BMJ Open

Predictive Modelling Methods of Hospital Readmission Risks for Patients with Chronic Obstructive Pulmonary Disease (COPD): A Systematic Review Protocol

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-093771
Article Type:	Protocol
Date Submitted by the Author:	16-Sep-2024
Complete List of Authors:	Tonde, Balkissa; Université Laval Faculté des Sciences de l'Administration, Department of Operations and Decision Systems; VITAM Center for Sustainable Health Research Traore, Metogara; Université Laval Faculté des Sciences de l'Administration, Department of Management; VITAM Center for Sustainable Health Research Landa, Paolo; Laval University, Operations and Decision Systems Department; Institut Universitaire de Cardiologie et de Pneumologie de Québec - Université Laval Côté, André; Université Laval, Département de management; VITAM Center for Sustainable Health Research, Laberge, Maude; Universite Laval Faculte de medecine, Social and Preventive Medicine; VITAM Center for Sustainable Health Research, Université Laval
Keywords:	Risk Factors, Hospitalization, Machine Learning, Pulmonary Disease, Chronic Obstructive, Prognosis

Title of the Manuscript:

Predictive Modelling Methods of Hospital Readmission Risks for Patients with Chronic Obstructive Pulmonary Disease (COPD): A Systematic Review Protocol

ABSTRACT

Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a significant chronic respiratory condition characterized by persistent airway obstruction, leading to substantial morbidity and mortality worldwide. Patients with COPD frequently experience hospital readmissions shortly after discharge, mainly due to acute exacerbations. This review evaluates and compare the effectiveness of various predictive modelling methods for hospital readmissions in COPD patients, focusing on recent advances, challenges, and limitations of prediction methods.

Method and analysis

This systematic review will adhere to PRISMA-P reporting guidelines. The review will include studies that develop or validate predictive models for hospital readmissions in COPD patients. We will perform a comprehensive search on PubMed, Embase, Cochrane Library, IEEE Xplore, Web of Science, and Google Scholar. Eligible studies will include those utilizing any predictive modelling method, focusing on unplanned readmissions within specified timeframes (30, 60, or 90 days). Two independent reviewers will screen titles, abstracts, and full texts, and will extract data using a standardized form. The methodological quality and risk of bias will be assessed using the Prediction Model Risk Of Bias Assessment Tool (PROBAST).

The results will be synthesized narratively, and a meta-analysis will be performed if the studies are sufficiently homogenous.

Ethics and dissemination

This research is a systematic review of published studies and does not involve direct patient data collection, thus not requiring ethics approval. Findings will be disseminated through publication in peer-reviewed journals and presentations at national and international conferences.

Prospero registration number

This protocol is registered with PROSPERO (**CRD42024579524**).

Keywords: COPD, hospital readmission, predictive modelling, risk factors

STRENGTHS AND LIMITATIONS OF A STUDY

⇒ This study's strengths lie in its comprehensive approach to systematically reviewing predictive models for hospital readmissions in COPD patients.

⇒ By directly comparing traditional statistical methods with contemporary machine learning techniques, this review assesses their effectiveness in predicting hospital readmissions among COPD patients, while highlighting the recent advances, challenges, and inherent limitations of these predictive modeling approaches.

⇛ The use of rigorous methodologies, applying of the CHARMS checklist and PROBAST tool, ensures a thorough assessment of the quality and risk of bias in the included studies.

⇒ The anticipated heterogeneity in study designs, patient populations, and predictive models may challenge the synthesis of results and limit the generalizability of the findings.

INTRODUCTION

According to the World Health Organization, Chronic Obstructive Pulmonary Disease (COPD) is a major chronic respiratory condition characterized by persistent airway obstruction, contributing to significant morbidity and mortality worldwide and accounting for nearly 5% of deaths in 2021[1]. Affecting over three million individuals globally, COPD imposes a substantial burden on both individuals and healthcare systems, contributing to high direct and indirect costs, frequent medical consultations, emergency visits, and hospitalizations[2,3,4]. A recent study by Chen et al. (2023)[5] estimates that COPD will cost the global economy USD 4.326 trillion between 2020 and 2050, representing an annual tax of 0.111% on global GDP. This highlights the substantial economic impact of the disease across different countries and regions. Moreover, patients with COPD often experience hospital readmissions shortly after discharge, primarily due to acute exacerbations. These readmissions highlight gaps in care management and coordination during transitions between care levels[6,7].

Frequent and often preventable readmissions significantly increase healthcare costs and negatively affect the quality of care provided[8,9,10]. For instance, in the United States, COPD readmissions are responsible for more than $15 billion annually in direct healthcare costs alone[8], representing up to 75% of the total costs associated with managing COPD exacerbations[11].

While progress has been made in understanding the factors contributing to these readmissions, the persistently high rates underscore the need for more effective prevention strategies and improved patient monitoring post-discharge. Indeed, enhanced coordination between hospital services and primary care, along with personalized management plans, has demonstrated the ability to reduce readmissions and improve patient clinical outcomes[12,13]. In response to the significant financial burden and high rates of readmissions associated with COPD, predictive modeling efforts have been undertaken to identify determinants of these readmissions, with the aim of improving interventions and care planning[14,15]. Predictive modelling is "the process of applying a statistical model or data mining algorithm to data to predict new observations or future observations"[16]. Current studies[14,17,18] reveal a wide variety of methods used to predict readmissions, reflecting the complexity of COPD and the multiple factors influencing patient outcomes.

Conventional approaches, such as LACE, PEARL, Elixhauser, and HOSPITAL indices [18,19,20], have been widely adopted in clinical practice. The LACE index predicts 30-day readmissions or death based on length of stay, acuity of admission, comorbidity (via the Charlson index), and emergency visits in the last six months[19]. The HOSPITAL score, identifies patients at risk of avoidable readmissions, focusing on hospital procedures and previous admissions[20]. The PEARL score predicts 90-day readmission or death in COPD based on five criteria: previous admissions, eMRCD score, age, and right- and left-sided heart failure[18]. The Elixhauser comorbidity index assesses assesses 29 comorbidities to predict hospital mortality and readmissions[21]. While these tools are useful, they often fail to fully address the complexity of COPD and its social determinants, which can limit their effectiveness in specific contexts or populations[18].

Machine learning models are emerging as promising alternatives that outperform traditional methods by leveraging electronic health record data for dynamic and precise risk analysis of readmissions. These advanced technologies enable a richer and more complex modelling of the interactions between risk factors due to their ability to integrate multiple dimensions and handle vast datasets of health information[14,15,17]. Several studies[14,15,17] have highlighted the potential of these models in terms of precision and their capacity to model the complex interactions between risk factors for hospital readmissions compared to traditional methods. However, each method may not be applicable or effective in every context, which highlights the importance of systematically comparing the effectiveness of these approaches, as well as their respective strengths and weaknesses.

