institut Pierre Louis d'Epidémiologie Et de Santé Publique, AP-HP, Hôpital Saint Antoine, Sorbonne Université, Service MIR, Paris, France

¹⁰ Department of Intensive Care Medicine, University Medical Center, University Utrecht, Utrecht, The Netherlands

¹¹ General Intensive Care, St George's University Hospitals NHS Foundation Trust, London, UK

¹² General Intensive Care Unit, Department of Anaesthesiology, Critical Care and Pain Medicine, Faculty of Medicine, Hebrew University and, Hadassah University Medical Center, Jerusalem, Israel

¹³ School of Computer Science and Engineering and Center for Computational Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel

¹⁴ Department of Cardiology, Pulmonology and Vascular Medicine, Faculty of Medicine, Heinrich-Heine-University, University Duesseldorf, Moorenstraße 5, 40225 Düsseldorf, Germany

Introduction

Identifying the patients with critical illnesses who will benefit from intensive care remains a challenge. Human judgement is imperfect [1, 2] and mistakes are made either by withholding or withdrawing intensive care in patients who might benefit from life-sustaining treatment (LST) or by exposing others to non-beneficial interventions. This problem is documented by the survival rates in patients with conditions deemed as unsurvivable and the variability of decisions to withdraw LST, resulting in ethical controversies [3–5]. Statistical models have been developed with empirical datasets to assist clinicians with prognostication to varying degrees of success [6]. These shortcomings are reflected by the growing interest in time-limited trials in the intensive care unit (ICU) to manage prognostic uncertainty in the individual patient [7].

Very old patients (chronological age ≥ 80 years) pose a particularly challenging problem for prognostication [8–11]. The age-related decline of physiological and cognitive reserves proceeds at different rates [12]. Its effect on the presentation and severity of acute disorders is significant but hard to measure [13]. Chronic co-morbidities and multi-morbidity, which are highly prevalent in old age, and the reduced resilience to acute stress (frailty) may influence the trajectory of critical conditions in various ways, before admission, during the stay in ICU and after discharge [14, 15]. The resulting heterogeneity within the group of very old patients compromises the performance of the current armamentarium for prognosticating the outcomes of intensive care [16, 17].

Whereas surviving a critical illness is the most crucial goal in younger patients, many very old patients value functional abilities and quality of life after discharge more than physical survival. Thus, younger and older individuals' views and expectations about intensive care may differ. Since the use of advanced directives is limited in many countries and the performance of tools for prognostications in these domains remains suboptimal [18], the ultimate prediction of ICU outcomes and the alignment of the subsequent decisions about LST with the patient's preferences are mostly left to the discretion of healthcare professionals or surrogate decision-makers. Their cultural and individual preferences can lead to substantial biases and misalignments of decisions regarding the patients' wishes [19–22]. As a result, societal norms may become self-fulfilling prophecies for ICU outcomes, which might be perpetuated in future prediction models trained on today's data.

Poor prognostication can eventually lead to inappropriate care for the individual patient and inefficient utilisation of ICU assets [23–25]. Substantial resources are being invested to improve predictive modelling and,

hence, decision-making about the level of LST that is best for the individual [26–28]. This article will discuss both classical and new approaches addressing these challenges.

Predictive modelling

The classical approach to building instruments for predicting (prognosticating) future events or states (outcome), such as survival or functional independence at a specific time, is fitting regression models to distributions of patient characteristics before these events, such as the severity of the critical condition, and the outcome of interest. Logistic regression is used when the outcome is dichotomous, such as survival vs non-survival.

There are only a small number of studies which focused on predictive modelling in ICU patients aged 80 years or older. Screening of Pubmed (www.pubmed.gov) for articles with a special emphasis on this particular cohort using the query “(older[ti] OR “very old”[ti] OR “oldest old”[ti] OR elderly[ti]) AND patient* AND (“intensive care” OR “critical care”) AND (outcome* OR survival OR mortality) AND (prognos* OR predict*) AND model*” identified five original studies published in the past 10 years [29–33], after the exclusion of studies using databases and those involving disease-specific subgroups or published by our own group.

The performance of predictive models is assessed by quantifying calibration and discrimination in different cohorts. Calibration is the measure of the statistical agreement of model predictions with observed outcomes. For example, if 80 out of 100 patients survive in a particular group, a perfectly calibrated model generates a mortality prediction of 80% for that group. Figure 1 demonstrates this relationship as a calibration plot for the hospital outcome cohort of the Simplified Acute Physiology Score (SAPS) 3 study which included 16 784 patients from 303 ICUs around the globe [34, 35]. Figure 2 illustrates the effect of age on calibration in a simple mortality model using the severity of organ dysfunction to predict ICU survival. A cohort of 24 489 adult patients with available sequential organ failure assessment (SOFA) scores was extracted from the eICU database [37]. ICU mortality was calculated for each SOFA score value in the younger training cohort and compared to that observed in a cohort of 5252 patients aged 80 years or older using the software R (version 4.1.1, www.r-project.org). The models trained on younger cohorts underestimate mortality for very old patients across parts or the whole of the spectrum of predictions (Fig. 2). Since there is no gold standard to evaluate calibration [38], the degree of miscalibration deemed acceptable is, to some degree, subjective and context-dependent. Clinicians should check

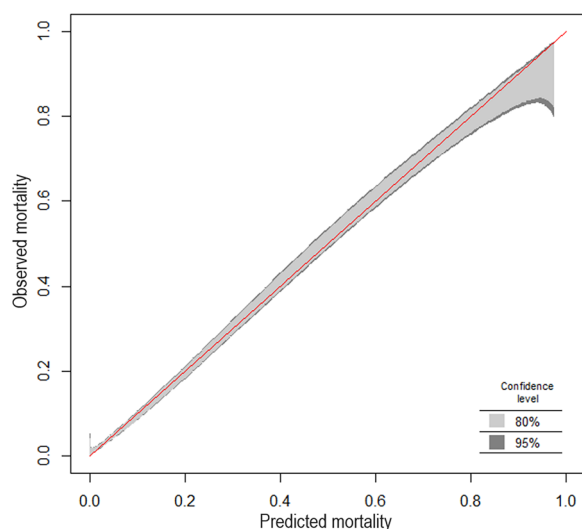


Fig. 1 Calibration belt for 16 784 patients from the SAPS 3 database [34, 35] describing the relationship between predicted and observed hospital mortality, which is depicted as the fraction of non-survivors. The plot was generated using the R package *givitiR* [36]. The overlay between the identity function (red line) and the calibration belt with narrow confidence intervals (shaded areas—see inset) indicates a good agreement between model predictions based on the SAPS 3 score and observed outcomes for most of the range of hospital mortality

whether the prediction model is well calibrated for the particular case-mix in their ICU.

Discrimination describes the model's ability to differentiate between groups of patients with distinct outcomes. This concept is crucial for assessing the merit of models in clinical practice, when outcomes are fundamentally different, such as survival and non-survival, and decisions are irreversible. Overall discrimination can be measured using the area under the curve of the receiver operating characteristic (AUC-ROC), which relates the sensitivity and specificity of a dichotomous prediction model. It ranges from 0.5, indicating discrimination no better than the random results after tossing a coin, to 1.0 for perfect discrimination. An AUC-ROC value of 0.9 implies that in 90% of the cases, the model correctly ranks a patient with a particular event, such as survival, higher than the one without. Figure 3 depicts AUC-ROCs for simulated distributions of ICU survivors and non-survivors having variable degrees of pre-existing frailty [39]. The values for these AUC-ROCs range from 0.69, which represents sub-optimal discrimination, to 0.9 which indicates very good discrimination. Despite good AUC-ROC values though, there can be a marked overlap of the distributions as illustrated by the simulated examples in Fig. 3. This complicates prognostication for those patients with a degree of frailty that is contained within the overlap. Thus, the discrimination of a predictive model should also be

verified for the range of values that is relevant to a particular patient group.

