



Review Article

The Art of Critical Data Mining — Statistical and Modeling Methods in Critical Care Database Research



Shengnan Kong^{1,2#}, Lin Chen^{3#}, Haojia Lyu⁴, Qin Lai², Xinya Li⁵, Yu Wang^{2,6,7*}  and Jun Lyu^{1,3*} 

¹Department of Pharmacy, The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong, China; ²School of Nursing, Jinan University, Guangzhou, Guangdong, China; ³Clinical Research Center, The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong, China; ⁴Department of Neurology, The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong, China; ⁵School of Nursing, Sun Yat-sen University, Guangzhou, Guangdong, China; ⁶Community Health Service Center of Jinan University, Guangzhou, Guangdong, China; ⁷Department of School Clinic, The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong, China

[#]These authors contributed equally to this work.

Received: July 22, 2026 | Revised: August 09, 2026 | Accepted: September 24, 2026 | Published online: September 28, 2026

Abstract

Large-scale critical care databases have expanded opportunities for clinical research. Previous reviews have summarized open-access intensive care unit databases and specific analytical methods, but integrated, practice-oriented guidance on commonly used and emerging methods and their selection remains limited. The objective of this narrative review is to provide a practice-oriented overview of commonly used and emerging statistical and modeling methods in critical care database research, with emphasis on their appropriate applications, strengths, limitations, and considerations for method selection. This narrative review summarizes commonly used and emerging methods in critical care database research, focusing on their applications, strengths, limitations, and practical considerations. The methods reviewed range from descriptive and regression analyses to prediction modeling, longitudinal analysis, causal inference, and artificial intelligence-based approaches. Common problems include missing data, temporal bias, confounding, overfitting, and limited external validation. Careful method selection and appropriate validation are therefore important when working with complex critical care data. Closer coordination of clinical questions, data structures, analytical methods, and reporting standards can strengthen research using intensive care unit databases. Paying closer attention to data quality, causal hypotheses, validation, reproducibility, and generalizability may improve the methodological rigor and translational value of these studies.

Introduction

The widespread adoption of electronic health record (EHR) systems and the rapid development of data-sharing platforms have substantially accelerated the use of large-scale clinical databases, such as the Medical Information Mart for Intensive Care (MIMIC)^{1,2} and the eICU Collaborative Research Database (eICU),³ in medical

research in recent years. As a result, the number of related studies has increased significantly in the last decade. These databases are characterized by large sample sizes, multidimensional data and comprehensive longitudinal information, providing a strong basis for studying disease mechanisms, evaluating therapeutic outcomes, and developing predictive models.⁴

Nevertheless, compared with traditional prospective studies, database-driven research continues to face several methodological challenges. First, the variety of statistical methods used is relatively limited, with logistic regression and conventional machine learning (ML) algorithms predominating, resulting in substantial methodological homogeneity across studies.⁵ Second, many studies still show limited methodological rigor in the development and validation of models. These limitations include a lack of external validation, insufficient control of overfitting, and reliance on a limited set of performance metrics.⁶ Previous reviews have offered valuable insights into open-access intensive care unit (ICU) databases and specific analytical methods, such as ML.^{5,6}

Keywords: Critical care database; Methodology; Causal inference; Machine learning; Predictive modeling; Target trial emulation.

***Correspondence to:** Jun Lyu, Department of Pharmacy, The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong 510630, China. ORCID: <https://orcid.org/0000-0002-2237-8771>. Tel: +86-13922274169, E-mail: lyujun2020@jnu.edu.cn; Yu Wang, Department of School Clinic, The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong 510630, China. ORCID: <https://orcid.org/0000-0003-1581-0818>. Tel: +86-18928903869, E-mail: hulibuwangyu@126.com

How to cite this article: Kong S, Chen L, Lyu H, Lai Q, Li X, Wang Y, et al. The Art of Critical Data Mining — Statistical and Modeling Methods in Critical Care Database Research. *J Transl Crit Care Med* 2026;8(3):e00011. DOI: 10.14218/JTCCM.2026.00011.

However, integrated, practice-oriented guidance on commonly used and emerging analytical methods, including their applications, strengths, limitations, and considerations for method selection, remains limited. This article presents a narrative review of commonly used and emerging methodological approaches in critical care database research. Rather than providing an exhaustive list of all available methods, the review focuses on frequently applied and increasingly relevant approaches to the main analytical tasks in this field, as reflected in the literature and the authors' experience. For organizational clarity, these approaches are grouped into five pragmatic, non-mutually exclusive domains according to their primary analytical objectives and the structure of the data being analyzed: descriptive and associational approaches; clinical prediction modeling; time-to-event, longitudinal, and dynamic methods; causal inference approaches; and artificial intelligence (AI) and advanced computational methods. The sequence progresses from characterizing data and estimating associations to predicting clinical outcomes, modeling temporal processes, estimating causal effects, and applying advanced computational techniques. Because a study design or analytical method may have multiple applications, each approach is categorized according to its predominant use without implying that the five domains are mutually exclusive. To provide a practical overview of the research process, Figure 1 presents an author-developed synthesis of the general workflow of critical care database research, informed by published methodological frameworks and guidance on clinical big data mining and the use of EHR data in critical care research.^{7,8} Table 1 summarizes the principal applications, advantages, and limitations of the methodological approaches discussed in this review.⁹⁻³⁹ Accordingly, the objective of the present narrative review is to provide clinicians and researchers with a practice-oriented overview of commonly used and emerging methodological approaches in critical care database research, focusing on the research questions they address, their analytical logic, strengths, limitations, and key considerations for method selection and implementation.

Common and emerging methodological approaches in critical care database research

Descriptive and associational approaches

One of the most common study designs in large-scale database research is the retrospective cohort study, which uses existing data to classify participants according to exposure and assess outcomes.⁴⁰ Depending on the research question, researchers may conduct descriptive, associational, predictive, or causal analyses within this design. Thus, this section focuses on descriptive and associational approaches rather than treating retrospective cohort studies as a separate analytical category.

This design is often used to examine associations between clinical characteristics and outcomes such as ICU mortality and complications in databases such as MIMIC. This design has been widely used in previous MIMIC-based studies of patients with various conditions, including ventilator-associated pneumonia,⁴¹ sepsis,⁴² ischemic stroke,⁴³ and dementia.⁴⁴

Retrospective cohort studies offer large sample sizes, high efficiency, and relatively low research costs. However, their validity depends on the quality and completeness of the available data and

may be limited by residual and unmeasured confounding.⁴⁵

Descriptive analysis

Descriptive studies are often used to summarize basic characteristics of a study population, such as demographic features, clinical indicators, and distributions of outcomes. Causal inference and predictive modeling are not included in these studies.^{46,47} This type of research is widely used in critical care database research to describe ICU patient populations, therapeutic interventions, and clinical outcomes.

For example, a MIMIC-based study of patients with ventilator-associated pneumonia used descriptive statistical analysis to characterize microbiological findings and patterns of antibiotic use.⁹ The main virtues of purely descriptive studies are simplicity, clear interpretability, and ease of implementation. These studies allow researchers to quickly outline the general characteristics of a study population and set the stage for further analyses. However, their limitations should also be recognized. Such studies cannot establish causality; they can only describe observed distributions and patterns and may be particularly susceptible to selection bias. The heterogeneity of patient populations included in large clinical databases may also limit the generalizability of the findings.⁴⁶

Multivariable association and risk-factor analysis

Multivariable regression models are commonly used in risk-factor analyses to identify variables independently associated with clinical outcomes. These models are commonly employed in critical care database research to examine associations between factors such as age, disease severity, and laboratory indicators and outcomes such as mortality, complications, and prognosis. For example, in studies based on the MIMIC cohort, multivariable Cox proportional hazards models have been used to examine the association between the Geriatric Nutritional Risk Index and 28-day mortality in elderly patients with sepsis,¹⁰ as well as the association between early prophylactic heparin use and mortality in critically ill patients with sepsis.¹¹ Multivariable regression offers several advantages: it enables the simultaneous examination of multiple predictors, allows adjustment for potential confounders, and quantifies associations using odds ratios or hazard ratios with corresponding confidence intervals. The three most common types—linear, logistic, and Cox proportional hazards regression—are applicable to continuous, binary, and time-to-event outcomes, respectively, and are readily implemented in standard statistical software packages.⁴⁸ However, this method may have limited capacity for causal inference when used in exploratory analyses without a prespecified exposure and adjustment strategy.⁴⁹

Nonlinear association modeling

Continuous variables are often dichotomized in traditional statistical analyses. However, this practice can lead to information loss and reduced statistical power.⁵⁰ Restricted cubic spline (RCS) modeling is a flexible method for modeling nonlinear relationships between continuous variables and outcomes. RCS has become widely used in database research.^{50,51}

RCS is commonly used to evaluate nonlinear dose-response relationships between biomarkers and clinical outcomes in critical care database research. For example, prior studies have applied RCS in patients with sepsis-associated acute kidney injury¹² and urosepsis.¹³ In the latter dual-cohort study, RCS showed a positive