This review aims to explore the current state of predictive modeling for hospital readmission risks in patients with COPD, with a particular focus on comparing traditional models to machine learning approaches. Unlike previous reviews, which may have focused on a limited range of methods or specific periods, this review systematically incorporates both heuristic and statistical approaches without any restriction on the period. By highlighting recent advances in the field, as well as the challenges and limitations of these prediction methods, we aim to provide a thorough evaluation of the effectiveness of various predictive models. Additionally, this review discusses the potential transition toward mixed prediction methods, which combine the strengths of both traditional and machine learning approaches, to improve accuracy and applicability in clinical settings. By synthesizing existing evidence, this review will offer valuable insights into the effectiveness of various predictive models and guide the transition towards more advanced, mixed-method prediction models.

The findings will inform future research and aid in identifying individuals at highest risk of readmission which could be the target of interventions to reduce hospital readmissions. Identifying these patients should enhance health outcomes and alleviate the burden on healthcare systems.

This systematic review aims to address the following primary question: What is the effectiveness of the different predictive modelling methods for hospital readmission risks in patients with COPD?

OBJECTIVES

The objectives of this review are to :

1.Identify the different predictive modelling methods used to predict readmission risks for patients with COPD.

2. Analyze and compare the performance and limitations of traditional predictive models with those using machine learning approaches.

3. Identify the strengths and weaknesses of using each method.

METHODS AND ANALYSIS

This review will follow a rigorous methodology outlined in the Cochrane Handbook for Systematic Reviews of Interventions[22,23], with a protocol registered in PROSPERO (number **CRD42024579524**).

Eligibility criteria

Complies with PRISMA-P guidelines[24,25], this systematic review will examine the following eligibility criteria based on the PICOS approach[26], which consists of the population, intervention/exposure, comparator, outcomes, and study design. The selection of these criteria is based on a brief review of the literature, including systematic evaluations of predictive models for hospital readmissions[27,28,29]. This approach ensures that the most relevant elements are included to effectively address the research question. Inclusion and exclusion criteria are described in table 1

Table 1 Summary of inclusion and exclusion criteria

Criteria	Inclusion	Exclusion
Population	Adult patients discharged after a hospital admission where the final diagnosis was COPD.	Studies on other populations. Studies not specifically concerning COPD readmissions. Studies focusing on elective readmissions (e.g. readmission after specific surgery).
Intervention	Any studies reporting the development/ validation of a readmission prediction model including the use of heuristics approach or statistical methods (machine learning, or other analytical techniques).	Studies on the risks of COPD readmissions without a focus on predictive modelling. Studies that do not provide a clear and detailed description of the variables used in the predictive models.
Comparator	No restriction on comparator type	NA

Outcomes	Unplanned readmission within 30, 60 or 90 days following patient discharge. Ability of models to predict hospital readmission risk (C-statistic, accuracy of predictive models, sensitivity, specificity, calibration, and validation of models). Associated risk factors include clinical variables, comorbidities, health behaviors, etc.	Outcomes not related to unplanned readmission or predictive modelling.
Study design	Quantitative empirical studies, including prospective and retrospective cohort studies, randomized controlled trials, case-controlled studies, model validation studies, and studies utilizing panel data.	Retracted publications, qualitative or mixed-methods studies, and non-empirical material such as knowledge syntheses, letters, perspectives, editorial, theses, conference abstracts, study protocols, books, book chapters, and scoping reviews.
Language	English or French	Other languages
Time	No restrictions	NA
Context	Healthcare studies related to hospital readmission for COPD.	Not health-related or not related to hospital readmission for COPD

Information sources

To identify studies on predictive models for COPD readmissions, an exhaustive search will be conducted on the following electronic databases: PubMed, Embase, Cochrane Library, IEEE Xplore, and Web of Science. The search period will cover from the inception of each database to the date of the search, in both English and French. In addition, a search for grey literature will be conducted using Google Scholar to include articles not published in traditional databases.

Search strategy

Our search strategy employs both MeSH terms and keywords to ensure comprehensive coverage of relevant studies. Key terms such as "patient readmission" and "rehospitalization" are specifically targeted to capture studies focused on hospital readmissions. Additionally, the strategy includes terms associated with predictive modeling, such as "Regression Analysis," "Algorithms," "Prognosis," "Support Vector Machine," and "machine learning." For COPD, we will use specific terms including "chronic obstructive pulmonary disease," "COPD," "emphysema," and "chronic bronchitis." We will also incorporate MeSH terms related to "risk" along with keywords such as "risk factors" and "risk assessment." These search criteria will be systematically combined using Boolean operators (AND, OR) to refine the final set of references in each database. The references cited in the included studies will also be reviewed to identify additional relevant articles. A proposed search strategy for PubMed is detailed in the Appendix 1.

Data collection and management

Selection of studies

Citations will be imported into the online systematic review platform Covidence (Veritas Health Innovation, Australia)[30,31], where duplicates will be automatically removed prior to the detailed study analysis. Two independent reviewers will conduct every step and meet to discuss discrepancies. If consensus is not reached, then a third investigator will adjudicate. The first step will consist of screening titles and abstracts to assess relevance based on predefined inclusion and exclusion criteria. We will then review the full article of each retained reference. At this step, we will note the reason for each excluded article.

Data collection and extraction process

A double data extraction will be performed for all included articles. Two reviewers will independently extract data from each article using a standardized data extraction form (Appendix 2) based on the CHARMS checklist (Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies)[32]. The data extraction form will be piloted on a 10% sample of articles and adjusted if necessary. Any disagreements in data extraction will be discussed between the two reviewers and, if unresolved, will be adjudicated by a third party .

The extracted data will include the following key elements:

- **Source and study design:** study title, authors, publication year, and country, along with the data source type (e.g., cohort, case-control).
- **Participant characteristics:** participant descriptions, including sample size and demographic details such as age, gender, and recruitment methods.

- **Modelling methods:** type of predictive modelling approach used in the study, as well as the analyses for the total population.
- **Handling of Missing Data:** Presence of missing data (on outcomes and/or variables) and method of handling missing data
- **Outcome measures:** Primary and secondary outcomes related to COPD readmission, with a focus on how these outcomes are measured and reported.
- **Predictive variables and model inputs:** number and names of predictive variables used (demographics, clinical), and measurement method.
- **Model development and validation:** Modeling methods used, selection of predictors, methods employed for model validation, (internal/external), performance metrics (C-index, sensitivity, specificity, calibration).
- **Results and interpretation:** final model's results, potential biases, limitations, and assumptions.