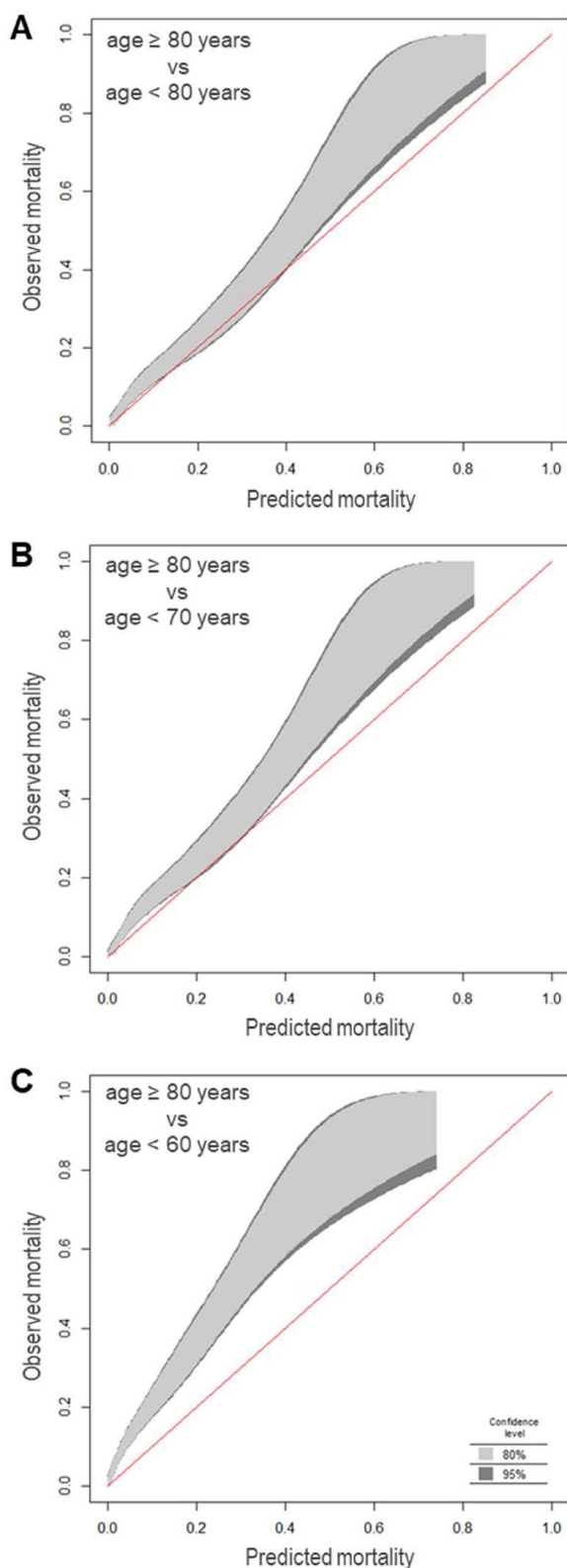
The generalisability of prediction models refers to their performance in patient groups or healthcare settings other than the one used for training. Variations of patient biographies according to geography and era as well as differences in cultural norms and practice patterns affect care processes and outcomes [4, 42–44], especially in very old patients with complex conditions. Thus, the generalisability of many prediction models is limited [45, 46] and the performance of these models was shown to degrade when deployed in different healthcare settings or during different periods in time [47, 48].

Disease severity scores

Disease severity scores, such as the Acute Physiology and Chronic Health Evaluation (APACHE) score or the Simplified Acute Physiology Score (SAPS), were developed to predict survival at discharge from the hospital. The underlying models are based on three types of variables: (1) chronological age and chronic health status including co-morbidities, (2) circumstances of ICU admission, e.g. elective vs acute, and admission diagnosis and (3) markers of organ dysfunction observed at admission to the ICU or within 24 h thereafter. A significant increase in the risk of death in patients older than 40 years in comparison to younger cohorts, even when controlling for other demographic and physiological variables, was observed while developing disease severity scores [34]. This is reflected by the points assigned to different age groups for the SAPS 3 score (Fig. 4).

In addition to scoring disease severity with data obtained at a single point in time, this method can also be applied in a dynamic way that includes the physiological response to interventions assessed at regular intervals in the ICU [49]. This approach has demonstrated the potential to outperform conventional scores with respect to the discrimination of survivors vs non-survivors [50]. Daily recordings of the SOFA score [51] allow the computation of time-dependent measures, such as the delta SOFA score, which quantifies the evolution of the critical illness in response to interventions and can improve outcome predictions [52].

Chronic conditions and impairments, which determine the health status in most very old patients, are established predictors of ICU survival [53, 54]. They were shown to have an increasing statistical impact on in-hospital mortality the longer patients are in intensive care [55] and a profound influence on survival at 6 months in very old patients [29]. However, frequent geriatric conditions, such as neurocognitive disorders, are still not taken into account by the classical disease severity scores. Moreover, the chronic health status in



◀ **Fig. 2** Illustration of the impact of chronological age on the calibration of prediction models. ICU mortality is depicted as the fraction of non-survivors and was predicted based on SOFA scores in patients from the eICU database [37]. The plots were generated using the R package givitiR [36]. The shaded regions depict calibration belts for various confidence levels (see inset). Panels **A–C** show the calibration characteristics for models trained on patients younger than the indicated threshold [**A**: ≤ 60 years (n = 7 928), **B**: ≤ 70 years (n = 13 612), **C**: ≤ 80 years (n = 19 237)] and applied to patients aged 80 years or older (n = 5 252). The extent and shape of the calibration belts' deviation from perfect calibration (red line) depend on the age gap between the training and deployment samples

very old patients should ideally be evaluated by a comprehensive geriatric assessment to capture the distinct risk profile in the individual [56, 57]. This includes additional dimensions, such as socio-economic features, which were shown to affect outcomes of critical conditions in this group [58]. Due to the time constraints in managing critically ill patients, this approach may only be feasible in the workup for planned ICU admissions in the context of elective surgery. However, it would be important to adapt fundamental aspects of a geriatric assessment to emergency care in the ICU as recently outlined by Jacobs et al. [59].

The overall contribution of chronic conditions and chronological age to the prediction of survival with disease severity scores increased in the past decades [35]. In contrast, the explanatory power of markers of acute organ dysfunction is decreasing. This can be related to the progress of intensive care in managing acute organ dysfunction. Yet, the ability to deal with the consequences of chronic conditions and impairments which progress with ageing and declining physiological reserves, including reduced cardiovascular fitness, immunosenescence and the impact of degenerative diseases, is still very limited. Importantly, the performance of disease severity scores is affected by ongoing demographic shifts, with today's very old patients not being adequately represented in historical samples for training models, and the fact that some conditions, which were imminently fatal in the past, are now managed as chronic conditions over an extended period. On this background, the discriminatory performance of the APACHE IV and SAPS II scores was found to be worse for very old patients [16, 60]. In addition to the above issues, the APACHE score requires choosing a principal admission diagnosis, which can lead to ambiguity in older patients with complex multimorbidity and further impair the model's performance.

There is an ongoing discussion about replacing chronological age with biological age to enhance the predictive

accuracy of disease severity scores in very old patients [61]. However, this requires a robust definition of biological age. Even if such a definition would focus on frailty as a marker of age-related vulnerability to stress [13], a comprehensive assessment of the individual patient beyond a simple screening might be necessary to produce a reliable and meaningful impact on the performance of prediction models. Similar considerations may apply to other geriatric conditions, notably multimorbidity [62]. Some studies have already combined measures of acute organ dysfunction with frailty screening and showed promising results for the predictive discrimination of short-term survival in sub-groups of very old ICU patients [63, 64]. In particular, phenotyping of patients 80 years or older based on a combination of SOFA scores and the clinical frailty scale (CFS) at ICU admission resulted in seven distinct sub-groups with different mortality at 30 days, ranging from 3 to 57%. Importantly, the phenotype representing mostly nonagenarians with a median CFS level of 4 (very mild frailty) and a median SOFA score of 4 had a mortality of only 7% at 30 days after ICU admission, in contrast to less old patients with more advanced frailty and similar SOFA scores [64].