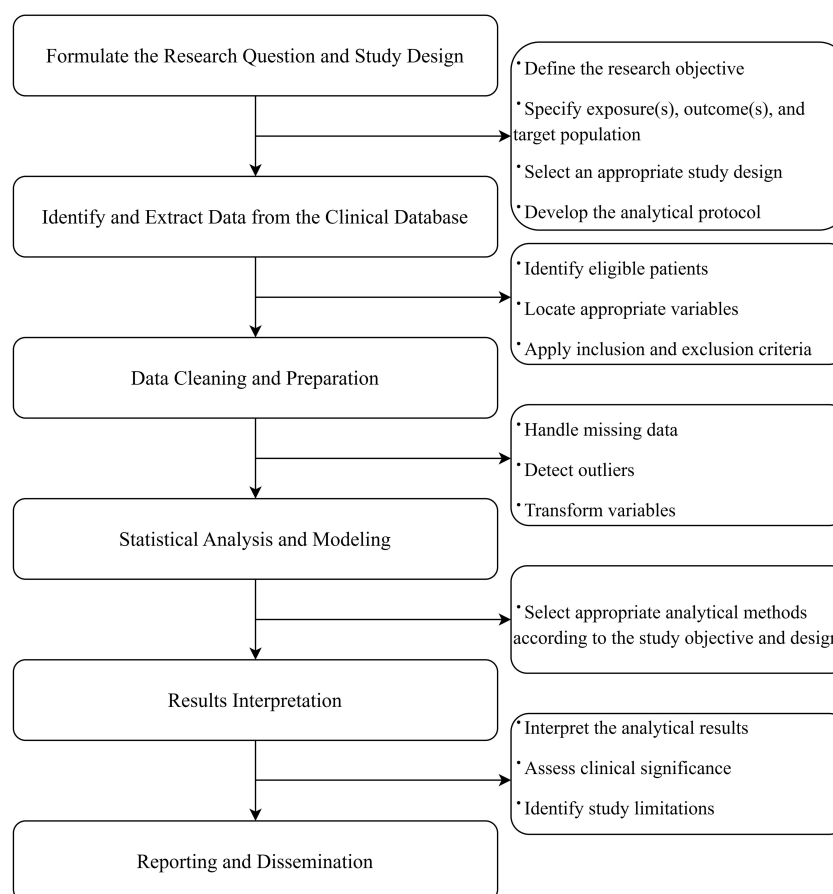


Fig. 1. Author-synthesized workflow for designing, conducting, analyzing, interpreting, and reporting critical care database research. This flowchart summarizes six major stages in critical care database research: formulation of the research question and study design; identification and extraction of data from the clinical database; data cleaning and preparation; statistical analysis and modeling; results interpretation; and reporting and dissemination. The callout boxes highlight key considerations across these stages, including defining the objective, specifying exposures and outcomes, identifying eligible patients and variables, handling missing data and outliers, selecting analytical methods aligned with the study objective and design, and assessing clinical significance and limitations. The workflow was synthesized by the authors and was conceptually informed by published methodological frameworks and guidance relevant to clinical big data mining and electronic health record research in critical care.^{7,8}

dose-response association between the red cell distribution width-to-albumin ratio and short-term mortality, illustrating how this approach can reveal more nuanced associations between continuous predictors and clinical outcomes.

The main advantage of RCS is that it can retain the full information contained in continuous variables and capture nonlinear trends flexibly and intuitively. However, this method can be sensitive to the number and location of spline knots and generally requires a sufficiently large sample size to obtain stable estimates. When the sample size is small or the data distribution is highly skewed, the fitted curve may exhibit instability.⁵¹ Therefore, knot selection and sensitivity analyses are important for the robustness and reliability of the findings.

Clinical prediction modeling

Diagnostic prediction models

Diagnostic prediction models estimate the probability that an individual currently has a specific health condition, usually a disease.⁵² Traditional diagnostic prediction models commonly use multivariable regression approaches, such as logistic regression for binary outcomes.^{53,54} More generally, predictive modeling follows a standardized workflow encompassing study design, data preprocessing, model development, and internal and external validation.⁵⁴

The model's performance is generally evaluated in three domains: discrimination, calibration, and clinical utility.

Table 1. Comparison of common methodological approaches in critical care database research

Methodological approach	Clinical scenarios	Advantages	Limitations and challenges
Descriptive and associational approaches			
Descriptive analysis	Describing microbiological findings and antibiotic use patterns in VAP ⁹	Simple and intuitive; easy to implement; provides basis for further analyses	Cannot infer causality; susceptible to selection bias; limited generalizability
Multivariable association and risk-factor analysis	Examining the associations of the Geriatric Nutritional Risk Index and early prophylactic heparin use with mortality in patients with sepsis ^{10,11}	Enables simultaneous examination of multiple variables; allows adjustment for potential confounders; quantifies associations using effect estimates with confidence intervals	Limited causal inference capacity in exploratory analyses without prespecified exposure and adjustment strategy
Nonlinear association modeling	Examining nonlinear associations of biomarkers with mortality in sepsis-associated acute kidney injury and urosepsis using restricted cubic splines ^{12,13}	Preserves continuous variable information; flexibly models nonlinear trends	Sensitive to the number and placement of knots; requires adequate sample size
Clinical prediction modeling			
Diagnostic prediction models	Early identification of bloodstream infection in critically ill patients using LASSO-based predictor selection, logistic regression, and a nomogram ¹⁴	Widely used for disease risk estimation and early identification of high-risk patients	Use of AUC alone may overestimate practical usefulness
Prognostic prediction models	Predicting short- and longer-term mortality in critically ill patients with coronary artery disease, acute pulmonary embolism, and COPD ¹⁵⁻¹⁷	Supports risk stratification and individualized prognostic assessment	Requires external validation before clinical implementation
Machine learning-based diagnostic prediction models	Early identification of sepsis-associated acute respiratory distress syndrome using machine learning models based on eICU and MIMIC-IV data ¹⁸	Can integrate multidimensional clinical information to generate individualized risk predictions and may support early detection and clinical decision-making	Risk of overfitting; risk of data leakage; requires external validation
Machine learning-based prognostic prediction models	Predicting new-onset atrial fibrillation in critically ill patients, critical outcomes in urinary tract infection, and prognosis in urosepsis and catheter-associated urinary tract infection ¹⁹⁻²²	May support prognostic risk stratification in complex patient populations	Survival time bias; competing risks; requires time-dependent performance metrics
Time-to-event, longitudinal, and dynamic methods			
Restricted mean survival time analysis	Evaluating survival time with early albumin plus crystalloid therapy in patients with sepsis ²³	Does not rely on proportional hazards assumption; results are clinically interpretable	Sensitive to truncation time (τ); may be unstable with insufficient follow-up
Longitudinal and joint modeling	Jointly modeling repeated daily SOFA scores and the competing risks of ICU death and discharge in patients with sepsis ²⁴	Captures temporal changes and within- and between-individual variability using repeated measurements; joint	Highly sensitive to data completeness; requires careful study design, appropriate measurement time point selection, and rigorous missing data handling

Table 1. Comparison of common methodological approaches in critical care database research (*continued*)

Methodological approach	Clinical scenarios	Advantages	Limitations and challenges
		models can link longitudinal and event-time processes while accounting for measurement error	
Trajectory analysis	Identifying longitudinal trajectories of lactate in sepsis, serum sodium in sepsis with lactic acidosis, and age-adjusted shock index in septic shock ²⁵⁻²⁷	Captures heterogeneity in longitudinal data	Sensitive to sample size, missing data, and the prespecified number of trajectory classes; requires robust model-selection procedures and sensitivity analyses
Causal inference approaches			
Propensity score methods and inverse probability of treatment weighting	Evaluating the associations of ramelteon exposure with survival in sepsis and albumin infusion with prognosis in ICU patients with cirrhosis and acute kidney injury ^{28,29}	Can improve balance in measured baseline covariates and reduce confounding due to measured characteristics	Sensitive to propensity-score model specification and extreme weights; cannot eliminate unmeasured confounding
Causal mediation analysis	Investigating potential mediating roles of inflammatory markers and metabolic or electrolyte-related biomarkers in associations with mortality ^{30,31}	Helps investigate potential mechanistic pathways	Causal interpretation requires strong assumptions regarding confounding, temporal ordering, and model specification; sensitivity analyses are important
Target trial emulation	Evaluating early albumin administration in relation to sepsis-associated acute kidney injury and corticosteroid treatment in patients with sepsis ^{32,33}	Aligns observational analyses with a prespecified target-trial framework and may reduce certain design- and time-related biases	Validity depends on data quality, variable availability, and accurate mapping of clinical pathways; prone to unmeasured confounding factors; prone to issues related to complex treatment pathways; robustness and credibility depend on well-specified trial protocols, meticulous operationalization of observational data, and comprehensive sensitivity analyses
Artificial intelligence and advanced computational methods			
Deep learning	Predicting ICU survival using multicohort data and estimating individualized mortality risks under different treatment strategies in patients with atrial fibrillation ^{34,35}	Applicable to outcome prediction and, when integrated with causal inference frameworks, individualized treatment-effect estimation	Requires large amounts of high-quality training data; limited interpretability due to complex, black-box model structures
Ensemble learning	Predicting sepsis-associated liver injury using ensemble learning with external multicenter validation ³⁶	Combines multiple base models and may improve predictive stability and generalizability	Increased computational complexity; risk of data leakage
Image-based deep learning	Classifying the progression of multiple thoracic abnormalities and localizing newly developed abnormalities on chest radiographs using MIMIC-CXR ³⁷	Can directly extract features from raw images for detection, classification, progression assessment, and lesion localization	Depends on large-scale, manually labeled datasets; labeling cost is high
Reinforcement	Optimizing sepsis treatment	Models sequential decision-	Limited prospective clinical evaluation;

Table 1. Comparison of common methodological approaches in critical care database research (*continued*)

Methodological approach	Clinical scenarios	Advantages	Limitations and challenges
learning	strategies using deep reinforcement learning with expert clinical knowledge ³⁸	making under uncertainty and may support personalized and adaptive treatment strategies	external validation and safety testing remain important
Natural language processing	Using clinical and nursing notes to predict hospital-acquired pressure injury through named-entity recognition ³⁹	Can process large volumes of unstructured clinical text; may support diagnostic and prognostic applications	Performance may vary across tasks, populations, and languages; limited by incomplete EHR data and poor interoperability with existing clinical systems

AUC, area under the receiver operating characteristic curve; COPD, chronic obstructive pulmonary disease; EHR, electronic health record; eICU, eICU Collaborative Research Database; ICU, intensive care unit; LASSO, least absolute shrinkage and selection operator; MIMIC-CXR, Medical Information Mart for Intensive Care Chest X-ray; MIMIC-IV, Medical Information Mart for Intensive Care IV; SOFA, Sequential Organ Failure Assessment; VAP, ventilator-associated pneumonia.

Discrimination reflects a model’s ability to distinguish between high- and low-risk individuals and is usually assessed using the area under the receiver operating characteristic curve (AUC) or the C-statistic. Calibration assesses agreement between predicted probabilities and observed outcome frequencies, typically via calibration plots. Decision curve analysis is frequently applied to assess clinical utility, estimating the net clinical benefit of using the model across various decision thresholds.⁵⁴ Although discrimination metrics are most often reported, using the AUC alone may overestimate the practical usefulness of a model. Therefore, a comprehensive evaluation that includes calibration and clinical utility is essential.^{54,55}

Predictive models are widely used in clinical database research for estimating disease risk and identifying high-risk patients early. For example, a MIMIC-IV-based study developed a nomogram for the early identification of bloodstream infection in critically ill patients using logistic regression with least absolute shrinkage and selection operator-based predictor selection. Model performance was assessed using discrimination, calibration, and decision curve analysis, with external validation in the eICU and an external hospital cohort.¹⁴

Prognostic prediction models

Prognostic prediction models are widely used in clinical research to estimate future risk of mortality, complications, and readmission. These models can support risk stratification and clinical decision-making. For example, MIMIC-based prognostic models have been developed for critically ill patients with coronary artery disease,¹⁵ acute pulmonary embolism,¹⁶ and chronic obstructive pulmonary disease¹⁷ to estimate short- and longer-term mortality risks. These models demonstrate how routinely collected clinical data can be used for individualized prognostic assessment and risk stratification.