Risk of bias in individual studies

The risk of bias will be assessed using the Prediction model Risk Of Bias Assessment Tool (PROBAST)[33,34], which is specifically designed for systematic reviews of diagnostic prediction models (Appendix 3). The assessment will focus on several key domains, including participants, predictors, outcomes, analyses, and overall methodological rigour. Each domain will be evaluated and classified as "high risk of bias," "unclear risk of bias," or "low risk of bias." Any discrepancy or conflict that arises during the assessment will be resolved through consensus discussions, and if necessary, with the involvement of a third reviewer.

Data synthesis

The synthesis will focus on both qualitative and quantitative aspects, acknowledging the potential heterogeneity across studies due to differences in study populations, model types, and implementation contexts.

Narrative synthesis

We will begin with a narrative synthesis to summarize and interpret the findings from the included studies. This approach allows us to describe the characteristics and performance of the predictive models in a detailed and systematic manner. We will highlight the key features of each study, including the population characteristics, the type of predictive models used, and the context in which the models were developed and validated. The narrative synthesis will also address the following critical aspects:

- Model discrimination: We will assess how well each model distinguishes between patients who are readmitted and those who are not. This will include evaluating metrics such as the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) and other relevant performance indicators.
- Cohort type: We will compare the types of COPD patient cohorts used across studies, considering factors such as the inclusion criteria, severity of the disease, and the presence of comorbidities.
- Practical implementation: We will examine the practical aspects of implementing these predictive models in clinical settings, including the ease of use, data availability, and the resources required for the implementation.

- Types of variables included: We will categorize and compare the types of variables included in the predictive models (e.g., demographic factors, clinical variables, biomarkers), as the choice of variables can significantly impact model accuracy and generalizability.
- Model performance indicators: Beyond discrimination, we will explore other performance indicators such as calibration, sensitivity, specificity, and predictive values.

Handling heterogeneity

Due to expected heterogeneity in study design, patient populations, and model types, a meta-analysis will only be conducted if sufficient homogeneity is present. In line with Kansagara & al. (2011)[27], we will qualitatively explore heterogeneity by synthesizing the differences and similarities in model discrimination, COPD patient cohort types, implementation practicalities, included variables, and performance indicators. This approach allows for a thorough and context-specific assessment of predictive models.

ETHICS AND DISSEMINATION

This research project is based on the analysis of published studies and does not involve direct patient data collection; hence no ethics approval is required. However, the study will adhere to ethical principles of scientific rigour, transparency, and respect for data sources. The findings will be disseminated through publication in scientific journals, presentations at national and international conferences, and specialized forums in health sciences and research methodology. This approach will maximize the impact of the findings and support evidence-informed public health policies.

Patient and public involvement

Not applicable.

DISCUSSION

The predictive modelling of hospital readmissions for COPD patients is an evolving field that has seen significant advancements, transitioning from traditional methods to sophisticated machine learning approaches. Each of these methodologies comes with its own set of strengths and weaknesses.

Traditional predictive tools, such as the LACE, PEARL, Elixhauser, and HOSPITAL indices, have been extensively used in clinical settings. These tools, while robust, often fall short of capturing the multifaceted nature of COPD patient care. For instance, the Elixhauser index, despite its comprehensive inclusion of comorbidities, is complex and lacks generalizability across diverse patient populations[8]. Similarly, the HOSPITAL score requires adjustments to reflect individual COPD patient characteristics, including social determinants of health[14]. COPD-specific indices like the PEARL score offer promising perspectives due to their specificity for acute COPD exacerbations but require rigorous evaluation for long-term application[18]. Furthermore, while general indices like LACE sometimes outperform disease-specific indices, they are not consistently reliable for predicting readmissions in COPD patients, highlighting the intricate nature of this task[35].

The adoption of more sophisticated modelling methods, such as machine learning techniques, offers a compelling alternative to traditional models, addressing some of their limitations.

Standard machine learning models, such as penalized logistic regression, random forests, and gradient-boosting decision trees, demonstrate an enhanced ability to process and analyze complex datasets, leading to a deeper understanding of readmission risk factors[14]. Additionally, deep machine learning models, such as convolutional and recurrent neural networks, provide nuanced insights into medical data sequences, which are crucial for understanding COPD's progression and outcomes[14]. However, these approaches demand large volumes of high-quality data and specialized expertise for their implementation, which can be a barrier in some clinical environments[17].

In conclusion, this research will assess and synthesize the current state of predictive modelling for hospital readmission risks in patients with COPD. By comparing traditional predictive models with emerging machine learning approaches, this review aims to provide a comprehensive understanding of the strengths and limitations of various methodologies. The insights gained may guide the development of more effective and context-sensitive predictive models that can improve the management of COPD patients, reduce hospital readmissions, and ease the burden on healthcare systems. The findings from this review could inform future research and support the transition towards integrating advanced prediction methods in clinical practice, offering personalized interventions that improve patient outcomes and optimize the use of healthcare resources.

Amendments

Any protocol amendments will be documented and made publicly accessible on the PROSPERO registration. Each amendment will include the date, a detailed description, and the rationale for the changes made.

Ethics approval and patient consent for publication

Not applicable.

REFERENCES

1. The top 10 causes of death, https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death (accessed 3 September 2024).

2. Momtazmanesh S, Moghaddam SS, Ghamari S-H, et al. Global burden of chronic respiratory diseases and risk factors, 1990–2019: an update from the Global Burden of Disease Study 2019. *eClinicalMedicine*; 59. Epub ahead of print 1 May 2023. DOI: 10.1016/j.eclinm.2023.101936.

3. Alshabanat A, Otterstatter MC, Sin DD, et al. Impact of a COPD comprehensive case management program on hospital length of stay and readmission rates. *Int J Chron Obstruct Pulmon Dis* 2017; 12: 961–971.

4. Alqahtani JS, Aldabayan YS, Aldhahir AM, et al. Predictors of 30- and 90-Day COPD Exacerbation Readmission: A Prospective Cohort Study. *Int J Chron Obstruct Pulmon Dis* 2021; 16: 2769–2781.

5. Chen S, Cao Z, Prettner K, et al. Estimates and Projections of the Global Economic Cost of 29 Cancers in 204 Countries and Territories From 2020 to 2050. *JAMA Oncol* 2023; 9: 465.

6. Fernández-García S, Represas-Represas C, Ruano-Raviña A, et al. Social and clinical predictors of short- and long-term readmission after a severe exacerbation of copd. *PLoS One* 2020; 15: e0229257.