Prediction of functional outcome and quality of life

Survival is an essential milestone of intensive care. However, recovering the pre-existing level of independence in daily life and self-perceived quality of life are other major goals [65]. They are especially important for very old ICU survivors with frailty, which confers an increased vulnerability to stress with potentially long-term sequelae for functional abilities. Few studies attempted to predict functional outcomes or quality of life for old individuals, since such investigations usually involve comprehensive assessments by geriatric teams [57]. Ferrante et al. [66] used age, frailty, pre-existing disabilities, depressive symptoms, previous hospitalisations and hospital length of stay to predict persistent functional impairment in the year after discharge from the ICU, though with only moderate discrimination (AUC-ROC 0.71). In another study, severe limitations of health-related quality of life were found in half of the patients with COVID-19, who

were 70 years and older, three months after ICU admission and to be associated with pre-existing frailty [67]. However, the generalisability of this particular finding might be limited by the admission biases caused by the constraints on ICU resources during the COVID-19 pandemic [68].

According to an international expert consensus, predicting functional outcomes and quality of life is generally compromised by obstacles to quantifying these measures robustly [69]. Future instruments to assess the quality of life in a very old person should consider the specific expectations and priorities in this age group, which may differ significantly from those in younger cohorts. There also is a broad spectrum of contextual confounders, ranging from individual preferences of decision-makers to access to healthcare services, which may influence these functional outcomes [21]. Moreover, patients' attitudes are known to fluctuate over time [70]. Therefore, the performance of current methods for predicting outcomes in the above domains does not appear to be robust enough for determining major decisions about LST.

New machine learning technologies

The number of patient characteristics found to correlate with outcomes of intensive care continues to rise, notably when considering the large set of features which can be provided by geriatric assessments [56] or dynamic predictions in the ICU [49]. The resulting expansion of patients' datasets in size and dimensionality constitutes a formidable challenge for predictive modelling [71]. New technologies in machine learning (ML) were designed to process large datasets to identify relationships between multiple variables describing patients, context and care processes. Until recently, most of these technologies were based on scaling-up classical methodologies, such as logistic regression in artificial neural networks or decision trees in random forests.

The quality and granularity of data for training remains pivotal for the performance of models, even when using 'big data' [26]. This concerns noise in recordings of physiological and biochemical variables, incorrect documentation of interventions and the omission of decisions about

(See figure on next page.)

Fig. 3 Illustration of discrimination of prediction models for simulated distributions of survival for various degrees of frailty, which was measured with the Clinical Frailty Scale (CFS) [39]. Percentages of ICU survivors and non-survivors per CFS category are depicted by light and dark bars, respectively, on the left side. The corresponding receiver operating characteristic (ROC) curves for survival predictions and the area under these curves (AUC-ROC) were obtained using the R package pROC [40] and are shown on the right side. Panel A depicts distributions of ICU survivors and non-survivors using information from the VIP2 study about the percentages in each CFS category [41]. Overall discrimination between the two outcomes (AUC-ROC = 0.69) is sub-optimal. Panel B shows a hypothetical scenario with a linear relationship between frailty (CFS) and ICU mortality. Despite the good AUC-ROC of 0.8, there is maximum uncertainty, i.e. a 50% chance of survival, for patients with CFS 5 (mild frailty). Panel C presents another hypothetical scenario resulting in an AUC-ROC of 0.9 which indicates a very good discrimination for the prediction model. However, the overlap at CFS 4 (very mild frailty) leads to substantial prognostic uncertainty for the patients in this particular frailty category

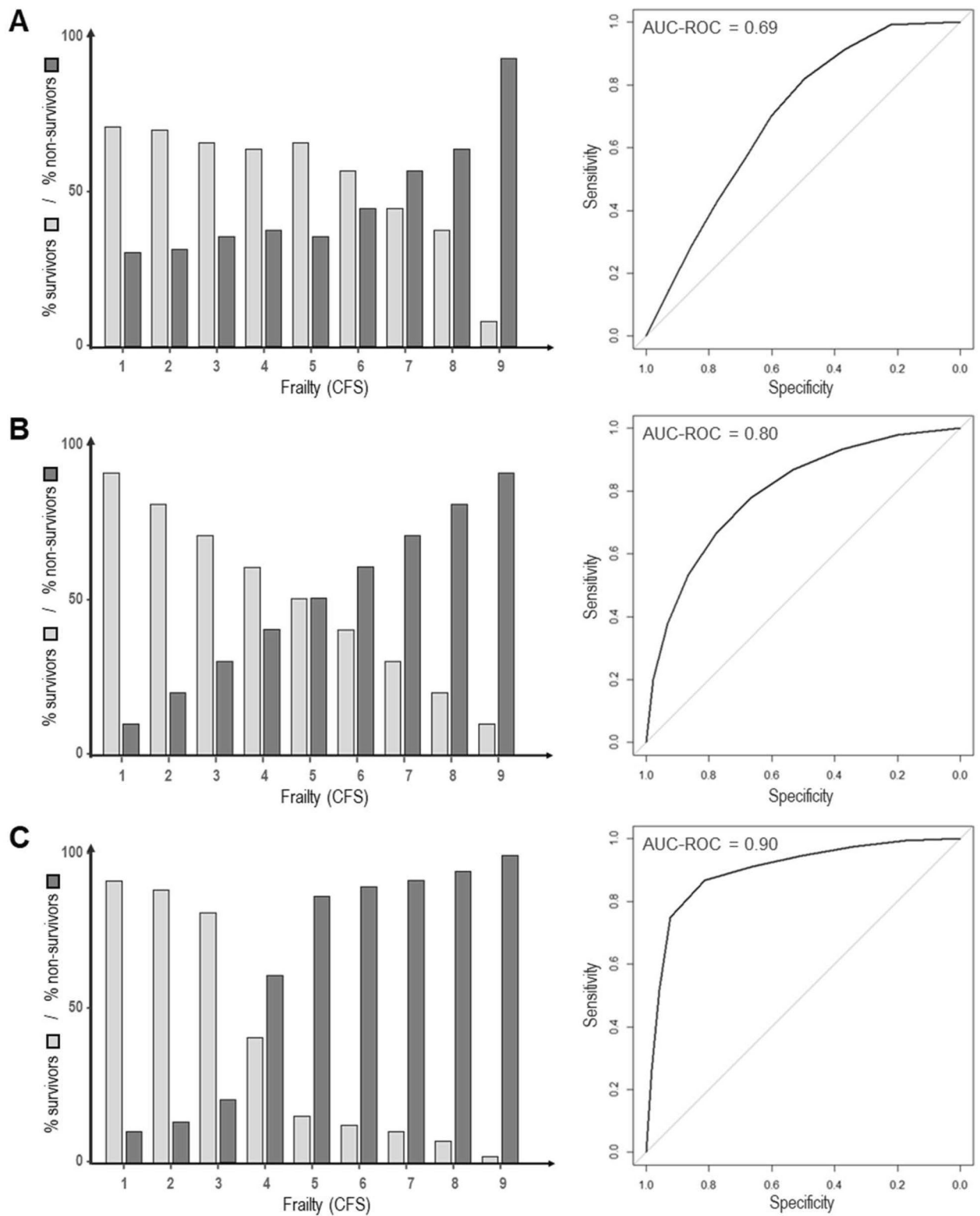


Fig. 3 (See legend on previous page.)

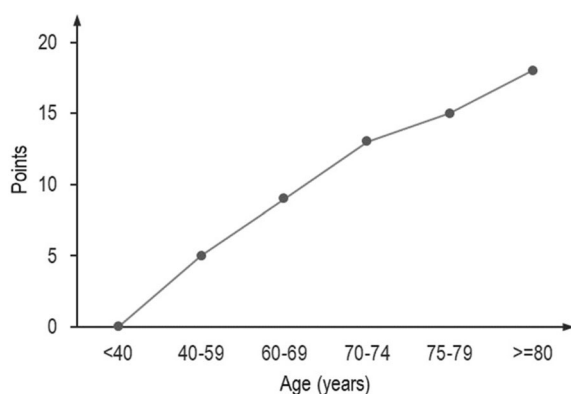


Fig. 4 Impact of chronological age on the SAPS 3 score. The plot shows the points added to the score for patients in specific age categories. Patients aged 70 years or older get more points for their chronological age than younger patients would receive for a diagnosis of cirrhosis or cancer, which would add 8 or 11 points, respectively

LST [20]. Moreover, cultural and professional norms, such as local policies of limiting LST, influence outcome and are implicitly integrated in historical datasets. They may affect the generalisability of prediction models if their association with outcome remains undetected during the development and deployment of these models.