These approaches may support more individualized risk assessment and adaptive strategies for clinical management. However, before such models can be used in clinical practice, they should undergo external validation using data from different populations that were not used during model development.⁵²

Machine learning-based diagnostic prediction models

ML is a major branch of AI that uses algorithmic learning to identify patterns and relationships in data without predefined rules.⁵⁶

ML approaches can generally be classified as supervised or unsupervised learning based on the learning paradigm. Supervised learning learns mappings from input features to outcomes from labeled data. It is used mainly for classification and regression tasks. In contrast, unsupervised learning finds latent structures in unlabeled data by clustering or dimensionality reduction, thus revealing hidden heterogeneity across patient populations.^{57,58}

During model development, datasets are often split into training and validation subsets to assess model performance and generalizability and reduce the risk of overfitting.⁵⁸ Compared with traditional regression-based methods, ML methods are less dependent on prespecified functional forms or distributional assumptions and instead emphasize predictive accuracy and pattern recognition.⁵⁷ ML algorithms can also integrate high-dimensional and heterogeneous clinical data, which can be advantageous in complex database environments, such as MIMIC and eICU.

ML methods have been widely used for disease prediction and early risk identification in clinical database research. For example, a study using the eICU and MIMIC-IV databases developed a ML diagnostic model for the early identification of sepsis-associated acute respiratory distress syndrome and evaluated several algorithms to distinguish affected from unaffected patients.¹⁸

Overall, ML-based diagnostic and predictive models can integrate multidimensional clinical information to generate individualized risk predictions and may support early detection and clinical decision-making.⁵⁷ However, special attention must still be paid to issues such as overfitting, data leakage, and external validation during model development and evaluation.

Machine learning-based prognostic prediction models

ML methods have also been widely employed for prognostic risk prediction in critically ill populations. For example, interpretable ML models have been developed to predict new-onset atrial fibril-

lation in critically ill patients using multicenter datasets.¹⁹ Other studies have used MIMIC-based data to predict critical outcomes in emergency department patients with urinary tract infection using algorithms such as XGBoost, random forests, and support vector machines,²⁰ and to develop interpretable prognostic models for patients with urosepsis and catheter-associated urinary tract infection.^{21,22} These studies illustrate the application of ML to risk stratification and outcome prediction in critically ill populations.

In summary, these studies suggest that ML-based prognostic models may support risk stratification and outcome prediction in heterogeneous critically ill populations. Such models may help inform clinical planning and healthcare resource allocation. However, special attention should be given to the methodological issues specific to prognostic modeling. These issues include survival time bias, competing risks, and the use of time-dependent performance metrics.

Time-to-event, longitudinal, and dynamic methods

Restricted mean survival time analysis

Restricted mean survival time (RMST) is an important summary measure in survival analysis that quantifies the average survival time from the start of a study to a pre-specified time point (τ). RMST is mathematically defined as the area under the survival curve within the specified follow-up interval.⁵⁹

Compared with traditional survival metrics, such as hazard ratios or median survival time, RMST has the major advantage of not requiring the proportional hazards assumption. As such, RMST can provide robust, interpretable comparisons between groups even when the proportional hazards assumption is violated.⁶⁰

In practical applications, the choice of τ is usually guided by clinical relevance or the objectives of the study design, for example, 2 years, 5 years, or the duration of ICU follow-up. Once chosen, τ should be kept constant throughout the analysis to avoid interpretive bias. RMST can be estimated directly from Kaplan-Meier survival curves or through regression-based methods (e.g., pseudo-value methods or rescaling techniques) to perform covariate-adjusted analyses.⁵⁹

RMST can be used to compare survival outcomes across therapeutic interventions or clinical characteristics in ICU database studies. For example, RMST was used in a study to compare survival time associated with early albumin plus crystalloid therapy in patients with sepsis.²³

The main advantage of RMST is that treatment effects can be expressed as absolute differences in survival time, which are often more clinically interpretable. However, this method has several limitations. Its results are sensitive to the choice of τ , and estimates may be unstable when follow-up is short or event rates are low.^{61,62}

Longitudinal and joint modeling

Longitudinal data refer to repeated measurements from the same subjects at different time points. Such data are used to characterize the temporal trajectories of variables and their associations with clinical outcomes.⁶³ Compared with analyses based on single measurements, longitudinal data analysis makes better use of time-dependent information and may improve the precision of risk assessment and prognostic prediction.

Joint modeling is one of the core methods in longitudinal data analysis. This framework generally consists of two interlinked components: a longitudinal submodel, which characterizes the temporal evolution of repeated measurements, and an event-time submodel, most commonly based on the Cox proportional hazards framework, which evaluates the influence of longitudinal changes on clinical outcomes. Joint models incorporate the unobserved true values of longitudinal processes into survival analysis to improve estimation accuracy and account for measurement error.⁶⁴

Joint modeling has been increasingly used for evaluating the association of dynamic physiological indicators with clinical outcomes in critical care database research. For example, a study of ICU patients with sepsis jointly modeled repeated daily Sequential Organ Failure Assessment scores and the competing risks of ICU death and discharge, showing that the evolution of Sequential Organ Failure Assessment scores was significantly associated with both risks and enabling dynamic prediction of ICU outcomes.²⁴

In summary, the advantage of longitudinal data analysis is the opportunity to quantify within- and between-individual variability simultaneously and to describe the temporal development of clinical variables. However, these methods are sensitive to data completeness because substantial missingness or irregular measurement intervals can adversely affect model stability and reliability.^{65,66} Thus, in practical applications, careful study design, appropriate selection of measurement time points, and rigorous handling of missing data are required.

Trajectory analysis

Trajectory analysis examines changes in one or more variables measured repeatedly over time or across age-related dimensions, such as behaviors, symptoms, or biomarkers.⁶⁷

Traditional mixed-effects models estimate average trajectories for the population, but trajectory analysis extends this concept by identifying subgroups with different developmental or temporal trajectories. These subgroups are often considered to represent different clinical “phenotypes” or developmental trajectories.⁶⁸

Trajectory analysis has been extensively applied in critical care database studies to identify dynamic clinical patterns and enable patient risk stratification. For example, MIMIC-based studies have used group-based trajectory modeling to characterize longitudinal patterns of lactate in patients with sepsis²⁵ and serum sodium in patients with sepsis and lactic acidosis.²⁶ In another study using both MIMIC-IV and eICU, 24-hour age-adjusted shock index trajectories were identified and shown to be associated with 30-day mortality in patients with septic shock.²⁷ These studies demonstrate how trajectory analysis can identify clinically significant heterogeneity in repeated measurements and inform prognostic stratification in critically ill populations.

The main advantage of trajectory analysis is that it identifies heterogeneity in longitudinal datasets and translates complex temporal patterns into clinically meaningful subgroup classifications. However, the robustness of the results depends on sample size, missing data, and the prespecified number of classes.⁶⁷ Therefore, robust model selection procedures and sensitivity analyses are needed to confirm the stability and validity of the findings.

Causal inference approaches

Propensity score methods and inverse probability of treatment weighting

Propensity score-based approaches such as inverse probability of treatment weighting (IPTW) are widely used to reduce measured confounding in retrospective cohort studies. IPTW is a causal inference method based on propensity scores used in observational studies to adjust for measured confounding and estimate causal effects of exposures or interventions under appropriate assumptions. The basis for IPTW is the concept of weighting individuals by their probability of exposure. Under adequate overlap, appropriate model specification, and relevant causal assumptions, IPTW can improve the balance of measured baseline covariates in the weighted sample.⁶⁹

In general, IPTW analysis includes several important steps. First, a propensity score model is constructed using measured confounders, with nonlinear or interaction terms added as needed to optimize model fit and covariate balance. Second, estimation instability is reduced by computing stabilized weights and truncating extreme weights when needed. Third, covariate balance is evaluated after weighting using standardized mean differences or graphical diagnostics to assess whether systematic differences in baseline characteristics remain between the exposure groups. Finally, weighted outcome models (e.g., weighted Cox proportional hazards regression or weighted linear regression) are fitted to estimate the causal effect of exposure or intervention on clinical outcomes.^{69,70}

IPTW has been widely used in critical care database research for treatment-effect estimation and exposure-outcome analyses. For example, MIMIC-IV-based studies have applied IPTW to examine the association between ramelteon exposure and survival outcomes in critically ill patients with sepsis,²⁸ as well as the association between albumin infusion and clinical outcomes in ICU patients with cirrhosis and acute kidney injury.²⁹ By creating weighted populations with more comparable baseline characteristics, IPTW can help reduce measured confounding in observational treatment-effect analyses.

The key advantage of IPTW is that it can approximate some features of randomized study designs, reduce measured confounding, and accommodate various outcome types. However, the validity of IPTW can be highly sensitive to the specification of the propensity score model. It is susceptible to extreme weights and cannot eliminate bias due to unmeasured confounding.⁷⁰

Causal mediation analysis

Mediation analysis seeks to examine the mechanisms by which an exposure or intervention may affect an outcome. Methodologically, mediation analysis breaks down the total effect of an exposure on an outcome into direct and indirect effects. The direct effect is the effect of the exposure on the outcome that is not mediated by the mediator whereas the indirect effect is the effect of the exposure on the outcome that is mediated by the mediator.⁷¹

To conceptualize the causal relationships between variables clearly, researchers often use directed acyclic graphs (DAGs). In DAGs, directed arrows represent the assumed causal direction, and the absence of cycles imposes an acyclic ordering on the hypothesized causal relationships.⁷² DAGs depict an investigator's hypotheses or assumptions about the systems that determine

exposure, the mechanisms by which exposure affects outcomes, and the factors that influence study inclusion.⁷³ DAGs distinguish three types of variables. A confounder causes both the exposure and the outcome, creating a non-causal association. A mediator lies on the causal pathway between the exposure and the outcome. A collider is a common effect of two variables on a path. If a collider is inadvertently adjusted for, it can introduce bias.⁷⁴

In mediation analysis, it is important to carefully consider multiple sources of confounding, including exposure-outcome confounding, exposure-mediator confounding, and mediator-outcome confounding.⁷⁵

Mediation analysis is a commonly used method for investigating potential biological and clinical mechanisms in studies utilizing clinical databases. For example, MIMIC-IV studies have applied causal mediation analysis to investigate the mediating role of inflammatory markers in the association between ondansetron pretreatment and mortality in mechanically ventilated ICU patients,³⁰ as well as the potential mediating effects of metabolic and electrolyte-related biomarkers in the relationship between the creatinine-to-albumin ratio and 28-day mortality after cardiac surgery.³¹

The core advantage of mediation analysis is that it helps identify potential mechanistic pathways thereby enhancing understanding of the causal processes linking exposures and outcomes.