7. Lindenauer PK, Dharmarajan K, Qin L, et al. Risk Trajectories of Readmission and Death in the First Year after Hospitalization for Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 2018; 197: 1009–1017.

8. Buhr RG, Jackson NJ, Kominski GF, et al. Readmission Rates for Chronic Obstructive Pulmonary Disease Under the Hospital Readmissions Reduction Program: an Interrupted Time Series Analysis. *J Gen Intern Med* 2020; 35: 3581–3590.

9. Shah AA, Devana SK, Lee C, et al. Machine learning driven identification of novel patient factors for prediction of major complications after posterior cervical spinal fusion. *Eur Spine J* 2022; 31: 1952–1959.

10. Njoku CM, Alqahtani JS, Wimmer BC, et al. Risk factors and associated outcomes of hospital readmission in COPD: A systematic review. *Respiratory Medicine* 2020; 173: 105988.

11. Jacobs DM, Noyes K, Zhao J, et al. Early Hospital Readmissions after an Acute Exacerbation of Chronic Obstructive Pulmonary Disease in the Nationwide Readmissions Database. *Ann Am Thorac Soc* 2018; 15: 837–845.

12. Adamson SL, Burns J, Camp PG, et al. Impact of individualized care on readmissions after a hospitalization for acute exacerbation of COPD. *Int J Chron Obstruct Pulmon Dis* 2016; 11: 61–71.

13. Bricard D, Or Z. Impact of early primary care follow-up after discharge on hospital readmissions. *Eur J Health Econ* 2019; 20: 611–623.

14. Min X, Yu B, Wang F. Predictive Modeling of the Hospital Readmission Risk from Patients' Claims Data Using Machine Learning: A Case Study on COPD. *Sci Rep* 2019; 9: 2362.

15. Li M, Cheng K, Ku K, et al. Modelling 30-day hospital readmission after discharge for COPD patients based on electronic health records. *npj Prim Care Respir Med* 2023; 33: 1–8.

16. Shmueli G. To Explain or to Predict? *Statist Sci*; 25. Epub ahead of print 1 August 2010. DOI: 10.1214/10-STS330.

17. Bonomo M, Hermsen MG, Kaskovich S, et al. Using Machine Learning to Predict Likelihood and Cause of Readmission After Hospitalization for Chronic Obstructive Pulmonary Disease Exacerbation. *Int J Chron Obstruct Pulmon Dis* 2022; 17: 2701–2709.

18. Echevarria C, Steer J, Heslop-Marshall K, et al. The PEARL score predicts 90-day readmission or death after hospitalisation for acute exacerbation of COPD. *Thorax* 2017; 72: 686–693.

19. van Walraven C, Dhalla IA, Bell C, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *CMAJ* 2010; 182: 551–557.

20. Donzé JD, Williams MV, Robinson EJ, et al. International Validity of the HOSPITAL Score to Predict 30-Day Potentially Avoidable Hospital Readmissions. *JAMA Intern Med* 2016; 176: 496–502.

21. Moore BJ, White S, Washington R, et al. Identifying Increased Risk of Readmission and In-hospital Mortality Using Hospital Administrative Data: The AHRQ Elixhauser Comorbidity Index. *Medical Care* 2017; 55: 698–705.

22. Higgins JPT, Cochrane Collaboration (eds). *Cochrane handbook for systematic reviews of interventions*. Second edition. Hoboken, NJ: Wiley-Blackwell, 2019.

23. Higgins JP, Green S. Guide to the Contents of a Cochrane Protocol and Review. In: *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons, Ltd, pp. 51–79.

24. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021; 372: n71.

25. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015; 349: g7647.

26. Methley AM, Campbell S, Chew-Graham C, et al. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Services Research* 2014; 14: 579.

27. Kansagara D, Englander H, Salanitro A, et al. Risk Prediction Models for Hospital Readmission: A Systematic Review. *JAMA* 2011; 306: 1688–1698.

28. Bellou V, Belbasis L, Konstantinidis AK, et al. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. *BMJ* 2019; 367: l5358.

29. Mahmoudi E, Kamdar N, Kim N, et al. Use of electronic medical records in development and validation of risk prediction models of hospital readmission: systematic review. *BMJ* 2020; 369: m958.

30. Covidence - Better systematic review management. *Covidence*, https://www.covidence.org/ (accessed 5 September 2024).

31. Kellermeyer L, Harnke B, Knight S. Covidence and Rayyan. *J Med Libr Assoc* 2018; 106: 580–583.

32. Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLoS Med* 2014; 11: e1001744.

33. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med* 2019; 170: W1–W33.

34. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 2019; 170: 51–58.

35. Bashir B, Schneider D, Naglak MC, et al. Evaluation of prediction strategy and care coordination for COPD readmissions. *Hospital Practice* 2016; 44: 123–128.

Annex

Appendix 1 PubMed search strategy

Search number	Query
13	#3 AND #6 AND #9 AND #12
12	#10 OR #11
11	"Predictive variabl*"[TIAB] OR "Risk factors"[TIAB] OR "Util*"[TIAB] OR "value*"[tiab] OR "tool*"[tiab] OR "index"[tiab] OR "indices"[tiab] OR "assess*"[tiab] OR "risk prediction"[TIAB] OR "risk score"[TIAB] OR "risk calculation"[TIAB] OR "risk assessment"[TIAB]
10	"Risk"[Mesh]
9	#7 OR #8
8	"Data mining*"[TIAB] OR "Forecasting*"[TIAB] OR "Explanatory modelling"[TIAB] OR "Neural Networks*"[TIAB] OR "Support Vector Machine"[TIAB] OR "Naïve Bayesian classifier"[TIAB] OR "Logistic regression"[TIAB] OR "heuristic approach*"[TIAB] OR "Statistical approach*"[TIAB] OR "c-statistic*"[tiab] OR "ROC"[tiab] OR "nomogram"[tiab] OR "indicat*"[tiab] OR "calibration"[tiab] OR "area under the curve"[tiab] OR "area under the receiver operator characteristic curve"[tiab] OR "Predictive Value of Tests"[TIAB]
7	"Regression Analysis"[Mesh] OR "Algorithms"[Mesh] OR "Nomograms"[mesh] OR "Prognosis"[Mesh:NoExp] OR "Support

	Vector Machine"[Mesh] OR "Area Under Curve"[Mesh]
6	#4 OR #5
5	"Rehosp*"[TIAB] OR "Discharge"[TIAB] OR "Unplanned readmission*"[TIAB] OR "Readmission*"[tiab] OR "Re-admission*"[tiab] OR "Readmit*"[tiab] OR "Repeat*"[tiab] OR "Admission*"[tiab]
4	"Patient Readmission"[Mesh] OR "Hospitalization*"[Mesh] OR Hospital*"[Mesh]
3	#1 OR #2
2	"Pulmonary Emphysema*"[TIAB] OR "Chronic Pulmonary disease*"[TIAB] OR "COPD*"[TIAB]
1	"Pulmonary Disease, Chronic Obstructive"[Mesh]

Appendix 2 CHARMS 2014 Relevant items to extract from individual studies in a systematic review of prediction models[32].