New and more sophisticated types of neural networks (transformers) were developed by enhancing the interconnections between their sub-units to detect more complex patterns in data [72, 73]. This technology is evolving rapidly, and we would like to refer the interested reader to the latest literature in this domain. Foundation models, often known as large language models, are based on the transformer architecture and represent today's frontier in ML technologies and artificial intelligence [74]. Factuality and adversarial safety of these models have been a problem but can be improved by adding verification with data from trusted sources [75–77]. The versatility of foundation models has been illustrated by the performance of models processing a broad spectrum of knowledge in different domains and, importantly, showing reasoning capabilities [78]. These capabilities appear to be an emergent feature of very large models and may eventually enhance the generalisability of the embedded information [79]. Smaller models can be effectively trained by focussing on domain-specific data, such as clinical patient notes [28]. This data source also contains information about professional norms and care processes, which is usually not included in classical prediction models. A large hospital system has applied this new technology to generate outcome predictions for an unselected cohort of almost 400 000 patients admitted to hospital during the period 2011 to 2020 [28]. Although discrimination for in-hospital mortality was excellent

(AUC-ROC 0.95), it remains to be seen if this performance level can be replicated with respect to the characteristics and outcomes in very old patients.

Although numerous ML models have been developed for outcome predictions in intensive care for unselected patient populations or disease-specific sub-groups [80], there is a scarcity of ML models designed for prognostication in very old patients using the specific set of geriatric characteristics in this cohort [57]. In the one study published in this field, an augmented version of decision tree analysis was employed to provide survival predictions based on six categories of variables, including demographic and acute patient characteristics as well as features of frailty and multi-morbidity. The resulting model showed only a slightly better discrimination than classical disease severity scores, such as SAPS II and APACHE IV, for patients older than 65 years [17]. However, sub-group analyses revealed a decreased discrimination for in-hospital mortality in patients older than 80 years, which confirms findings in previous studies using conventional methods [16].

We have previously discussed the benefits and pitfalls of ML technologies and the related uncertainties about predictions from an ethics point of view [81]. The recent advent of large foundation models has only heightened the concerns underlined in that article. This especially applies to the selection of unbiased datasets for training, e.g. by assessing the incidence of decisions to limit LST which is expected to be very variable [25], and the framework for an oversight of algorithms to detect inappropriate outputs [75]. Moreover, the importance of transparency and explainability of ML models in defining their utility in clinical practice remains a matter of debate [82]. Methods which estimate the effect of a specific patient characteristic on outcome predictions can provide a solution for the issue of transparency [83]. Yet, the size of the effect of a variable on the model's output does not explain the parameter's specific biological role, which would be useful for causal reasoning.

From outcome predictions to decisions—the challenge of uncertainty

Current prediction models are designed to generate a probability for a specific outcome derived from the statistical properties of training cohorts. The resulting number, however, can only be integrated into the decision-making for individual patients with additional considerations. First, decision-makers have to verify that a prediction model is applicable within the specific context of the case since variations of care processes between healthcare settings may substantially alter the calibration and discrimination of models [47]. In particular, recommendations for specific cut-offs with respect to major decisions about LST should be scrutinised for unwanted consequences

[84]. Second, even if the patient and context variables matched the training cohort, the probabilistic nature of model predictions requires dichotomisation of the model output into categories of actionable decisions when, for example, contemplating withholding or withdrawing LST in the individual patient. This process necessitates an understanding of the uncertainty regarding both the predictions of the model in general and the specific risk given for an individual patient [85, 86]. Uncertainty can be quantified using the concept of entropy [87] and monitored over time to understand how the confidence in outcome predictions evolves during critical care [88]. This might be particularly relevant to the cohort of older patients due to their inter-individual heterogeneity, which results in a high degree of uncertainty in statistical predictions when provided only at a single point in time.

Prognostication is a prelude to decision-making in intensive care. Decisions should be based on an objective description of the critical condition to be fair and robust. They should also be aligned with the patient's individual preferences which is especially relevant to individuals with complex co-morbidities [89, 90]. However, discussing predictions about likely outcomes with patients or surrogates can be challenging. This step benefits from suitable communication skills to translate numerical data, i.e. probabilities, into information that can be processed by lay people during decision-making. The finding that better numeracy in patients is associated with better outcome [91] indicates the need to further improve these skills among healthcare professionals.

The impact of predictive uncertainty on the decision-making in clinical practice has been acknowledged by clinicians as well as professional institutions [9, 92]. A time-limited trial (TLT) in the ICU can be a suitable intervention if the uncertainty about the predicted outcome is uncomfortably high. A TLT is a collaborative plan among clinicians and the patient or surrogate decision makers to use LST for a defined duration, after which the patient's response to this treatment informs the decision to continue or withdraw LST [7]. On this background, a TLT provides the opportunity to gather additional information, notably about the patient's response to interventions [93], which can enhance the accuracy of prognostication [50]. However, a TLT may also result in increased uncertainty about survival, e.g. when new data obtained after stabilisation of organ function makes the predicted mortality drop from 80 to 50%, which represents the peak of uncertainty. These situations require a careful and comprehensive assessment as well as discussions with other stakeholders about individual goals of care and, thereby, the individual objective for prognostication [94]. Especially in very old patients, the predicted benefit of intensive care has to be weighed

against its physiological and psychological burden in the context of an enhanced vulnerability to stress (frailty). Shared decision-making [89] may also include reflections on the impact of a potentially suboptimal outcome on the wellbeing of caregivers, when assessed from the patient's perspective.

Discussion

The prediction of outcomes for the individual patient determines major decisions along the pathway of critical care [11]. This concerns decisions about admission to the ICU as well as those about continuation, escalation or withdrawal of LST in the ICU. Whereas prognostication at the point of a potential ICU admission is usually based on a fixed amount of information and made under time constraints, decision-making in the ICU can benefit from flexibility in the extent of time available for collecting more data and improve prognostication for the individual patient. Considering time as an additional dimension in predictive modelling, however, requires more sophisticated methods than those currently used in clinical practice.

When statistical models are employed for the purpose of prognostication, particular attention needs to be paid to the calibration and discrimination of these models in a specific healthcare setting. These factors are influenced by the generalisability of a model's underlying assumptions and contextual parameters, such as cultural norms for limiting LST or the availability of early rehabilitation for ICU patients [59]. Failure to acknowledge these issues may lead to faulty predictions and inappropriate care.

Additional assessments and reflections are necessary to integrate model-based predictions into robust decision-making about critical care. Reflections about prognostic uncertainty should go beyond a simple acknowledgement [9] and extend to quantification [85, 95]. Understanding the degree of the uncertainty of predictions is particularly relevant to dichotomous and irreversible decisions about LST. However, it remains to be determined what specific degree of uncertainty might be acceptable when translating model predictions into actions [92]. Additional steps might be necessary to manage situations of enhanced uncertainty. These include obtaining second opinions from experts outside the ICU team, such as from geriatricians, and engaging in a robust communication with patients or surrogate decision-makers.

The problem of the uncertainty of initial prognostic assessments can be mitigated by adding information about the patient's response to treatment over time [7]. This approach is of particular value for the cohort of very old patients with a high prevalence of various degrees of frailty, which confers an increased but poorly predictable vulnerability to the stress caused by critical conditions.