Importantly, causal interpretations in mediation analyses are based on several strong assumptions that are more demanding than those required for conventional association analyses.⁷⁵ For the effect estimates from a mediation analysis to be interpreted as causal, researchers must adequately control for confounding. The assumptions include: (1) no unmeasured exposure-outcome confounding, (2) no unmeasured mediator-outcome confounding, (3) no unmeasured exposure-mediator confounding, and (4) no exposure-induced mediator-outcome confounding.⁷⁶ Although assumptions 1 and 3 are generally satisfied in randomized exposure studies, assumption 2 can still be violated because the mediator is usually not randomized.⁷⁶ Additionally, mediation analysis requires that the exposure, mediator, and outcome occur in a clear temporal order. Otherwise, the direction of the causal relationship cannot be reliably determined.⁷⁷ The use of DAGs based on subject-matter knowledge is recommended to support the identification and selection of relevant confounders.⁷⁸

However, the validity of its results depends heavily on adequate adjustment for confounders and appropriate model specification. Bias may still be introduced due to unmeasured or residual confounding. Mediation analyses are therefore often conducted within a formal causal inference framework and accompanied by sensitivity analyses to assess the robustness of the results.^{79,80}

Target trial emulation

Target trial emulation (TTE) is a framework for causal inference in the analysis of observational data. TTE seeks to improve the validity of causal inference by explicitly emulating a hypothetical randomized controlled trial, or "target trial," at the design and analysis stages when actual randomized trials are not feasible or ethical.⁸¹⁻⁸⁴ TTE usually consists of two main steps. First, the target trial is explicitly defined in terms of its key components. These include causal estimands, assumptions, eligibility criteria, treatment strategies, follow-up periods and analytical approaches.

Second, these predefined elements are mapped onto observational data for implementation and analysis.⁸⁵ Aligning observational analyses with the structure of a target trial may reduce selection bias and time-related biases, such as immortal time and survival bias.^{86,87}

TTE is increasingly applied in clinical database research to estimate treatment effects in real-world settings. For example, a MIMIC-IV-based TTE evaluated the association between early albumin administration and the risk of sepsis-associated acute kidney injury.³² More recently, a multicenter TTE using MIMIC, eICU, and an external clinical cohort estimated the effect of corticosteroid treatment on outcomes in patients with sepsis.³³ These studies illustrate the application of TTE to treatment-effect evaluation using observational critical care data.

The main advantage of TTE is its ability to systematically align observational analyses with the conceptual framework of randomized controlled trials, which may reduce certain design-related and time-related biases when its assumptions are adequately addressed. The validity of the findings depends critically on the quality of the underlying data, the availability of relevant variables, and accurate mapping of clinical pathways. In addition, TTE remains vulnerable to unmeasured confounding and challenges posed by complex treatment pathways. Therefore, the robustness and credibility of TTE-based findings depend on well-specified trial protocols, meticulous operationalization of observational data, and comprehensive sensitivity analyses.^{85,88}

Artificial intelligence and advanced computational methods

Deep learning

Deep learning encompasses computational models with multiple hierarchical processing layers that learn increasingly abstract representations of data. Deep learning is essentially a subset of the broader class of representation learning, where each nonlinear layer transforms the representation produced by the previous layer into a more abstract and informative form.⁸⁹ Deep learning methods are often applied to outcome prediction and treatment optimization in clinical database research. For example, a multicohort study developed and validated a deep learning model to predict ICU survival using MIMIC-III, eICU, and an external hospital dataset.³⁴

Deep learning approaches have also been increasingly integrated with causal inference frameworks for estimating individualized treatment effects extending beyond purely predictive applications. For example, a study using the Dragonnet architecture assessed individualized mortality risks under different treatment strategies for ICU patients with atrial fibrillation.³⁵ This work highlights the transition of deep learning from purely predictive tasks to broader applications in causal inference and precision medicine. Deep-learning models often require large amounts of high-quality training data and may have limited interpretability because of their complex, black-box nature, which can hinder their clinical implementation.⁹⁰

Ensemble learning

Ensemble learning is a set of methods that combine multiple base models to improve overall predictive performance. Combining different algorithms or similar models can reduce bias and variance, thereby improving model stability and generalizability.⁹¹

Common approaches include bagging (e.g., random forests), boosting (e.g., XGBoost and LightGBM), and stacking (model ensembles based on meta-learners).⁹¹

Ensemble learning has been widely used in risk prediction in clinical research. Its performance and generalizability have also been assessed through external validation in multicenter settings. For instance, a multicenter study of sepsis-associated liver injury externally validated an ensemble-learning model, although its discrimination was lower in the external validation cohort than in the training and internal validation cohorts.³⁶ However, ensemble learning also entails increased computational complexity and the risk of data leakage.⁹¹

Image-based deep learning

Medical images provide information about anatomical structures and pathological changes. They are important for lesion identification, disease diagnosis, treatment planning, and surgical procedures. In recent years, the rapid development of AI has advanced medical image analysis.⁹² These AI systems employ deep learning models that can learn complex patterns from raw images and apply them to new cases. Deep learning models have achieved substantial progress in medical image analysis⁹³ because of their strong feature-extraction capabilities and generalizability. In medical image analysis, models such as convolutional neural networks can directly extract features from raw images (e.g., X-ray, computed tomography, and magnetic resonance imaging) for lesion detection, classification, and severity estimation.⁹³

For example, a MIMIC-CXR-based study developed a weakly supervised deep learning model to classify the progression of multiple thoracic abnormalities on chest radiographs and to localize newly developed abnormalities for several of these conditions.³⁷ However, the success of deep learning often depends on large, manually labeled datasets, for which annotation can be costly.⁹²

Reinforcement learning

Reinforcement learning (RL) is a ML paradigm in which an agent learns to make optimal decisions through repeated interactions with the environment. RL aims to maximize the cumulative expected reward over time in the context of Markov decision processes.⁹⁴

In clinical research, RL has been used to optimize treatment strategies for critically ill patients. For example, therapeutic decision-making processes have been modeled for conditions such as sepsis by combining deep RL algorithms with expert clinical knowledge to derive personalized treatment strategies.³⁸ RL can model clinical uncertainty and sequential decision-making and may support personalized treatment strategies and real-time adaptation of interventions. In intensive care, the availability of high-granularity patient data makes RL a potentially useful framework for modeling patient states and optimizing treatment pathways; however, evidence from prospective clinical evaluation remains limited.⁹⁵ Further progress in this field will require external validation, publicly available source code, and prospective safety testing of RL.⁹⁶

Natural language processing

Natural language processing (NLP) is a set of computational methods that can be used to extract meaningful information from

unstructured text. In medical database research, NLP is often applied to clinical notes, nursing documentation, and biomedical literature to generate additional variables that augment structured datasets. NLP converts free-form text into machine-readable structured features, thereby facilitating subsequent statistical analysis and ML applications.⁹⁷ A typical NLP workflow consists of two main stages: text processing and classification. In text processing, EHR data⁹⁸ are analyzed to extract keywords, concepts, and risk factors related to diseases. Next, the extracted features are evaluated for their ability to discriminate among clinical states. Finally, informative variables are converted into structured representations for downstream analysis and predictive modeling.⁹⁹

In practice, NLP has been extensively applied to intensive care databases such as MIMIC. For example, Gu *et al.*³⁹ applied NLP-based named-entity recognition to clinical and nursing notes from MIMIC-III to predict hospital-acquired pressure injury, illustrating how unstructured ICU documentation can be transformed into features for predictive modeling. This approach can expand the information available from clinical databases and provide additional ways of examining clinical questions. NLP can process large volumes of unstructured text and may support diagnostic and prognostic applications, although its performance can vary across tasks, populations, and languages.¹⁰⁰ However, NLP algorithms face inherent linguistic challenges. For example, many algorithms struggle with negation; “does not smoke” could be misinterpreted as “smoker”. Their performance is also limited by incomplete EHR data and poor interoperability with existing clinical systems.⁹⁹

Common methodological issues and challenges in current research

Data-related issues

Large clinical databases such as MIMIC and eICU are critical resources for observational research and predictive modeling. However, data quality has a major influence on the reliability and validity of study findings. Missing data are among the most common challenges in these databases. Critical care databases often contain missing values in laboratory measurements, vital signs, and nursing notes. Data may be missing not at random. Thus, simple casewise deletion of missing observations can introduce substantial selection bias.¹⁰¹ Multiple imputation and model-based handling strategies can partially address this issue, but they may still compromise the stability of estimates in high-dimensional settings or when missingness is extensive. Studies should therefore clearly report the proportion of missing data, the assumptions made regarding the mechanism of missingness, and the data handling strategies employed.¹⁰²

A second major challenge is inconsistency in data standards. Differences in diagnostic coding practices, laboratory measurement units, and clinical documentation standards across healthcare institutions or ICU systems may impede variable consistency and limit comparability across studies.^{103,104}

In addition, precise temporal definitions are essential for longitudinal and causal inference analyses, particularly TTE. Imprecise definitions of exposure periods or follow-up windows may introduce immortal time bias or exposure misclassification

and thus substantially bias causal-effect estimates.⁸⁶

Missing data, inconsistent data standards, and inaccurate temporal definitions are among the greatest challenges in large-scale database research. To ensure reproducibility and scientific credibility, researchers should describe data sources, variable definitions, missing-data handling, and time windows in the Methods section. They should also address potential biases and their implications for interpretation.

Bias and confounding

Bias and confounding are major methodological challenges in large-scale clinical database studies, with direct implications for the validity of causal inference. Bias-related concerns include selection bias, information bias, and confounding. Confounding occurs when a third variable affects both the exposure and the outcome, thus distorting the estimated causal relationship. Information bias is usually linked to measurement errors, such as inaccurate diagnostic coding or incomplete clinical records. Selection bias occurs when there are systematic differences between the study sample and the target population.¹⁰⁵

These issues occur at all stages of the research process, including study design, data preprocessing, statistical analysis, and interpretation of results. Common strategies to mitigate bias include clearly defining inclusion and exclusion criteria to minimize selection bias, rigorously standardizing exposure, outcome, and covariate definitions to reduce information bias, and using approaches such as multivariable regression, propensity score methods, IPTW, instrumental variable analysis, or TTE to address confounding. In addition, sensitivity analyses should be performed to evaluate the robustness of the findings. Potential sources of bias and their possible impact should be reported transparently.