Domain	Key items	Reported

Source of data	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)
Participants	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria) Participant description Details of treatments received, if relevant Study dates
Outcome(s) to be predicted	Definition and method for measurement of outcome Was the same outcome definition (and method for measurement) used in all patients? Type of outcome (e.g., single or combined endpoints) Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)? Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)? Time of outcome occurrence or summary of duration of follow-up
Candidate predictors (or index tests)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics) Definition and method for measurement of candidate predictors Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment initiation) Were predictors assessed blinded for outcome, and for each other (if relevant)?

	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)
Sample size	Number of participants and number of outcomes/events Number of outcomes/events in relation to the number of candidate predictors (Events Per Variable)
Missing data	Number of participants with any missing value (include predictors and outcomes) Number of participants with missing data for each predictor Handling of missing data (e.g., complete-case analysis, imputation, or other methods)
Model development	Modelling method (e.g., logistic, survival, neural network, or machine learning techniques) Modelling assumptions satisfied Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome) Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike Information Criterion) Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized
Model	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination

performance	Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used
Model Evaluation	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators) In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)
Results	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals) Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance Comparison of the distribution of predictors (including
Interpretation and discussion	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)

	Comparison with other studies, discussion of generalizability, strengths and limitations.

Appendix 3 PROBAST: Assessment of Risk of Bias and Concerns Regarding Applicability[34]

1. Participants	2. Predictors	3. Outcome	4. Analysis
Signaling questions			
1.1. Were appropriate data sources used, e.g., cohort, RCT, or nested case-control study data?	**2.1.** Were predictors defined and assessed in a similar way for all participants?	**3.1.** Was the outcome determined appropriately?	**4.1.** Were there a reasonable number of participants with the outcome?
1.2. Were all inclusions and exclusions of participants appropriate?	**2.2.** Were predictor assessments made without knowledge of outcome data?	**3.2.** Was a prespecified or standard outcome definition used?	**4.2.** Were continuous and categorical predictors handled appropriately?
	2.3. Are all predictors available at the time the model is intended to be used?	**3.3.** Were predictors excluded from the outcome definition?	**4.3.** Were all enrolled participants included in the analysis?
		3.4. Was the outcome defined and determined in a similar way for all	**4.4.** Were participants with missing data handled appropriately?

		participants? **3.5.** Was the outcome determined without knowledge of predictor information? **3.6.** Was the time interval between predictor assessment and outcome determination appropriate?	**4.5.** Was selection of predictors based on univariable analysis avoided? **4.6.** Were complexities in the data (e.g., censoring, competing risks, sampling of control participants) accounted for appropriately? **4.7.** Were relevant model performance measures evaluated appropriately? **4.8.** Were model overfitting, underfitting, and optimism in model performance accounted for? **4.9.** Do predictors and their

			assigned weights in the final model correspond to the results from the reported multivariable analysis?
ROB Selection of participants	Predictors or their assessment	Outcome or its determination	Analysis
Applicability Included participants or setting does not match the review question	Definition, assessment, or timing of predictors does not match the review question	Its definition, timing, or determination does not match the review question	

RCT = randomized controlled trial; ROB = risk of bias. Signaling questions are answered as yes, probably yes, probably no, no, or no information. ROB and concerns for applicability are rated as low, high, or unclear.

BMJ Open

Predictive Modelling Methods of Hospital Readmission Risks for Patients with Chronic Obstructive Pulmonary Disease (COPD): A Systematic Review Protocol

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-093771.R1
Article Type:	Protocol
Date Submitted by the Author:	13-Mar-2025
Complete List of Authors:	Tonde, Balkissa; Université Laval Faculté des Sciences de l'Administration, Department of Operations and Decision Systems; VITAM Center for Sustainable Health Research Traore, Metogara; Université Laval Faculté des Sciences de l'Administration, Department of Management; VITAM Center for Sustainable Health Research Landa, Paolo; Laval University, Operations and Decision Systems Department; Institut Universitaire de Cardiologie et de Pneumologie de Québec - Université Laval Côté, André; Université Laval, Département de management; Université Laval Faculté des Sciences de l'Administration, Department of Operations and Decision Systems Laberge, Maude; Universite Laval Faculte de medecine, Social and Preventive Medicine; VITAM Center for Sustainable Health Research, Université Laval
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Respiratory medicine, Health informatics
Keywords:	Risk Factors, Machine Learning, Pulmonary Disease, Chronic Obstructive, Prognosis, Hospitalization

Title of the Manuscript:

Predictive Modelling Methods of Hospital Readmission Risks for Patients with Chronic Obstructive Pulmonary Disease (COPD): A Systematic Review Protocol

ABSTRACT

Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a significant chronic respiratory condition characterized by persistent airway obstruction, leading to substantial morbidity and mortality worldwide. Patients with COPD frequently experience hospital readmissions shortly after discharge, mainly due to acute exacerbations. This review aims to identify and synthesize the reported performance metrics and methodological limitations of different predictive modelling methods for hospital readmissions in COPD patients.

Method and analysis

This protocol adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols (PRISMA-P) guidelines. The review will include studies that develop or validate predictive models for hospital readmissions in COPD patients. A comprehensive search will be conducted across PubMed, Embase, Cochrane Library, IEEE Xplore, Web of Science, and Google Scholar using predefined keywords. Eligible studies will include those utilizing any predictive modelling method, focusing on unplanned readmissions within specified timeframes (30, 60, or 90 days). Two independent reviewers will screen titles, abstracts, full texts, selecting studies based on predefined inclusion criteria.

Data extraction will be conducted based on the CHARMS checklist, and the methodological quality and risk of bias will be assessed using the Prediction Model Risk Of Bias Assessment Tool (PROBAST).

The results will be synthesized narratively. A meta-analysis using a random-effects model will be conducted if at least five external validation studies are available for the same prediction model.

Ethics and dissemination

This research is based exclusively on published studies and does not involve the collection of primary data collection from patients. Therefore, ethical approval is not required. Findings will be disseminated through publications in peer-reviewed journals and presentations at national and international conferences.