Table 1 Problems of integrating prognostic modelling into decision-making processes in the ICU

Problem	Solutions
Generalisability of predictive models	Add variables describing specific legal, cultural and professional norms as well as available resources, which influence care processes, to prediction models
Prediction of functional outcome and quality of life in very old patients	Suitable training samples with respect to chronic conditions (e.g. frailty, multi-morbidity) and impairments (e.g., assisted living) typical for very old age
Interpretability of model predictions	Incorporate explainability metrics
Alignment of data-informed decisions with patient's preferences	Document patient's goals and expectations, share decision-making
Uncertainty of outcome predictions for individual patients	Time-limited trials with sequentially updated predictions, shared decision-making integrating the burden of treatment

Analysing time-dependent (dynamic) scores for organ failure and disease severity is a promising approach to enhance predictive accuracy. As data acquisition and processing are becoming fully automated, dynamic methods for prognostication may be developed into efficient tools for decision support. However, the burden of care as perceived by the individual patient, which is influenced by age-related conditions, has to be integrated into decisions about LST. Since there are no robust instruments for quantifying symptoms in this domain, the decision-making still requires a prudent judgment by the ICU team.

Conclusions

Current methods for predicting outcomes of intensive care are imperfect and require careful implementation, especially for very old patients with complex chronic health status and variable personal expectations. The additional consideration of geriatric and time-dependent patient characteristics and a better generalisability of prediction models can further refine prognostication to support individualised decisions about LST in this patient cohort.

From our point of view, adjustments of the methods for prognostic modelling and their careful integration into the decision-making processes in the ICU may provide the opportunity for actionable prognostication in clinical practice (Table 1). The next generations of intensivists will have to acquire advanced knowledge about probabilistic reasoning to better understand the strengths and limitations of these technologies [96]. However, regardless of the refinement of data science technologies, the predictability of future events for the individual patient will remain limited due to the probabilistic nature of these methods. Therefore, conversations about prognostic uncertainty with patients and families are a crucial component of decision-making before admission to the ICU or in the ICU [94]. The variable preferences and wishes of stakeholders have to

be balanced with predicted outcomes and their uncertainty when making decisions about LST [97]. In these situations, the trade of decision-making still remains an art.

Abbreviations

APACHE	Acute physiology and chronic health evaluation (score)
AUC-ROC	Area under the curve of the receiver operating characteristic
CFS	Clinical frailty scale
ICU	Intensive care unit
LST	Life-sustaining treatment
ML	Machine learning
SAPS	Simplified acute physiology score
SOFA	Sequential organ failure assessment (score)
TLT	Time-limited trial

Author contributions

MB, RM, JF, YK, BG and WS conceived this study. All authors revised the manuscript and approved the final version.

Funding

Open Access funding enabled and organized by Projekt DEAL. Institutional funding.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

None.

Received: 16 February 2024 Accepted: 10 June 2024

Published online: 22 June 2024

References

- Detsky ME, Harhay MO, Bayard DF, et al. Discriminative accuracy of physician and nurse predictions for survival and functional outcomes 6 months after an ICU admission. *JAMA*. 2017;317(21):2187–95. <https://doi.org/10.1001/jama.2017.4078>.
- Arkes HR, Abernethy SK, Arpin KA. Analysis of physicians' probability estimates of a medical outcome based on a sequence of events. *JAMA Netw*