Issues related to statistical analysis and modeling

In large-scale database research, the statistical analyses and models themselves can introduce important methodological problems. The most common problems are the inappropriate variable selection and the lack of model validation.

For example, in high-dimensional data settings, inadequate selection of candidate variables may lead to substantial multicollinearity, which in turn results in unstable parameter estimates and degraded predictive performance. Regularization techniques commonly used for variable selection and dimensionality reduction include least absolute shrinkage and selection operator regression and elastic net, which may improve model robustness and generalizability.¹⁰⁶

Second, many studies are limited by poor model validation, especially the lack of external validation and incomplete reporting of model performance metrics. For example, some studies report only discrimination measures, such as AUC, but do not report calibration performance and clinical utility, thereby limiting assessment of the model’s clinical applicability.¹⁰⁷ Model reliability can be improved using internal validation methods such as bootstrap resampling or cross-validation during model development. Where possible, researchers should also conduct external validation. In addition, studies should report discrimination, calibration, and overall model performance metrics systematically and comprehensively.¹⁰⁷

Issues specific to causal inference

Causal inference in database research presents several specific methodological challenges. The primary limitation is unmeasured confounding. While approaches such as propensity score adjustment and IPTW can enhance balance on observed covariates, they cannot account for unmeasured confounders. This may result in biased estimates of the causal effect.⁷⁰

Temporal biases, especially immortal time bias and survival bias, are also common when exposures change over time or when the intervention starts at different time points. For example, if exposure is defined as “receiving medication during hospitalization,” patients who die before treatment begins may be excluded from the exposed group, which can lead to an overestimation of the treatment effect.⁸⁶

To improve the validity of causal inference, researchers should clearly define exposures, outcomes, and temporal windows, select analytical methods carefully, evaluate covariate balance rigorously, and conduct sensitivity analyses to evaluate the potential impact of unmeasured confounding.

Future directions and recommendations

As research using large clinical databases expands, future progress may center on three major areas: data integration, methodological refinement, and multimodal applications.

First, analyses based on a single database, such as MIMIC or eICU, are insufficient to ensure generalizability. Therefore, cross-database integration and multicenter collaborative analysis will be important for improving external validity and evaluating transportability. For instance, Chen *et al.*¹⁰⁸ used an “in-house development plus dual external validation” framework integrating data from the Peking Union Medical College Hospital cohort, MIMIC-IV, and eICU to assess the robustness of the blood pressure response index as a marker of vasopressor responsiveness, illustrating multi-database collaborative modeling.

Future cross-database studies should clearly report the following for each data source: database version, data collection period, cohort selection criteria, and variable extraction procedures.¹⁰⁹ Before combining datasets from different sources, researchers should address heterogeneity at the syntactic, structural, and semantic levels by aligning variable definitions, coding systems, units of measurement, and temporal frameworks.¹¹⁰

Second, refining the methodology should entail more than adopting individual statistical techniques. It should also promote closer alignment among research questions, analytical frameworks, and relevant reporting standards. From a causal inference perspective, frameworks such as TTE and DAGs have considerable potential to improve study design, clarify causal assumptions, and minimize biases resulting from inappropriate temporal alignment or covariate adjustment. However, these frameworks are underutilized in critical care database research. Therefore, their broader and more rigorous application should be encouraged to improve the interpretability of causal observational findings. Meanwhile, the increasing use of advanced analytical methods should be accompanied by reporting guidelines appropriate to the study design and analytical objective. For instance, retrospective cohort studies should adhere to the

STROBE statement,¹¹¹ and those based on routinely collected health data should also follow the RECORD extension.¹⁰⁹ Studies explicitly emulating a target trial should adhere to the TARGET statement,⁸⁵ and those developing clinical prediction models using regression or machine-learning methods should adhere to the TRIPOD+AI statement.¹¹² For example, a MIMIC-IV-based study used a TTE framework combined with the clone-censor-weighting approach to evaluate the association between early albumin administration and the risk of sepsis-associated acute kidney injury, while addressing time-related biases such as immortal time bias and prevalent-user bias.³²

Multimodal AI has emerged as a promising approach to critical care database research. Healthcare data are inherently multimodal, including structured EHRs, medical imaging, physiological waveforms, and unstructured clinical text. Each of these provides complementary information about a patient’s clinical status.^{113,114} Notably, multimodal data integration may improve predictive performance and more closely approximate human expert decision-making by providing a more valid basis for clinical decisions.¹¹⁵ For example, one study using the MIMIC-CXR database used a deep learning and transfer learning approach to automatically extract the labels from the radiology reports via weakly supervised learning. This method may help identify dynamic changes in pulmonary lesions and provide useful information for clinical decision-making.³⁷

Despite these promising findings, the number of studies developing multimodal AI models in clinical settings remains limited. Unimodal approaches continue to dominate the field.¹¹⁶ The adoption of multimodal AI in medical research also faces several challenges, including incomplete and heterogeneous data, as well as limited standardization and model interpretability.^{116,117} Furthermore, critical care EHR databases pose additional challenges, including missingness, irregular sampling, and potential measurement or classification errors, which may introduce information bias.^{118,119}

To address the challenges, future efforts should focus on several areas. First, data quality and standardization should be strengthened through rigorous preprocessing and harmonization. Second, model interpretability and clinical trustworthiness should be enhanced. More studies are adopting explainable ML techniques, such as Shapley additive explanations, to elucidate model predictions. Such interpretability is important for fostering clinician acceptance and enabling safe deployment.^{120,121} Third, rigorous external validation is important. External validation helps assess whether models generalize beyond their development cohorts and perform reliably across diverse populations and clinical settings, facilitating the transition from single-center to multicenter applications.¹²² Targeted solutions are required to address the specific challenges posed by critical care databases. To address high missingness and irregular sampling, deep learning architectures that directly model irregularly sampled time series may complement conventional imputation methods such as multiple imputation.^{123,124} Additionally, causal inference frameworks, such as TTE, can be used to address time-varying treatments and time-dependent confounding. These frameworks structure observational analyses to emulate randomized trials, potentially reducing biases, including immortal time and selection bias.¹²⁵

Strengths and limitations

This review integrates commonly used and emerging methods across five pragmatic domains and emphasizes method selection, validation, and reporting considerations. It also has several limitations. First, because it is a narrative review rather than a systematic one, the selection of topics was guided by the existing literature and the authors' expertise rather than by a formal search and selection protocol. Therefore, relevant methods or studies may have been omitted. Second, the review focuses primarily on methods commonly or increasingly used in critical care database research. It does not cover all specialized causal inference or emerging computational approaches in depth. Third, while clinical text, medical imaging, and multimodal data are discussed, the review focuses more heavily on structured clinical data and provides less detailed coverage of physiological waveforms and other high-frequency data. Fourth, the five-domain framework is intended to be a practical organizational structure, rather than mutually exclusive categories, so some methods may fit within more than one domain. Finally, space constraints precluded a detailed discussion of implementation procedures, diagnostics, and software-specific considerations.

Conclusions

Future database research should focus on continued advances in data integration, refinement of causal inference methods, standardization of statistical modeling, and expansion of multimodal AI applications to improve the scientific rigor and translational value of large-scale clinical database research.

Acknowledgments

Not applicable.

Funding

This work was supported by the Science and Technology Projects in Guangzhou (2025A03J4246).

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

Writing - original draft (SK), methodology (SK), conceptualization (SK), writing - original draft (LC), methodology (LC), literature search and selection (LC); literature extraction and organization (HL), literature search and selection (HL); literature search and selection (QL); literature search and selection (XL); writing - review and editing (YW), supervision (YW), conceptualization (YW); writing - review and editing (JL), supervision (JL), conceptualization (JL). All authors have approved the final version and publication of the manuscript.

References

- [1] Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, *et al.* MIMIC-IV, a freely accessible electronic health record

dataset. *Sci Data* 2023;10(1):1. DOI: 10.1038/s41597-022-01899-x, PMID: 36596836.

- [2] Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, *et al.* MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;3:160035. DOI: 10.1038/sdata.2016.35, PMID: 27219127.
- [3] 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;5:180178. DOI: 10.1038/sdata.2018.178, PMID: 30204154.
- [4] Ke Y, Yang R, Liu N. Comparing Open-Access Database and Traditional Intensive Care Studies Using Machine Learning: Bibliometric Analysis Study. *J Med Internet Res* 2024;26:e48330. DOI: 10.2196/48330, PMID: 38630522.
- [5] Kallout J, Lamer A, Grosjean J, Kerdelhué G, Bouzillé G, Clavier T, *et al.* Contribution of Open Access Databases to Intensive Care Medicine Research: Scoping Review. *J Med Internet Res* 2025;27:e57263. DOI: 10.2196/57263, PMID: 39787600.
- [6] Shillan D, Sterne JAC, Champneys A, Gibbison B. Use of machine learning to analyse routinely collected intensive care unit data: a systematic review. *Crit Care* 2019;23(1):284. DOI: 10.1186/s13054-019-2564-9, PMID: 31439010.
- [7] Wu WT, Li YJ, Feng AZ, Li L, Huang T, Xu AD, *et al.* Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Mil Med Res* 2021;8(1):44. DOI: 10.1186/s40779-021-00338-z, PMID: 34380547.
- [8] Heavner SF, Kumar VK, Anderson W, Al-Hakim T, Dasher P, Armaignac DL, *et al.* Critical Data for Critical Care: A Primer on Leveraging Electronic Health Record Data for Research From Society of Critical Care Medicine's Panel on Data Sharing and Harmonization. *Crit Care Explor* 2024;6(11):e1179. DOI: 10.1097/CCE.0000000000001179, PMID: 39559555.
- [9] Liu Q, Yang J, Zhang J, Zhao F, Feng X, Wang X, *et al.* Description of Clinical Characteristics of VAP Patients in MIMIC Database. *Front Pharmacol* 2019;10:62. DOI: 10.3389/fphar.2019.00062, PMID: 30778301.
- [10] Li L, Lu X, Qin S, Huang D. Association between geriatric nutritional risk index and 28 days mortality in elderly patients with sepsis: a retrospective cohort study. *Front Med (Lausanne)* 2023;10:1258037. DOI: 10.3389/fmed.2023.1258037, PMID: 37822466.
- [11] Zou ZY, Huang JJ, Luan YY, Yang ZJ, Zhou ZP, Zhang JJ, *et al.* Early prophylactic anticoagulation with heparin alleviates mortality in critically ill patients with sepsis: a retrospective analysis from the MIMIC-IV database. *Burns Trauma* 2022;10:tkac029. DOI: 10.1093/burnst/tkac029, PMID: 36168402.
- [12] Xu H, Mo R, Liu Y, Niu H, Cai X, He P. L-shaped association between triglyceride-glucose body mass index and short-term mortality in ICU patients with sepsis-associated acute kidney injury. *Front Med (Lausanne)* 2024;11:1500995. DOI: 10.3389/fmed.2024.1500995, PMID: 39712174.
- [13] Yu Y, He Z, Li W, Wang K, Zhu D, Zhang L, *et al.* Red cell distribution width to albumin ratio predicts short-term mortality in urosepsis: a dual-cohort study. *Front Nutr* 2026;13:1709663. DOI: 10.3389/fnut.2026.1709663, PMID: 41743068.
- [14] Qi Z, Dong L, Lin J, Duan M. Development and validation a nomogram prediction model for early diagnosis of bloodstream infections in the intensive care unit. *Front Cell Infect Microbiol* 2024;14:1348896. DOI: 10.3389/fcimb.2024.1348896, PMID: 38500500.
- [15] Tao Y, Huang G, Huang M, Yao Q, Wang Z, Han L, *et al.* A study on