Prospero registration number

This protocol is registered with PROSPERO (**CRD42024579524**).

Keywords: COPD, hospital readmission, predictive modelling, risk factors

STRENGTHS AND LIMITATIONS OF A STUDY

⇒ This study's strengths lie in its comprehensive approach to systematically reviewing predictive models for hospital readmissions in COPD patients.

⇒ This review will critically assess the reported performance metrics across studies while accounting for methodological variations, including data heterogeneity and model specifications.

⇒ The use of rigorous methodologies, applying of the CHARMS checklist and PROBAST tool, ensures a thorough assessment of the quality and risk of bias in the included studies.

⇒ The anticipated heterogeneity in study designs, patient populations, and predictive models may challenge the synthesis of results and limit the generalizability of the findings.

INTRODUCTION

According to the World Health Organization, Chronic Obstructive Pulmonary Disease (COPD) is a major chronic respiratory condition characterized by persistent airway obstruction, contributing to significant morbidity and mortality worldwide and accounting for nearly 5% of deaths in 2021[1]. In addition to its impact on mortality, COPD imposes substantial costs, further exacerbated by frequent medical consultations, emergency visits, and hospitalizations[2,3,4]. A recent study by Chen et al. (2023)[5] estimates that COPD will cost the global economy USD 4.326 trillion between 2020 and 2050, representing an annual tax of 0.111% on global GDP. This highlights the substantial economic impact of the disease across different countries and regions. Moreover, patients with COPD often experience hospital readmissions shortly after discharge, primarily due to acute exacerbations. These readmissions highlight gaps in care management and coordination during transitions between care levels[6,7]. Frequent and often preventable readmissions significantly increase healthcare costs and negatively affect the quality of care provided[8,9,10].

For instance, in the United States, COPD readmissions are responsible for more than $15 billion annually in direct healthcare costs alone[8], representing up to 75% of the total costs associated with managing COPD exacerbations[11]. Despite progress in understanding readmission factors, high rates persist, emphasizing the need for better prevention and post-discharge monitoring. Improved coordination between hospitals and primary care, combined with personalized management, has proven effective in reducing readmissions and improving patient outcomes[12,13]. Given COPD significant financial burden and high readmission rates, predictive modeling helps identify key risk factors to optimize interventions and care planning[14,15]. Predictive modelling is "the process of applying a statistical model or data mining algorithm to data to predict new observations or future observations"[16]. Current studies[14,17,18] highlight diverse methods for predicting readmissions, reflecting COPD's complexity and its multiple influencing factors.

Conventional approaches, such as LACE, PEARL, Elixhauser, and HOSPITAL indices [18,19,20], have been widely adopted in clinical practice. The LACE index predicts 30-day readmissions or death based on length of stay, acuity of admission, comorbidity, and emergency visits in the last six months[19]. The HOSPITAL score, identifies patients at risk of avoidable readmissions, focusing on hospital procedures and previous admissions[20]. The PEARL score predicts 90-day readmission or death in COPD based on five criteria: previous admissions, eMRCD score, age, and right- and left-sided heart failure[18]. The Elixhauser comorbidity index assesses assesses 29 comorbidities to predict hospital mortality and readmissions[21]. While these tools are useful, they often fail to fully address the complexity of COPD and its social determinants, which can limit their effectiveness in specific contexts or populations[18].

Machine learning models are promising alternatives that outperform traditional methods by leveraging electronic health record for dynamic and precise readmission risk analysis. These technologies enhance risk factor modeling by integrating multiple dimensions and processing large health datasets[14,15,17]. Several studies[14,15,17] highlighted these models precision and ability to capture complex risk factor interactions in hospital readmissions better than traditional methods. Effectiveness varies by context, emphasizing the need for systematic evaluation of each methods strengths and weaknesses.

This review aims to explore the current state of predictive modeling for COPD hospital readmission risks, focusing on evaluating the performance of traditional models and machine learning approaches. Unlike previous reviews focusing on specific methods or periods, this review systematically incorporates heuristic and statistical approaches without time restrictions. By highlighting recent advances, challenges, and limitations, it evaluates the effectiveness of predictive models. Additionally, this review discusses the transition to mixed prediction methods, combining traditional and machine learning approaches for improved accuracy and clinical use. By synthesizing existing evidence, it provides insights into predictive model effectiveness and supports the shift toward advanced hybrid models. The findings will inform future research and aid in identifying individuals at highest risk of readmission which could be the target of interventions to reduce hospital readmissions. Identifying these patients should enhance health outcomes and alleviate the burden on healthcare systems.

This systematic review aims to address the following primary question: What is the effectiveness of the different predictive modelling methods for hospital readmission risks in patients with COPD?

OBJECTIVES

The objectives of this review are to :

1.Identify the different predictive modelling approaches used to predict readmission risks for patients with COPD.

2. Synthesize and analyze the reported performance metrics and methodological limitations of these predictive models.

3. Identify the strengths and weaknesses of using each method.

METHODS AND ANALYSIS

This review will follow a rigorous methodology outlined in the Cochrane Handbook for Systematic Reviews of Interventions[22,23], with a protocol registered in PROSPERO (number **CRD42024579524**).

Eligibility criteria

This protocol follows PRISMA-P guidelines[24,25] and the PICOS approach[26] to define eligibility criteria, including the population, intervention/exposure, comparator, outcomes, and study design. These criteria is based on a brief literature review, including systematic evaluations of predictive models for hospital readmissions[27,28,29]. This approach ensures that the most relevant elements are included to effectively address the research question. Inclusion and exclusion criteria are described in table 1

Table 1 Summary of inclusion and exclusion criteria

Criteria	Inclusion	Exclusion
Population	Adult patients discharged after a hospital admission where the final diagnosis was COPD.	Studies on other populations. Studies not specifically concerning COPD readmissions. Studies focusing on elective readmissions (e.g. readmission after specific surgery).
Intervention	Any studies reporting the development/ validation of a readmission prediction model including the use of heuristics approach or statistical methods (machine learning, or other analytical techniques).	Studies on the risks of COPD readmissions without a focus on predictive modelling. Studies that do not provide a clear and detailed description of the variables used in the predictive models.
Comparator	No restriction on comparator type	NA
Outcomes	Unplanned readmission within 30, 60 or 90 days following patient discharge. Ability of models to predict hospital readmission risk (C-statistic, accuracy of predictive models, sensitivity, specificity, calibration, and validation of models). Associated risk factors include clinical variables, comorbidities, health behaviors,	Outcomes not related to unplanned readmission or predictive modelling.

	etc.	
Study design	Quantitative empirical studies, including prospective and retrospective cohort studies, randomized controlled trials, case-controlled studies, model validation studies, and studies utilizing panel data.	Retracted publications, qualitative or mixed-methods studies, and non-empirical material such as knowledge syntheses, letters, perspectives, editorial, theses, conference abstracts, study protocols, books, book chapters, and scoping reviews.
Language	English or French	Other languages
Time	No restrictions	NA
Context	Healthcare studies related to hospital readmission for COPD.	Not health-related or not related to hospital readmission for COPD

Information sources

To identify studies on predictive models for COPD readmissions, an exhaustive search will be conducted on the following electronic databases: PubMed, Embase, Cochrane Library, IEEE Xplore, and Web of Science. The search period will cover from the inception of each database to the date of the search, in both English and French. In addition, a search for grey literature will be conducted using Google Scholar to include articles not published in traditional databases.