- Open. 2022;5(6): e2218804. <https://doi.org/10.1001/jamanetworkopen.2022.18804>.
3. Meadow W, Pohlman A, Frain L, Ren Y, Kress JP, Teuteberg W, Hall J. Power and limitations of daily prognostications of death in the medical intensive care unit. *Crit Care Med*. 2011;39(3):474–9. <https://doi.org/10.1097/CCM.0b013e318205df9b>.
 4. Vlachos S, Rubenfeld G, Menon D, Harrison D, Rowan K, Maharaj R. Early and late withdrawal of life-sustaining treatment after out-of-hospital cardiac arrest in the United Kingdom: institutional variation and association with hospital mortality. *Resuscitation*. 2023;193: 109956. <https://doi.org/10.1016/j.resuscitation.2023.109956>.
 5. Graham M. Burying our mistakes: dealing with prognostic uncertainty after severe brain injury. *Bioethics*. 2020;34:612–9. <https://doi.org/10.1111/bioe.12737>.
 6. Labarère J, Bertrand R, Fine MJ. How to derive and validate clinical prediction models for use in intensive care medicine. *Intensive Care Med*. 2014;40:513–27. <https://doi.org/10.1007/s00134-014-3227-6>.
 7. Kruser JM, Ashana DC, Courtright KR, Kross EK, Neville TH, Rubin E, Schenker Y, et al. Defining the time-limited trial for patients with critical illness: an official american thoracic society workshop report. *Ann Am Thorac Soc*. 2024;21(2):187–99. <https://doi.org/10.1513/annalsats.202310-925st>.
 8. Sprung CL, Artigas A, Kesecioglu J, Pezzi A, Wiis J, Pirracchio R, Baras M, Edbrooke DL, Pesenti A, Bakker J, Hargreaves C, Gurman G, Cohen SL, Lippert A, Payen D, Corbella D, Iapichino G. The Eldicus prospective, observational study of triage decision making in European intensive care units. part II: intensive care benefit for the elderly. *Crit Care Med*. 2012;40(1):132–8. <https://doi.org/10.1097/CCM.0b013e318232d6b0>.
 9. Beil M, van Heerden PV, Joynt GM, Lapinsky S, Flaatten H, Guidet B, de Lange D, Leaver S, Jung C, Forte DN, Bin D, Elhadi M, Szczeklik W, Sviri S. Limiting life-sustaining treatment for very old ICU patients: cultural challenges and diverse practices. *Ann Intensive Care*. 2023;13(1):107. <https://doi.org/10.1186/s13613-023-01189-8>.
 10. Aliberti MJR, Bailly S, Anstey M. Tailoring treatments to older people in intensive care a way forward. *Intensive Care Med*. 2022;48(12):1775–7. <https://doi.org/10.1007/s00134-022-06916-6>.
 11. Guidet B, Vallet H, Flaatten H, Joynt G, Bagshaw SM, Leaver SK, Beil M, Du B, Forte DN, Angus DC, Sviri S, de Lange D, Herridge MS, Jung C. The trajectory of very old critically ill patients. *Intensive Care Med*. 2024. <https://doi.org/10.1007/s00134-023-07298-z>.
 12. Oh HS, Rutledge J, Nachun D, Pálóvic R, Abiose O, Moran-Losada P, Channappa D, Urey DY, et al. Organ aging signatures in the plasma proteome track health and disease. *Nature*. 2023;624(7990):164–72. <https://doi.org/10.1038/s41586-023-06802-1>.
 13. Howlett SE, Rutenberg AD, Rockwood K. The degree of frailty as a translational measure of health in aging. *Nat Aging*. 2021;1(8):651–65. <https://doi.org/10.1038/s43587-021-00099-3>.
 14. Beil M, Flaatten H, Guidet B, Sviri S, Jung C, de Lange D, Leaver S, Fjølner J, Szczeklik W, van Heerden PV. The management of multi-morbidity in elderly patients: ready yet for precision medicine in intensive care? *Crit Care*. 2021;25(1):330. <https://doi.org/10.1186/s13054-021-03750-y>.
 15. Ferrucci L, Kuchel GA. Heterogeneity of aging: individual risk factors, mechanisms, patient priorities, and outcomes. *J Am Geriatr Soc*. 2021;69(3):610–2. <https://doi.org/10.1111/jgs.17011>.
 16. Flaatten H, de Lange DW, Artigas A, Bin D, Moreno R, Christensen S, Joynt GM, Bagshaw SM, Sprung CL, Benoit D, Soares M, Guidet B. The status of intensive care medicine research and a future agenda for very old patients in the ICU. *Intensive Care Med*. 2017;43(9):1319–28. <https://doi.org/10.1007/s00134-017-4718-z>.
 17. Liu X, Hu P, Yeung W, Zhang Z, Ho V, Liu C, Dumontier C, Thorat PJ, Mao Z, Cao D, Mark RG, Zhang Z, Feng M, Li D, Celi LA. Illness severity assessment of older adults in critical illness using machine learning (ELDER-ICU): an international multicentre study with subgroup bias evaluation. *Lancet Digit Health*. 2023;5(10):e657–67. [https://doi.org/10.1016/S2589-7500\(23\)00128-0](https://doi.org/10.1016/S2589-7500(23)00128-0).
 18. van Sleetuwen D, Zegers M, Ramjith J, Crujtsberg JK, Simons KS, van Bommel D, Burgers-Bonthuis D, Koeter J, Bisschops LLA, Janssen I, Rettig TCD, van der Hoeven JG, van de Laar FA, van den Boogaard M. Prediction of long-term physical, mental, and cognitive problems following critical illness: development and external validation of the PROSPECT prediction model. *Crit Care Med*. 2024;52(2):200–9. <https://doi.org/10.1097/CCM.0000000000006073>.
 19. Mousai O, Tafoureau L, Yovell T, Flaatten H, Guidet B, Beil M, de Lange D, Leaver S, Szczeklik W, Fjølner J, Nachshon A, van Heerden PV, Joskowicz L, Jung C, Hyams G, Sviri S. The role of clinical phenotypes in decisions to limit life-sustaining treatment for very old patients in the ICU. *Ann Intensive Care*. 2023;13(1):40. <https://doi.org/10.1186/s13613-023-01136-7>.
 20. Flaatten H, deLange D, Jung C, Beil M, Guidet B. The impact of end-of-life care on ICU outcome. *Intensive Care Med*. 2021;47:624–5. <https://doi.org/10.1007/s00134-021-06365-7>.
 21. Nordenskjöld Syrous A, Malmgren J, Odenstedt Hergés H, Olausson S, Kock-Redfors M, Ågård A, Block L. Reasons for physician-related variability in end-of-life decision-making in intensive care. *Acta Anaesthesiol Scand*. 2021;65(8):1102–8. <https://doi.org/10.1111/aas.13842>.
 22. Vallet H, Schwarz GL, Flaatten H, de Lange DW, Guidet B, Dechartres A. Mortality of older patients admitted to an ICU: a systematic review. *Crit Care Med*. 2021;49(2):324–34. <https://doi.org/10.1097/CCM.00000000000004772>.
 23. Curtis JR, Barnato AE. Variability in decisions to limit life-sustaining treatments: is it all about the physician? *Chest*. 2014;146(3):532–4. <https://doi.org/10.1378/chest.14-0636>.
 24. Ceconi M, Spies CD, Moreno R. Economic sustainability of intensive care in Europe. *Intensive Care Med*. 2024;50(1):136–40. <https://doi.org/10.1007/s00134-023-07268-5>.
 25. De-Arteaga M, Elmer J. Self-fulfilling prophecies and machine learning in resuscitation science. *Resuscitation*. 2023;183: 109622. <https://doi.org/10.1016/j.resuscitation.2022.10.014>.
 26. Morris AH, Horvat C, Stagg B, Grainger DW, Lanspa M, Orme J, Clemmer TP, Weaver LK, et al. Computer clinical decision support that automates personalized clinical care: a challenging but needed healthcare delivery strategy. *J Am Med Inform Assoc*. 2022;30(1):178–94. <https://doi.org/10.1093/jamia/ocac143>.
 27. Chang DW, Neville TH, Parrish J, et al. Evaluation of time-limited trials among critically ill patients with advanced medical illnesses and reduction of nonbeneficial ICU treatments. *JAMA Intern Med*. 2021;181(6):786–94. <https://doi.org/10.1001/jamainternmed.2021.1000>.
 28. Jiang LY, Liu XC, Nejatian NP, Nasir-Moin M, Wang D, Abidin A, Eaton K, Riina H, Laufer I, Punjabi P, et al. Health system-scale language models are all-purpose prediction engines. *Nature*. 2023;619(7969):357–62. <https://doi.org/10.1038/s41586-023-06160-y>.
 29. Poole D, Bianchin A, Caria FC, Dal Cero P, Cian DE, S, Galeotti E, Menato L, Peta M, Turchet F, Zamperoni A. Differences in early, intermediate, and long-term mortality among elderly patients admitted to the ICU: results of a retrospective observational study. *Minerva Anestesiol*. 2022;88(6):479–89. <https://doi.org/10.23736/S0375-9393.22.16002-5>.
 30. Ball IM, Bagshaw SM, Burns KE, Cook DJ, Day AG, Dodek PM, Kutsogiannis DJ, Mehta S, Muscedere JG, Stelfox HT, Turgeon AF, Wells GA, Stiell IG. A clinical prediction tool for hospital mortality in critically ill elderly patients. *J Crit Care*. 2016;35:206–12. <https://doi.org/10.1016/j.jcrr.2016.05.026>.
 31. Heyland DK, Garland A, Bagshaw SM, Cook D, Rockwood K, Stelfox HT, Dodek P, Fowler RA, Turgeon AF, Burns K, Muscedere J, Kutsogiannis J, Albert M, Mehta S, Jiang X, Day AG. Recovery after critical illness in patients aged 80 years or older: a multi-center prospective observational cohort study. *Intensive Care Med*. 2015;41(11):1911–20. <https://doi.org/10.1007/s00134-015-4028-2>.
 32. Sim YS, Jung H, Shin TR, Kim DG, Park SM. Mortality and outcomes in very elderly patients 90 years of age or older admitted to the ICU. *Respir Care*. 2015;60:347–55. <https://doi.org/10.4187/respcare.03155>.
 33. Zampieri FG, Colombari F. The impact of performance status and comorbidities on the short-term prognosis of very elderly patients admitted to the ICU. *BMC Anesthesiol*. 2014;22(14):59. <https://doi.org/10.1186/1471-2253-14-59>.
 34. Metnitz PG, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR. SAPS 3 investigators. SAPS 3—from evaluation of the patient to evaluation of the intensive care unit. part 1: objectives, methods and cohort description. *Intensive Care Med*. 2005. <https://doi.org/10.1007/s00134-005-2762-6>.
 35. Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR. SAPS 3 investigators. SAPS 3—from evaluation of the patient to evaluation of the intensive care unit. part 2: development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Med*. 2005;31(10):1345–55. <https://doi.org/10.1007/s00134-005-2763-5>.