- predicted in-hospital mortality in critically ill patients with coronary heart disease: analysis of the MIMIC-IV database. *BMC Med Inform Decis Mak* 2025;26(1):21. DOI: 10.1186/s12911-025-03319-7, PMID: 41413523.
- [16] Ding CW, Liu C, Zhang ZP, Cheng CY, Pei GS, Jing ZC, *et al.* Development and external validation of a nomogram for predicting short-term prognosis in patients with acute pulmonary embolism. *Int J Cardiol* 2024;407:132065. DOI: 10.1016/j.ijcard.2024.132065, PMID: 38642720.
 - [17] Guo Y, Zuo C, Yan J, Ban C. Nomogram model of mortality risk in patients with chronic obstructive pulmonary disease in intensive care unit: based on MIMIC-IV database and external validation study. *Front Med (Lausanne)* 2025;12:1547047. DOI: 10.3389/fmed.2025.1547047, PMID: 40766059.
 - [18] Bai Y, Xia J, Huang X, Chen S, Zhan Q. Using machine learning for the early prediction of sepsis-associated ARDS in the ICU and identification of clinical phenotypes with differential responses to treatment. *Front Physiol* 2022;13:1050849. DOI: 10.3389/fphys.2022.1050849, PMID: 36579020.
 - [19] Guan C, Gong A, Zhao Y, Yin C, Geng L, Liu L, *et al.* Interpretable machine learning model for new-onset atrial fibrillation prediction in critically ill patients: a multi-center study. *Crit Care* 2024;28(1):349. DOI: 10.1186/s13054-024-05138-0, PMID: 39473013.
 - [20] Yen CC, Ma CY, Tsai YC. Interpretable Machine Learning Models for Predicting Critical Outcomes in Patients with Suspected Urinary Tract Infection with Positive Urine Culture. *Diagnostics (Basel)* 2024;14(17):1974. DOI: 10.3390/diagnostics14171974, PMID: 39272758.
 - [21] Wei Y, Xu W, Yang S, Zhang C, Wang J, Wan X. Significant adverse prognostic events in patients with urosepsis: a machine learning based model development and validation study. *Front Cell Infect Microbiol* 2025;15:1623109. DOI: 10.3389/fcimb.2025.1623109, PMID: 40861482.
 - [22] Liu L, Yu X, Chen Z, Zhang Q, Zhuang D. Predicting mortality in intensive care unit patients with CAUTI using an interpretable machine learning model: a retrospective cohort study from MIMIC-IV database. *Front Med (Lausanne)* 2025;12:1665035. DOI: 10.3389/fmed.2025.1665035, PMID: 40995080.
 - [23] Zhou S, Zeng Z, Wei H, Sha T, An S. Early combination of albumin with crystalloids administration might be beneficial for the survival of septic patients: a retrospective analysis from MIMIC-IV database. *Ann Intensive Care* 2021;11(1):42. DOI: 10.1186/s13613-021-00830-8, PMID: 33689042.
 - [24] Lavalley-Morelle A, Timsit JF, Mentré F, Mullaert J, OUTCOMEREA network. Joint modeling under competing risks: application to survival prediction in patients admitted in Intensive Care Unit for sepsis with daily Sequential Organ Failure Assessment score assessments. *CPT Pharmacometrics Syst Pharmacol* 2022;11(11):1472–1484. DOI: 10.1002/psp4.12856, PMID: 36201150.
 - [25] Wei Y, Zhuang J, Li J, Wang Z, Wang J, Zhang X, *et al.* Lactate trajectories and outcomes in patients with sepsis in the intensive care unit: group-based trajectory modeling. *Front Public Health* 2025;13:1610220. DOI: 10.3389/fpubh.2025.1610220, PMID: 40589821.
 - [26] Li H, Zhou Q, Nan Y, Liu C, Zhang Y. Group-based Trajectory Modeling of Serum Sodium and Survival in Sepsis Patients with Lactic Acidosis: results from MIMIC-IV Database. *Tohoku J Exp Med* 2025;265(3):123–134. DOI: 10.1620/tjem.2024.J091, PMID: 39261081.
 - [27] Yue S, Hou X, Wang Y, Xu Z, Li X, Wang J, *et al.* Influence of age-adjusted shock index trajectories on 30-day mortality for critical patients with septic shock. *Front Med (Lausanne)* 2025;12:1534706. DOI: 10.3389/fmed.2025.1534706, PMID: 40417677.
 - [28] Han YY, Tian Y, Zhao BC, Liu KX. Ramelteon exposure and survival of critically ill sepsis patients: a retrospective study from MIMIC-IV. *BMC Anesthesiol* 2024;24(1):454. DOI: 10.1186/s12871-024-02851-9, PMID: 39696003.
 - [29] Li M, Ge Y, Wang J, Chen W, Li J, Deng Y, *et al.* Impact of albumin infusion on prognosis in ICU patients with cirrhosis and AKI: insights from the MIMIC-IV database. *Front Pharmacol* 2024;15:1467752. DOI: 10.3389/fphar.2024.1467752, PMID: 39434907.
 - [30] Zhou S, Tao L, Zhang Z, Zhang Z, An S. Mediators of neutrophil-lymphocyte ratio in the relationship between ondansetron pre-treatment and the mortality of ICU patients on mechanical ventilation: causal mediation analysis from the MIMIC-IV database. *Br J Clin Pharmacol* 2022;88(6):2747–2756. DOI: 10.1111/bcp.15204, PMID: 34964162.
 - [31] Shi P, Rui S, Meng Q. Association between serum creatinine-to-albumin ratio and 28-day mortality in intensive care unit patients following cardiac surgery: analysis of MIMIC-IV data. *BMC Cardiovasc Disord* 2025;25(1):100. DOI: 10.1186/s12872-025-04505-1, PMID: 39953440.
 - [32] Li XY, Chen WS, Qu ZK, Chen JG, Li L, Li SN, *et al.* Early use of albumin may increase the risk of sepsis-associated acute kidney injury in sepsis patients: a target trial emulation. *Mil Med Res* 2025;12(1):51. DOI: 10.1186/s40779-025-00641-z, PMID: 40841985.
 - [33] Rajendran S, Xu Z, Pan W, Zang C, Siempos I, Torres L, *et al.* Multicenter target trial emulation to evaluate corticosteroids for sepsis stratified by predicted organ dysfunction trajectory. *Nat Commun* 2025;16(1):4450. DOI: 10.1038/s41467-025-59643-z, PMID: 40360520.
 - [34] Tang H, Jin Z, Deng J, She Y, Zhong Y, Sun W, *et al.* Development and validation of a deep learning model to predict the survival of patients in ICU. *J Am Med Inform Assoc* 2022;29(9):1567–1576. DOI: 10.1093/jamia/ocac098, PMID: 35751440.
 - [35] Kang MW, Ahn SY, Kang Y. Determining optimal strategies for personalized atrial fibrillation treatment in intensive care unit patients using a deep learning-based causal inference approach: rhythm and/or rate control. *J Am Med Inform Assoc* 2026;33(3):679–689. DOI: 10.1093/jamia/ocaf203, PMID: 41317034.
 - [36] Lei J, Zhai J, Zhang Y, Qi J, Sun C. Supervised Machine Learning Models for Predicting Sepsis-Associated Liver Injury in Patients With Sepsis: development and validation study based on a multicenter cohort study. *J Med Internet Res* 2025;27:e66733. DOI: 10.2196/66733, PMID: 40418571.
 - [37] Yu K, Ghosh S, Liu Z, Deible C, Poynton CB, Batmanghelich K. Anatomy-specific Progression Classification in Chest Radiographs via Weakly Supervised Learning. *Radiol Artif Intell* 2024;6(5):e230277. DOI: 10.1148/ryai.230277, PMID: 39046325.
 - [38] Wu X, Li R, He Z, Yu T, Cheng C. A value-based deep reinforcement learning model with human expertise in optimal treatment of sepsis. *NPJ Digit Med* 2023;6(1):15. DOI: 10.1038/s41746-023-00755-5, PMID: 36732666.
 - [39] Gu S, Lee EW, Zhang W, Simpson RL, Hertzberg VS, Ho JC. Evaluating Natural Language Processing Packages for Predicting Hospital-Acquired Pressure Injuries From Clinical Notes. *Comput Inform Nurs* 2024;42(3):184–192. DOI: 10.1097/CIN.0000000000001053, PMID: 37607706.
 - [40] Klebanoff MA, Snowden JM. Historical (retrospective) cohort