Search strategy

Our search strategy employs both MeSH terms and keywords to ensure comprehensive coverage of relevant studies. Key terms such as "patient readmission" and "rehospitalization" are specifically targeted to capture studies focused on hospital readmissions. Additionally, the strategy includes terms associated with predictive modeling, such as "Regression Analysis," "Algorithms," "Prognosis," "Support Vector Machine," and "machine learning." For COPD, we will use specific terms including "chronic obstructive pulmonary disease," "COPD," "emphysema," and "chronic bronchitis." We will also incorporate MeSH terms related to "risk" along with keywords such as "risk factors" and "risk assessment." These search criteria will be systematically combined using Boolean operators (AND, OR) to refine the final set of references in each database. The references cited in the included studies will also be reviewed to identify additional relevant articles. A proposed search strategy for PubMed is detailed in the Appendix 1.

Data collection and management

Selection of studies

Citations will be imported into the online systematic review platform Covidence, where duplicates will be automatically removed prior to the detailed study analysis[30,31]. Two independent reviewers will conduct every step and meet to discuss discrepancies. If consensus is not reached, then a third investigator will adjudicate. The first step will consist of screening titles and abstracts to assess relevance based on predefined inclusion and exclusion criteria. We will then review the full article of each retained reference.

At this step, we will note the reason for each excluded article.

Data collection and extraction process

A double data extraction will be performed for all included articles. Two reviewers will independently extract data from each article using a standardized data extraction form (Appendix 2) based on the CHARMS checklist[32]. The data extraction form will be piloted on a 10% sample of articles and adjusted if necessary. Any disagreements in data extraction will be discussed between the two reviewers and, if unresolved, will be adjudicated by a third party . The extracted data will include the following key elements:

- **Source and study design:** study title, authors, publication year, and country, along with the data source type (e.g., cohort, case-control).
- **Participant characteristics:** participant descriptions, including sample size and demographic details such as age, gender, and recruitment methods.
- **Modelling methods:** type of predictive modelling approach used in the study, as well as the analyses for the total population.
- **Handling of Missing Data:** Presence of missing data (on outcomes and/or variables) and method of handling missing data
- **Outcome measures:** Primary and secondary outcomes related to COPD readmission, with a focus on how these outcomes are measured and reported.
- **Predictive variables and model inputs:** number and names of predictive variables used (demographics, clinical), and measurement method.
- **Model development and validation:** Modeling methods used, selection of predictors, methods employed for model validation, (internal/external), performance metrics (C-index, sensitivity, specificity, calibration).

- **Results and interpretation:** final model's results, potential biases, limitations, and assumptions.

Risk of bias in individual studies

The risk of bias will be assessed using the Prediction model Risk Of Bias Assessment Tool (PROBAST)[33,34], which is specifically designed for systematic reviews of diagnostic prediction models (Appendix 3). The assessment will focus on several key domains, including participants, predictors, outcomes, analyses, and overall methodological rigour. Each domain will be evaluated and classified as "high risk of bias," "unclear risk of bias," or "low risk of bias." Any discrepancy or conflict that arises during the assessment will be resolved through consensus discussions, and if necessary, with the involvement of a third reviewer.

Data synthesis

The synthesis will focus on both qualitative and quantitative aspects, acknowledging the potential heterogeneity across studies due to differences in study populations, model types, and implementation contexts.

Narrative synthesis

We will begin with a narrative synthesis to summarize and interpret the findings from the included studies. This approach allows us to describe the characteristics and performance of the predictive models in a detailed and systematic manner. We will highlight the key study features, including population characteristics, predictive model types, and their development and validation contexts.

The narrative synthesis will also address the following critical aspects:

- Model discrimination: We will assess how well each model distinguishes between patients who are readmitted and those who are not. This will include evaluating metrics such as the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) and other relevant performance indicators.
- Cohort type: We will compare the types of COPD patient cohorts used across studies, considering factors such as the inclusion criteria, severity of the disease, and the presence of comorbidities.
- Practical implementation: We will examine the practical aspects of implementing these predictive models in clinical settings, including the ease of use, data availability, and the resources required for the implementation.
- Predictor variables : We will categorize and compare the types of variables included in the predictive models, as the choice of variables can significantly impact model accuracy and generalizability.
- Model performance indicators: Beyond discrimination, we will explore other performance indicators such as calibration, sensitivity, specificity, and predictive values.

Meta-analysis and heterogeneity management

Due to variations in study design, patient populations, and model types, heterogeneity is expected in this review. Therefore, a random-effects model with restricted maximum likelihood estimation will be used to account for inter-study variability[35]. The Hartung-Knapp-Sidik-Jonkman (HKSJ) method will be used to calculate the 95% confidence intervals[36].

The primary outcome will be the C-statistic, which will be transformed using the logit function, and standard errors will be calculated accordingly. A forest plot will summarize predictive performance, displaying AUROC, sensitivity, and specificity with corresponding 95% confidence intervals. Statistical heterogeneity assessed using the I^2 statistic, will be further examined through a leave-one-out sensitivity analysis if $I^2 > 50\%$[37]. If more than 10 studies are available, a meta-regression will explore the impact of study characteristics on model performance[35]. Subgroup analyses will be conducted based on region, participant sex, study design, model performance and validation method[38]. Publication bias will be assessed using funnel plots, and Egger's regression test will analyze asymmetry.

ETHICS AND DISSEMINATION

This research project is based on the analysis of published studies and does not involve direct patient data collection; hence no ethics approval is required. However, the study will adhere to ethical principles of scientific rigour, transparency, and respect for data sources. The findings will be disseminated through publication in scientific journals, presentations at national and international conferences, and specialized forums in health sciences and research methodology. This approach will maximize the impact of the findings and support evidence-informed public health policies.

Patient and public involvement

No patients or the public were involved in the study design.