36. Nattino G, Finazzi S, Bertolini G. A new test and graphical tool to assess the goodness of fit of logistic regression models. *Stat Med*. 2016;35(5):709–20. <https://doi.org/10.1002/sim.6744>.
37. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU collaborative research database, a freely available multi-center database for critical care research. *Sci Data*. 2018;11(5): 180178. <https://doi.org/10.1038/sdata.2018.178>.
38. Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L. A tutorial on calibration measurements and calibration models for clinical prediction models. *J Am Med Inform Assoc*. 2020;27(4):621–33. <https://doi.org/10.1093/jamia/oc228>.
39. Pulok MH, Theou O, van der Valk AM, Rockwood K. The role of illness acuity on the association between frailty and mortality in emergency department patients referred to internal medicine. *Age Ageing*. 2020;49(6):1071–9. <https://doi.org/10.1093/ageing/afaa089>.
40. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform*. 2011;17(12):77. <https://doi.org/10.1186/1471-2105-12-77>.
41. Guidet B, de Lange DW, Boumendil A, Leaver S, Watson X, Boulanger C, et al. The contribution of frailty, cognition, activity of daily life and comorbidities on outcome in acutely admitted patients over 80 years in European ICUs: the VIP2 study. *Intensiv Care Med*. 2020;46(1):57–69. <https://doi.org/10.1007/s00134-019-05853-1>.
42. Bassford CR, Krucien N, Ryan M, Griffiths FE, Svantesson M, Fritz Z, Perkins GD, Quinton S, Slowther AMUK. Intensivists' preferences for patient admission to ICU: evidence from a choice experiment. *Crit Care Med*. 2019;47(11):1522–30. <https://doi.org/10.1097/CCM.0000000000003903>.
43. Avidan A, Sprung CL, Scheffold JC, Ricou B, Hartog CS, Nates JL, Jaschinski U, Lobo SM, Joynt GM, Lesieur O, et al. Variations in end-of-life practices in intensive care units worldwide (ethicus-2): a prospective observational study. *Lancet Respir Med*. 2021;9(10):1101–10. [https://doi.org/10.1016/S2213-2600\(21\)00261-7](https://doi.org/10.1016/S2213-2600(21)00261-7).
44. Naik H, Murray TM, Khan M, Daly-Grafstein D, Liu G, Kassen BO, Onrot J, Sutherland JM, Staples JA. Population-based trends in complexity of hospital inpatients. *JAMA Intern Med*. 2024;184(2):183–92. <https://doi.org/10.1001/jamainternmed.2023.7410>.
45. Van Calster B, Steyerberg EW, Wynants L, van Smeden M. There is no such thing as a validated prediction model. *BMC Med*. 2023;21(1):70. <https://doi.org/10.1186/s12916-023-02779-w>.
46. Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. *Lancet Digit Health*. 2020;2(9):e489–92. [https://doi.org/10.1016/S2589-7500\(20\)30186-2](https://doi.org/10.1016/S2589-7500(20)30186-2).
47. Wong A, Otles E, Donnelly JP, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med*. 2021;181(8):1065–70. <https://doi.org/10.1001/jamainternmed.2021.2626>.
48. Yek C, Wang J, Fintzi J, Mancera AG, Keller MB, Warner S, Kadri SS. Impact of surge strain and pandemic progression on prognostication by an established COVID-19-specific severity score. *Crit Care Explor*. 2023;5(12): e1021. <https://doi.org/10.1097/CCE.0000000000001021>.
49. Lapp L, Roper M, Kavanagh K, Bouamrane MM, Schraag S. Dynamic prediction of patient outcomes in the intensive care unit: a scoping review of the state-of-the-art. *J Intensiv Care Med*. 2023;38(7):575–91. <https://doi.org/10.1177/08850666231166349>.
50. Meiring C, Dixit A, Harris S, MacCallum NS, Brealey DA, Watkinson PJ, Jones A, Ashworth S, Beale R, Brett SJ, Singer M, Ercole A. Optimal intensive care outcome prediction over time using machine learning. *PLoS ONE*. 2018;13(11): e0206862. <https://doi.org/10.1371/journal.pone.0206862>.
51. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. on behalf of the working group on sepsis-related problems of the European society of intensive care medicine. *Intensiv Care Med*. 1996;22(7):707–10. <https://doi.org/10.1007/BF01709751>.
52. Moreno R, Vincent JL, Matos R, Mendonça A, Cantraine F, Thijs L, Takala J, Sprung C, Antonelli M, Bruining H, Willatts S. The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. *Intensiv Care Med*. 1999;25(7):686–96. <https://doi.org/10.1007/s001340050931>.
53. Ferrante LE, Pisani MA, Murphy TE, Gahbauer EA, Leo-Summers LS, Gill TM. Functional trajectories among older persons before and after critical illness. *JAMA Intern Med*. 2015;175(4):523–9. <https://doi.org/10.1001/jamainternmed.2014.7889>.
54. Bruno RR, Wernly B, Bagshaw SM, van den Boogaard M, Darvall JN, De Geer L, de Gopegui Miguelena PR, Heyland DK, et al. The clinical frailty scale for mortality prediction of old acutely admitted intensive care patients: a meta-analysis of individual patient-level data. *Ann Intensiv Care*. 2023;13(1):37. <https://doi.org/10.1186/s13613-023-01132-x>.
55. Iwashyna TJ, Hodgson CL, Pilcher D, Bailey M, van Lint A, Chavan S, Bellomo R. Timing of onset and burden of persistent critical illness in Australia and New Zealand: a retrospective, population-based, observational study. *Lancet Respir Med*. 2016;4(7):566–73. [https://doi.org/10.1016/S2213-2600\(16\)30098-4](https://doi.org/10.1016/S2213-2600(16)30098-4).
56. Veronese N, Custodero C, Demurtas J, Smith L, Barbagallo M, Maggi S, Cella A, Vanacore N, Aprile PL, Ferrucci L, Pilotto A. Special interest group in systematic reviews of the European geriatric medicine society (EuGMS); special interest group in meta-analyses and comprehensive geriatric assessment of the European geriatric medicine society (EuGMS). comprehensive geriatric assessment in older people: an umbrella review of health outcomes. *Age Ageing*. 2022. <https://doi.org/10.1093/ageing/afac104>.
57. Vallet H, Guidet B, Boumendil A, De Lange DW, Leaver S, Szczeklik W, Jung C, Sviril S, Beil M, Flaatten H. The impact of age-related syndromes on ICU process and outcomes in very old patients. *Ann Intensiv Care*. 2023;13(1):68. <https://doi.org/10.1186/s13613-023-01160-7>.
58. Jain S, Murphy TE, O'Leary JR, Leo-Summers L, Ferrante LE. Association between socioeconomic disadvantage and decline in function, cognition, and mental health after critical illness among older adults: a cohort study. *Ann Intern Med*. 2022;175(5):644–55. <https://doi.org/10.7326/M21-3086>.
59. Jacobs MJ, Rahamim A, Beil M, Guidet B, Vallet H, Flaatten H, et al. Critical care beyond organ support: the importance of geriatric rehabilitation. *Ann Intensiv Care*. 2024. <https://doi.org/10.1186/s13613-024-01306-1>.
60. Liu X, Shen M, Lie M, Zhang Z, Li D, Liu C, Mark R, Zhang Z, Celi LA. Evaluating prognostic bias of critical illness severity scores based on age, gender, and primary language in the USA. *medRxiv*. 2022. <https://doi.org/10.1101/2022.08.01.22277736>.
61. Moreno R, Rhodes A, Piquilloud L, Hernandez G, Takala J, Gershengorn HB, Tavares M, Coopersmith CM, et al. The sequential organ failure assessment (SOFA) score: has the time come for an update? *Crit Care*. 2023;27(1):15. <https://doi.org/10.1186/s13054-022-04290-9>.
62. Langenberg C, Hingorani AD, Whitty CJM. Biological and functional multi-morbidity-from mechanisms to management. *Nat Med*. 2023;29(7):1649–57. <https://doi.org/10.1038/s41591-023-02420-6>.
63. de Lange DW, Brinkman S, Flaatten H, Boumendil A, Morandi A, Andersen FH, Artigas A, Bertolini G, Cecconi M, et al. Cumulative prognostic score predicting mortality in patients older than 80 years admitted to the ICU. *J Am Geriatr Soc*. 2019;67(6):1263–7. <https://doi.org/10.1111/jgs.15888>.
64. Mousai O, Tafoureau L, Yovell T, Flaatten H, Guidet B, Jung C, de Lange D, Leaver S, Szczeklik W, Fjølner J, van Heerden PV, Joskowicz L, Beil M, Hyams G, Sviril S. Clustering analysis of geriatric and acute characteristics in a cohort of very old patients on admission to ICU. *Intensiv Care Med*. 2022;48(12):1726–35. <https://doi.org/10.1007/s00134-022-06868-x>.
65. Herridge MS, Azoulay E. Outcomes after critical illness. *N Engl J Med*. 2023;388(10):913–24. <https://doi.org/10.1056/NEJMra2104669>.
66. Ferrante LE, Murphy TE, Leo-Summers LS, O'Leary JR, Vander Wyk B, Pisani MA, Gill TM. Development and validation of a prediction model for persistent functional impairment among older ICU survivors. *J Am Geriatr Soc*. 2023;71(1):188–97. <https://doi.org/10.1111/jgs.18075>.
67. Soliman IW, Leaver S, Flaatten H, Fjølner J, Wernly B, Bruno RR, Artigas A, Bollen Pinto B, Scheffold JC, et al. Health-related quality of life in older patients surviving ICU treatment for COVID-19: results from an international observational study of patients older than 70 years. *Age Ageing*. 2022. <https://doi.org/10.1093/ageing/afab278>.
68. Beil M, van Heerden PV, Sviril S. Decision-making under resource constraints. In: Flaatten H, Guidet B, Vallet H, editors. *The very old critically ill patients. lessons from the ICU*. Cham: Springer; 2022. p. 251–65.
69. Mikkelsen ME, Still M, Anderson BJ, Bienvenu OJ, Brodsky MB, Brummel N, Butcher B, Clay AS, Felt H, Ferrante LE et al. Society of critical care medicine's international consensus conference on prediction and