- studies and other epidemiologic study designs in perinatal research. *Am J Obstet Gynecol* 2018;219(5):447–450. DOI: 10.1016/j.ajog.2018.08.044, PMID: 30194051.
- [41] Yu H, Chen S. Association between anion gap and the 30-day mortality of patients with ventilator-associated pneumonia: a study of the MIMIC-III database. *J Thorac Dis* 2024;16(5):2994–3006. DOI: 10.21037/jtd-23-1735, PMID: 38883665.
- [42] Xu J, Tong L, Yao J, Guo Z, Lui KY, Hu X, *et al.* Association of Sex With Clinical Outcome in Critically Ill Sepsis Patients: A Retrospective Analysis of the Large Clinical Database MIMIC-III. *Shock* 2019;52(2):146–151. DOI: 10.1097/SHK.0000000000001253, PMID: 30138298.
- [43] Cai W, Xu J, Wu X, Chen Z, Zeng L, Song X, *et al.* Association between triglyceride-glucose index and all-cause mortality in critically ill patients with ischemic stroke: analysis of the MIMIC-IV database. *Cardiovasc Diabetol* 2023;22(1):138. DOI: 10.1186/s12933-023-01864-x, PMID: 37312120.
- [44] Lieberman OJ, Lee S, Zabinski J. Donepezil treatment is associated with improved outcomes in critically ill dementia patients via a reduction in delirium. *Alzheimers Dement* 2023;19(5):1742–1751. DOI: 10.1002/alz.12807, PMID: 36218161.
- [45] Wang X, Kattan MW. Cohort Studies: Design, Analysis, and Reporting. *Chest* 2020;158(1S):S72–S78. DOI: 10.1016/j.chest.2020.03.014, PMID: 32658655.
- [46] Grimes DA, Schulz KF. Descriptive studies: what they can and cannot do. *Lancet* 2002;359(9301):145–149. DOI: 10.1016/S0140-6736(02)07373-7, PMID: 11809274.
- [47] Kesmodel US. Cross-sectional studies - what are they good for? *Acta Obstet Gynecol Scand* 2018;97(4):388–393. DOI: 10.1111/aogs.13331, PMID: 29453895.
- [48] Grant SW, Hickey GL, Head SJ. Statistical primer: multivariable regression considerations and pitfalls. *Eur J Cardiothorac Surg* 2019;55(2):179–185. DOI: 10.1093/ejcts/ezy403, PMID: 30596979.
- [49] Lewer D, Brothers T, O'Nions E, Pickavance J. Factors associated with: problems of using exploratory multivariable regression to identify causal risk factors. *BMJ Med* 2025;4(1):e001375. DOI: 10.1136/bmjmed-2025-001375, PMID: 41181819.
- [50] Discacciati A, Palazzolo MG, Park JG, Melloni GEM, Murphy SA, Bellavia A. Estimating and presenting non-linear associations with restricted cubic splines. *Int J Epidemiol* 2025;54(4):dyaf088. DOI: 10.1093/ije/dyaf088, PMID: 40527479.
- [51] Desquilbet L, Mariotti F. Dose-response analyses using restricted cubic spline functions in public health research. *Stat Med* 2010;29(9):1037–1057. DOI: 10.1002/sim.3841, PMID: 20087875.
- [52] van Smeden M, Reitsma JB, Riley RD, Collins GS, Moons KG. Clinical prediction models: diagnosis versus prognosis. *J Clin Epidemiol* 2021;132:142–145. DOI: 10.1016/j.jclinepi.2021.01.009, PMID: 33775387.
- [53] Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162(1):W1–W73. DOI: 10.7326/M14-0698, PMID: 25560730.
- [54] Wynants L, Collins GS, Van Calster B. Key steps and common pitfalls in developing and validating risk models. *BJOG* 2017;124(3):423–432. DOI: 10.1111/1471-0528.14170, PMID: 27362778.
- [55] Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol* 2019;110:12–22. DOI: 10.1016/j.jclinepi.2019.02.004, PMID: 30763612.
- [56] Gautam N, Ghanta SN, Clausen A, Saluja P, Sivakumar K, Dhar G, *et al.* Contemporary Applications of Machine Learning for Device Therapy in Heart Failure. *JACC Heart Fail* 2022;10(9):603–622. DOI: 10.1016/j.jchf.2022.06.011, PMID: 36049812.
- [57] Gautam N, Mueller J, Alqaisi O, Gandhi T, Malkawi A, Tarun T, *et al.* Machine Learning in Cardiovascular Risk Prediction and Precision Preventive Approaches. *Curr Atheroscler Rep* 2023;25(12):1069–1081. DOI: 10.1007/s11883-023-01174-3, PMID: 38008807.
- [58] Oikonomou EK, Khera R. Machine learning in precision diabetes care and cardiovascular risk prediction. *Cardiovasc Diabetol* 2023;22(1):259. DOI: 10.1186/s12933-023-01985-3, PMID: 37749579.
- [59] Hasegawa T, Misawa S, Nakagawa S, Tanaka S, Tanase T, Ugai H, *et al.* Restricted mean survival time as a summary measure of time-to-event outcome. *Pharm Stat* 2020;19(4):436–453. DOI: 10.1002/pst.2004, PMID: 32072769.
- [60] Uno H, Claggett B, Tian L, Fu H, Huang B, Kim DH, *et al.* Adding a new analytical procedure with clinical interpretation in the tool box of survival analysis. *Ann Oncol* 2018;29(5):1092–1094. DOI: 10.1093/annonc/mdy109, PMID: 29617717.
- [61] Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, *et al.* Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol* 2014;32(22):2380–2385. DOI: 10.1200/JCO.2014.55.2208, PMID: 24982461.
- [62] Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol* 2013;13:152. DOI: 10.1186/1471-2288-13-152, PMID: 24314264.
- [63] Hu J, Szymczak S. A review on longitudinal data analysis with random forest. *Brief Bioinform* 2023;24(2):bbad002. DOI: 10.1093/bib/bbad002, PMID: 36653905.
- [64] Asar Ö, Ritchie J, Kalra PA, Diggle PJ. Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *Int J Epidemiol* 2015;44(1):334–344. DOI: 10.1093/ije/dyu262, PMID: 25604450.
- [65] Li Y, Feng D, Sui Y, Li H, Song Y, Zhan T, *et al.* Analyzing longitudinal binary data in clinical studies. *Contemp Clin Trials* 2022;115:106717. DOI: 10.1016/j.cct.2022.106717, PMID: 35240309.
- [66] Ouko RK, Mukaka M, Ohuma EO. Joint modelling of longitudinal data: a scoping review of methodology and applications for non-time to event data. *BMC Med Res Methodol* 2025;25(1):40. DOI: 10.1186/s12874-025-02485-6, PMID: 39962370.
- [67] Nagin DS, Jones BL, Elmer J. Recent Advances in Group-Based Trajectory Modeling for Clinical Research. *Annu Rev Clin Psychol* 2024;20(1):285–305. DOI: 10.1146/annurev-clinpsy-081122-012416, PMID: 38382118.
- [68] Herle M, Micali N, Abdulkadir M, Loos R, Bryant-Waugh R, Hübel C, *et al.* Identifying typical trajectories in longitudinal data: modelling strategies and interpretations. *Eur J Epidemiol* 2020;35(3):205–222. DOI: 10.1007/s10654-020-00615-6, PMID: 32140937.
- [69] Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med* 2015;34(28):3661–3679. DOI: 10.1002/sim.6607, PMID: 26238958.
- [70] Chesnaye NC, Stel VS, Tripepi G, Dekker FW, Fu EL, Zoccali C, *et al.* An introduction to inverse probability of treatment weighting in