- identification of long-term impairments after critical illness. *Crit Care Med*. 2020; 48(11):1670–1679.
70. Dürst AV, Spencer B, Büla C, Fustinoni S, Mazzocato C, Rochat E, Rubli Truchard E, Monod S, Jox RJ. Wish to die in older patients: development and validation of two assessment instruments. *J Am Geriatr Soc*. 2020;68(6):1202–9. <https://doi.org/10.1111/jgs.16392>.
 71. Tsui A, Tudosiu PD, Brudfors M, Jha A, Cardoso J, Ourselin S, Ashburner J, Rees G, Davis D, Nachev P. Predicting mortality in acutely hospitalised older patients: the impact of model dimensionality. *BMC Med*. 2023;21(1):10. <https://doi.org/10.1186/s12916-022-02698-2>.
 72. Bubeck S, Chandrasekaran V, Eldan R et al. Sparks of artificial general intelligence: early experiments with GPT-4. [arXiv:2303.12712v5](https://arxiv.org/abs/2303.12712v5)
 73. Sibilano E, Buongiorno D, Lassi M, Grippo A, Bessi V, Sorbi S, et al. Understanding the role of self-attention in a Transformer model for the discrimination of SCD from MCI using resting-state EEG. *IEEE J Biomed Health Inform*. 2024. <https://doi.org/10.1109/JBHI.2024.3390606>.
 74. Bommasani R, Hudson DA, Adeli E, et al. On the opportunities and risks of foundation models. [arXiv](https://arxiv.org/abs/2021.10.48550). 2021. <https://doi.org/10.48550/arXiv.2108.07258>.
 75. Anderljung M, Barnhart J, Korinek A, et al. Frontier AI regulation: managing emerging risks to public safety. [arXiv](https://arxiv.org/abs/2023.2307.03718). 2023. <https://doi.org/10.48550/arXiv.2307.03718>.
 76. Tu T, Palepu A, Schaeckermann M, et al. Towards conversational diagnostic AI. [arXiv](https://arxiv.org/abs/2023.10.48550). 2023. <https://doi.org/10.48550/arXiv.2401.05654>.
 77. Zakka C, Shad R, Chaurasia A, Dalal AR, Kim JL, Moor M, Fong R, Phillips C, Alexander K, Ashley E, Boyd J, Boyd K, Hirsch K, Langlotz C, Lee R, Melia J, Nelson J, Sallam K, Tullis S, Vogelsohn MA, Cunningham JP, Hiesinger W. Almanac—retrieval-augmented language models for clinical medicine. *NEJM AI*. 2024. <https://doi.org/10.1056/aioa2300068>.
 78. Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models. *Adv Neural Inf Process Syst*. 2022;35:24824–37.
 79. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, Scales N, Tanwani A, Cole-Lewis H, et al. Large language models encode clinical knowledge. *Nature*. 2023;620:172–80. <https://doi.org/10.1038/s41586-023-06291-2>.
 80. Barboi C, Tzavelis A, Muhammad LN. Comparison of severity of illness scores and artificial intelligence models that are predictive of intensive care unit mortality: meta-analysis and review of the literature. *JMIR Med Inform*. 2022;10(5): e35293. <https://doi.org/10.2196/35293>.
 81. Beil M, Proft I, van Heerden D, Sviri S, van Heerden PV. Ethical considerations about artificial intelligence for prognostication in intensive care. *Intensiv Care Med Exp*. 2019;10(7):70. <https://doi.org/10.1186/s40635-019-0286-6>.
 82. McCoy LG, Brenna CTA, Chen SS, Vold K, Das S. Believing in black boxes: machine learning for healthcare does not need explainability to be evidence-based. *J Clin Epidemiol*. 2022;142:252–7. <https://doi.org/10.1016/j.jclinepi.2021.11.001>.
 83. Chen H, Lundberg S, Lee SI. Explaining models by propagating shapley values of local components. In: Shaban-Nejad A, Michalowski M, Buckridge DL, editors. *Explainable AI in healthcare and medicine: studies in computational intelligence*. Cham: Springer; 2021. p. 261–70.
 84. Wunsch H, Hill AD, Bosch N, et al. Comparison of 2 triage scoring guidelines for allocation of mechanical ventilators. *JAMA Netw Open*. 2020;3(12): e2029250. <https://doi.org/10.1001/jamanetworkopen.2020.29250>.
 85. Riley RD, Pate A, Dhiman P, Archer L, Martin GP, Collins GS. Clinical prediction models and the multiverse of madness. *BMC Med*. 2023;21(1):502. <https://doi.org/10.1186/s12916-023-03212-y>.
 86. Ehrmann DE, Joshi S, Goodfellow SD, Mazwi ML, Eytan D. Making machine learning matter to clinicians: model actionability in medical decision-making. *NPJ Digit Med*. 2023;6(1):7. <https://doi.org/10.1038/s41746-023-00753-7>.
 87. Beil M, van Heerden PV, de Lange DW, Szczekliki W, Leaver S, Guidet B, Flaatten H, Jung C, Sviri S, Joskowicz L. Contribution of information about acute and geriatric characteristics to decisions about life-sustaining treatment for old patients in intensive care. *BMC Med Inform Decis Mak*. 2023;23:1. <https://doi.org/10.1186/s12911-022-02094-z>.
 88. Beil M, Flaatten H, Guidet B, Joskowicz L, Jung C, de Lange D, Leaver S, Fjølner J, Szczekliki W, Sviri S, van Heerden PV. Time-dependent uncertainty of critical care transitions in very old patients—lessons for time-limited trials. *J Crit Care*. 2022;71: 154067. <https://doi.org/10.1016/j.jcrc.2022.154067>.
 89. Kon AA, Davidson JE, Morrison W, Danis M, White DB. Shared decision-making in intensive care units. executive summary of the American college of critical care medicine and american thoracic society policy statement. *Am J Respir Crit Care Med*. 2016;193(12):1334–6. <https://doi.org/10.1164/rccm.201602-0269ED>.
 90. Boyd C, Smith CD, Masoudi FA, Blaum CS, Dodson JA, Green AR, et al. Decision making for older adults with multiple chronic conditions: executive summary for the american geriatrics society guiding principles on the care of older adults with multimorbidity. *J Am Geriatr Soc*. 2019;67(4):665–73. <https://doi.org/10.1111/jgs.15809>.
 91. Congelosi PD, Carroll MC, Wong SL. Numeracy levels influence shared decision-making and surgical outcomes: a scoping review of the literature. *Am J Surg*. 2023;225(6):967–74. <https://doi.org/10.1016/j.amjsurg.2023.01.002>.
 92. General Medical Council, UK. Treatment and care towards the end of life: good practice in decision making. 2022. <https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/treatment-and-care-towards-the-end-of-life>
 93. Kruser JM, Nadig NR, Viglianti EM, Clapp JT, Secunda KE, Halpern SD. Time-limited trials for patients with critical illness: a review of the literature. *Chest*. 2023;S0012–3692(23):05855–65. <https://doi.org/10.1016/j.chest.2023.12.014>.
 94. Jackson VA, Emanuel L. Navigating and communicating about serious illness and end of life. *N Engl J Med*. 2024;390(1):63–9. <https://doi.org/10.1056/NEJMcp2304436>.
 95. Begoli E, Bhattacharya T, Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. *Nat Mach Intell*. 2019;1:20–3. <https://doi.org/10.1038/s42256-018-0004-1>.
 96. Goodman KE, Rodman AM, Morgan DJ. Preparing physicians for the clinical algorithm era. *N Engl J Med*. 2023;389(6):483–7. <https://doi.org/10.1056/NEJMp2304839>.
 97. Sviri S, Garb Y, Stav I, Rubinow A, Linton DM, Caine YG, Marcus EL. Contradictions in end-of-life decisions for self and other, expressed by relatives of chronically ventilated patients. *J Crit Care*. 2009;24(2):293–301. <https://doi.org/10.1016/j.jcrc.2009.01.011>.

Publisher's Note

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