- observational research. *Clin Kidney J* 2022;15(1):14–20. DOI: 10.1093/ckj/sfab158, PMID: 35035932.
- [71] Tönnies T, Schlesinger S, Lang A, Kuss O. Mediation Analysis in Medical Research. *Dtsch Arztebl Int* 2023;120(41):681–687. DOI: 10.3238/arztebl.m2023.0175, PMID: 37584228.
- [72] Williamson EJ, Aitken Z, Lawrie J, Dharmage SC, Burgess JA, Forbes AB. Introduction to causal diagrams for confounder selection. *Respirology* 2014;19(3):303–311. DOI: 10.1111/resp.12238, PMID: 24447391.
- [73] Digitale JC, Martin JN, Glymour MM. Tutorial on directed acyclic graphs. *J Clin Epidemiol* 2022;142:264–267. DOI: 10.1016/j.jclinepi.2021.08.001, PMID: 34371103.
- [74] Gaskell A, Sleigh J. DAGs for dummies: how to extract causation from correlation. *Br J Anaesth* 2025;135(4):837–839. DOI: 10.1016/j.bja.2025.07.001, PMID: 40707288.
- [75] VanderWeele TJ. Mediation Analysis: A Practitioner's Guide. *Annu Rev Public Health* 2016;37:17–32. DOI: 10.1146/annurev-publheal-032315-021402, PMID: 26653405.
- [76] Li Y, Yoshida K, Kaufman JS, Mathur MB. A brief primer on conducting regression-based causal mediation analysis. *Psychol Trauma* 2023;15(6):930–938. DOI: 10.1037/tra0001421, PMID: 36701540.
- [77] Schuler MS, Coffman DL, Stuart EA, Nguyen TQ, Vegetabile B, McCaffrey DF. Practical challenges in mediation analysis: a guide for applied researchers. *Health Serv Outcomes Res Methodol* 2025;25(1):57–84. DOI: 10.1007/s10742-024-00327-4, PMID: 39958808.
- [78] Tennant PWG, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, *et al.* Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol* 2021;50(2):620–632. DOI: 10.1093/ije/dyaa213, PMID: 33330936.
- [79] Albert JM, Li Y, Sun J, Woyczynski WA, Nelson S. Continuous-time causal mediation analysis. *Stat Med* 2019;38(22):4334–4347. DOI: 10.1002/sim.8300, PMID: 31286536.
- [80] Rudolph KE, Williams NT, Diaz I. Practical causal mediation analysis: extending nonparametric estimators to accommodate multiple mediators and multiple intermediate confounders. *Biostatistics* 2024;25(4):997–1014. DOI: 10.1093/biostatistics/kxae012, PMID: 38576206.
- [81] Hernán MA, Dahabreh IJ, Dickerman BA, Swanson SA. The Target Trial Framework for Causal Inference From Observational Data: Why and When Is It Helpful? *Ann Intern Med* 2025;178(3):402–407. DOI: 10.7326/ANNALS-24-01871, PMID: 39961105.
- [82] Dahabreh IJ, Bibbins-Domingo K. Causal Inference About the Effects of Interventions From Observational Studies in Medical Journals. *JAMA* 2024;331(21):1845–1853. DOI: 10.1001/jama.2024.7741, PMID: 38722735.
- [83] Hernán MA, Wang W, Leaf DE. Target Trial Emulation: a framework for causal inference from observational data. *JAMA* 2022;328(24):2446–2447. DOI: 10.1001/jama.2022.21383, PMID: 36508210.
- [84] Hernán MA. Methods of Public Health Research - Strengthening Causal Inference from Observational Data. *N Engl J Med* 2021;385(15):1345–1348. DOI: 10.1056/NEJMp2113319, PMID: 34596980.
- [85] Cashin AG, Hansford HJ, Hernán MA, Swanson SA, Lee H, Jones MD, *et al.* Transparent Reporting of Observational Studies Emulating a Target Trial-The TARGET Statement. *JAMA* 2025;334(12):1084–1093. DOI: 10.1001/jama.2025.13350, PMID: 40899949.
- [86] Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol* 2016;79:70–75. DOI: 10.1016/j.jclinepi.2016.04.014, PMID: 27237061.
- [87] Hernán MA, Sterne JAC, Higgins JPT, Shrier I, Hernández-Díaz S. A Structural Description of Biases That Generate Immortal Time. *Epidemiology* 2025;36(1):107–114. DOI: 10.1097/EDE.0000000000001808, PMID: 39494894.
- [88] Ren Y, Jia Y, Liu L, Lyv H, Tao L, Li Y, *et al.* Design and Implementation of Observational Studies Emulating a Target Trial. *JAMA Netw Open* 2026;9(2):e2558262. DOI: 10.1001/jamanetworkopen.2025.58262, PMID: 41712213.
- [89] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–444. DOI: 10.1038/nature14539, PMID: 26017442.
- [90] Jang HJ, Cho KO. Applications of deep learning for the analysis of medical data. *Arch Pharm Res* 2019;42(6):492–504. DOI: 10.1007/s12272-019-01162-9, PMID: 31140082.
- [91] Mahajan P, Uddin S, Hajati F, Moni MA. Ensemble Learning for Disease Prediction: A Review. *Healthcare (Basel)* 2023;11(12):1808. DOI: 10.3390/healthcare11121808, PMID: 37372925.
- [92] Wang H, Jin Q, Li S, Liu S, Wang M, Song Z. A comprehensive survey on deep active learning in medical image analysis. *Med Image Anal* 2024;95:103201. DOI: 10.1016/j.media.2024.103201, PMID: 38776841.
- [93] Zhou SK, Greenspan H, Davatzikos C, Duncan JS, van Ginneken B, Madabhushi A, *et al.* A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proc IEEE Inst Electr Electron Eng* 2021;109(5):820–838. DOI: 10.1109/JPROC.2021.3054390, PMID: 37786449.
- [94] Tan RK, Liu Y, Xie L. Reinforcement learning for systems pharmacology-oriented and personalized drug design. *Expert Opin Drug Discov* 2022;17(8):849–863. DOI: 10.1080/17460441.2022.2072288, PMID: 35510835.
- [95] Jayaraman P, Desman J, Sabounchi M, Nadkarni GN, Sakhuja A. A Primer on Reinforcement Learning in Medicine for Clinicians. *NPJ Digit Med* 2024;7(1):337. DOI: 10.1038/s41746-024-01316-0, PMID: 39592855.
- [96] Otten M, Jagesar AR, Dam TA, Biesheuvel LA, den Hengst F, Ziesemer KA, *et al.* Does Reinforcement Learning Improve Outcomes for Critically Ill Patients? A systematic review and level-of-readiness assessment. *Crit Care Med* 2024;52(2):e79–e88. DOI: 10.1097/CCM.0000000000006100, PMID: 37938042.
- [97] Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, *et al.* Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA* 2011;306(8):848–855. DOI: 10.1001/jama.2011.1204, PMID: 21862746.
- [98] Afzal N, Sohn S, Abram S, Scott CG, Chaudhry R, Liu H, *et al.* Mining peripheral arterial disease cases from narrative clinical notes using natural language processing. *J Vasc Surg* 2017;65(6):1753–1761. DOI: 10.1016/j.jvs.2016.11.031, PMID: 28189359.
- [99] Eguía H, Sánchez-Bocanegra CL, Vinciarelli F, Alvarez-Lopez F, Saigí-Rubió F. Clinical Decision Support and Natural Language Processing in Medicine: Systematic Literature Review. *J Med Internet Res* 2024;26:e55315. DOI: 10.2196/55315, PMID: 39348889.
- [100] Le Glaz A, Haralambous Y, Kim-Dufor DH, Lenca P, Billot R, Ryan TC, *et al.* Machine Learning and Natural Language Processing in Mental Health: systematic review. *J Med Internet Res* 2021;23(5):e15708.

- DOI: 10.2196/15708, PMID: 33944788.
- [101] Perez-Lebel A, Varoquaux G, Le Morvan M, Josse J, Poline JB. Benchmarking missing-values approaches for predictive models on health databases. *Gigascience* 2022;11:giac013. DOI: 10.1093/gigascience/giac013, PMID: 35426912.
- [102] Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, *et al.* Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393. DOI: 10.1136/bmj.b2393, PMID: 19564179.
- [103] Penev YP, Buchanan TR, Ruppert MM, Liu M, Shekouhi R, Guan Z, *et al.* Electronic Health Record Data Quality and Performance Assessments: Scoping Review. *JMIR Med Inform* 2024;12:e58130. DOI: 10.2196/58130, PMID: 39504136.
- [104] Shi X, Zhai Y, Yu X, Li X, Hazlehurst BL, Nyongesa DB, *et al.* Statistical methods to harmonize electronic health record data across healthcare systems: case study and lessons learned. *Bioinformatics* 2026;42(3):bttag107. DOI: 10.1093/bioinformatics/btag107, PMID: 41769828.
- [105] Brown JP, Hunnicutt JN, Ali MS, Bhaskaran K, Cole A, Langan SM, *et al.* Core Concepts in Pharmacoepidemiology: Quantitative Bias Analysis. *Pharmacoepidemiol Drug Saf* 2024;33(10):e70026. DOI: 10.1002/pds.70026, PMID: 39375940.
- [106] Sanchez-Pinto LN, Venable LR, Fahrenbach J, Churpek MM. Comparison of variable selection methods for clinical predictive modeling. *Int J Med Inform* 2018;116:10–17. DOI: 10.1016/j.ijmedinf.2018.05.006, PMID: 29887230.
- [107] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594. DOI: 10.1136/bmj.g7594, PMID: 25569120.
- [108] Chen Y, Jiang H, Wei Y, Qiu Y, Su L, Chen J, *et al.* Blood pressure response index and clinical outcomes in patients with septic shock: a multicenter cohort study. *EBioMedicine* 2024;106:105257. DOI: 10.1016/j.ebiom.2024.105257, PMID: 39059317.
- [109] Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, *et al.* The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015;12(10):e1001885. DOI: 10.1371/journal.pmed.1001885, PMID: 26440803.
- [110] Cheng C, Messerschmidt L, Bravo I, Waldbauer M, Bhavikatti R, Schenk C, *et al.* A General Primer for Data Harmonization. *Sci Data* 2024;11(1):152. DOI: 10.1038/s41597-024-02956-3, PMID: 38297013.
- [111] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007;147(8):573–577. DOI: 10.7326/0003-4819-147-8-200710160-00010, PMID: 17938396.
- [112] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, *et al.* TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385:e078378. DOI: 10.1136/bmj-2023-078378, PMID: 38626948.
- [113] Mohsen F, Ali H, El Hajj N, Shah Z. Artificial intelligence-based methods for fusion of electronic health records and imaging data. *Sci Rep* 2022;12(1):17981. DOI: 10.1038/s41598-022-22514-4, PMID: 36289266.
- [114] Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med* 2022;28(9):1773–1784. DOI: 10.1038/s41591-022-01981-2, PMID: 36109635.
- [115] Kline A, Wang H, Li Y, Dennis S, Hutch M, Xu Z, *et al.* Multimodal machine learning in precision health: A scoping review. *NPJ Digit Med* 2022;5(1):171. DOI: 10.1038/s41746-022-00712-8, PMID: 36344814.
- [116] Schouten D, Nicoletti G, Dille B, Chia C, Vendittelli P, Schuurmans M, *et al.* Navigating the landscape of multimodal AI in medicine: A scoping review on technical challenges and clinical applications. *Med Image Anal* 2025;105:103621. DOI: 10.1016/j.media.2025.103621, PMID: 40482561.
- [117] Liu J, Cen X, Yi C, Wang FA, Ding J, Cheng J, *et al.* Challenges in AI-driven Biomedical Multimodal Data Fusion and Analysis. *Genomics Proteomics Bioinformatics* 2025;23(1):qzaf011. DOI: 10.1093/gpb/jnl/qzaf011, PMID: 40036568.
- [118] Papapanagiotou I, Karalis A, Kokkoris S, Vrettou CS, Kampouroupoulou O, Giannopoulou V, *et al.* Machine learning for early detection and prediction of sepsis: explainability and key sepsis biomarkers representation-A systematic review. *Int J Med Inform* 2026;214:106420. DOI: 10.1016/j.ijmedinf.2026.106420, PMID: 41962401.
- [119] Bots SH, Groenwold RHH, Dekkers OM. Using electronic health record data for clinical research: a quick guide. *Eur J Endocrinol* 2022;186(4):E1–E6. DOI: 10.1530/EJE-21-1088, PMID: 35157608.
- [120] Alkhanbouli R, Matar Abdulla Almadhaani H, Alhosani F, Simsekler MCE. The role of explainable artificial intelligence in disease prediction: a systematic literature review and future research directions. *BMC Med Inform Decis Mak* 2025;25(1):110. DOI: 10.1186/s12911-025-02944-6, PMID: 40038704.
- [121] Caterson J, Lewin A, Williamson E. The application of explainable artificial intelligence (XAI) in electronic health record research: A scoping review. *Digit Health* 2024;10:20552076241272657. DOI: 10.1177/20552076241272657, PMID: 39493635.
- [122] Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol* 2016;69:245–247. DOI: 10.1016/j.jclinepi.2015.04.005, PMID: 25981519.
- [123] Sun C, Song M, Cai D, Zhang B, Li H, Hong S. A Review of Deep Learning Methods for Irregularly Sampled Medical Time Series Data. *Health Data Sci* 2026;6:0456. DOI: 10.34133/hds.0456, PMID: 42088656.
- [124] Ren W, Liu Z, Wu Y, Zhang Z, Hong S, Liu H; Missing Data in Electronic health Records (MINDER) Group. Moving Beyond Medical Statistics: A Systematic Review on Missing Data Handling in Electronic Health Records. *Health Data Sci* 2024;4:0176. DOI: 10.34133/hds.0176, PMID: 39635227.
- [125] Reep CAT, Wils EJ, Heunks L. Opportunities, challenges and future perspectives for target trial emulation in critical care clinical research. *Crit Care* 2025;29(1):484. DOI: 10.1186/s13054-025-05723-x, PMID: 41225646.