DISCUSSION

The evolution of predictive models for COPD hospital readmissions has transitioned from traditional risk scores to more sophisticated machine learning approaches. Established indices such as LACE, PEARL, Elixhauser, and HOSPITAL remain widely used in clinical practice. However, these conventional models often struggle with generalizability and adaptability to individual patient characteristics, limiting their ability to capture the complexity of readmissions[14,39,40,41].

In contrast, machine learning based models leverage large datasets to identify intricate interactions between risk factors, improving prediction accuracy[42]. Despite these advantages, their clinical adoption faces significant challenges, including high data requirements, computational demands, and the need for specialized expertise[43]. Furthermore, generalizability remains a major concern, as many models are developed on specific cohorts, restricting their applicability across diverse healthcare settings[44,45]. Multicenter validation and pilot studies on varied patient populations are essential to enhance robustness and ensure clinical reliability[46]. Another critical barrier to adoption is integration into Electronic Health Records (EHRs). Many predictive models rely on structured data, while valuable clinical insights remain embedded in unstructured physician notes and free-text reports. Natural Language Processing (NLP) techniques could address this issue by automating data extraction and enhancing real-time model usability in clinical decision-making[47,48]. Additionally, model interpretability remains a crucial factor influencing clinical acceptance. While deep learning models demonstrate high predictive performance, their "black-box" nature limits transparency and trust among healthcare professionals[17,49].

Hybrid models, which combine traditional risk scores with AI-driven predictions, offer a promising solution by balancing accuracy and explainability[50]. Another key consideration is bias and fairness in predictive modeling. Socioeconomic and behavioral factors are often underrepresented in datasets, leading to disparities in prediction quality across different patient populations[51,52]. Incorporating these factors into machine learning models could improve representativeness and mitigate biases in care delivery. The next phase of predictive modeling for COPD readmissions must focus on dynamic integration into clinical workflows. Evaluation should extend beyond standard metrics like AUC to include calibration analysis and net benefit assessment, ensuring real-world clinical relevance[14,18,53].

Future research should explore the feasibility of real-world implementation by conducting prospective validation studies within routine care settings[54]. Additionally, visualization techniques, such as comparative charts and performance graphs, can help illustrate differences between models and facilitate clinical decision-making[55].

This review will synthesize the current state of predictive modeling for COPD readmissions, highlighting the strengths and limitations of different approaches. The findings could inform future research and support the transition towards integrating advanced prediction methods in clinical practice.

Amendments

Any protocol amendments will be documented and made publicly accessible on the PROSPERO registration. Each amendment will include the date, a detailed description, and the rationale for the changes made.

Ethics approval and patient consent for publication

Not applicable.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding statement

Balkissa Tonde is supported by the VITAM strategic research development fund and the postgraduate scholarship fund from the Faculty of Business Administration at Université Laval.

Contribution Statement:

Balkissa Tonde designed the protocol, developed the search strategy, and drafted and revised the manuscript. Balkissa Tonde and Metegora Mohamed Traore planned the data extraction process, the methodological assessment of the studies, and the statistical analysis. Maude Laberge provided guidance and critical insights at each step and revised the manuscript. Paolo Landa and André Côté also provided valuable feedback and critical revisions to the manuscript. All authors reviewed, approved, and contributed to the final version of the manuscript. The guarantor of the study is Balkissa Tonde.

REFERENCES

1. WHO. Chronic obstructive pulmonary disease (COPD). 2024. Accessed March 3, 2025. https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd)

2. Momtazmanesh S, Moghaddam SS, Ghamari SH, et al. Global burden of chronic respiratory diseases and risk factors, 1990–2019: an update from the Global Burden of Disease Study 2019. *eClinicalMedicine*. 2023;59. doi:10.1016/j.eclinm.2023.101936

3. Alshabanat A, Otterstatter MC, Sin DD, et al. Impact of a COPD comprehensive case management program on hospital length of stay and readmission rates. *Int J Chron Obstruct Pulmon Dis*. 2017;12:961-971. doi:10.2147/COPD.S124385

4. Alqahtani JS, Aldabayan YS, Aldhahir AM, Al Rajeh AM, Mandal S, Hurst JR. Predictors of 30- and 90-Day COPD Exacerbation Readmission: A Prospective Cohort Study. *Int J Chron Obstruct Pulmon Dis*. 2021;16:2769-2781. doi:10.2147/COPD.S328030

5. Chen S, Cao Z, Prettner K, et al. Estimates and Projections of the Global Economic

Cost of 29 Cancers in 204 Countries and Territories From 2020 to 2050. *JAMA Oncol*. 2023;9(4):465. doi:10.1001/jamaoncol.2022.7826

6. Fernández-García S, Represas-Represas C, Ruano-Raviña A, et al. Social and clinical predictors of short- and long-term readmission after a severe exacerbation of copd. *PLoS One*. 2020;15(2):e0229257. doi:10.1371/journal.pone.0229257

7. Lindenauer PK, Dharmarajan K, Qin L, Lin Z, Gershon AS, Krumholz HM. Risk Trajectories of Readmission and Death in the First Year after Hospitalization for Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med*. 2018;197(8):1009-1017. doi:10.1164/rccm.201709-1852OC

8. Buhr RG, Jackson NJ, Kominski GF, Dubinett SM, Mangione CM, Ong MK. Readmission Rates for Chronic Obstructive Pulmonary Disease Under the Hospital Readmissions Reduction Program: an Interrupted Time Series Analysis. *J Gen Intern Med*. 2020;35(12):3581-3590. doi:10.1007/s11606-020-05958-0

9. Shah AA, Devana SK, Lee C, et al. Machine learning driven identification of novel patient factors for prediction of major complications after posterior cervical spinal fusion. *Eur Spine J*. 2022;31(8):1952-1959. doi:10.1007/s00586-021-06961-7

10. Njoku CM, Alqahtani JS, Wimmer BC, et al. Risk factors and associated outcomes of hospital readmission in COPD: A systematic review. *Respiratory Medicine*. 2020;173:105988. doi:10.1016/j.rmed.2020.105988

11. Jacobs DM, Noyes K, Zhao J, et al. Early Hospital Readmissions after an Acute Exacerbation of Chronic Obstructive Pulmonary Disease in the Nationwide Readmissions Database. *Ann Am Thorac Soc*. 2018;15(7):837-845. doi:10.1513/AnnalsATS.201712-913OC

12. Adamson SL, Burns J, Camp PG, Sin DD, van Eeden SF. Impact of individualized care on readmissions after a hospitalization for acute exacerbation of COPD. *Int J Chron Obstruct Pulmon Dis*. 2016;11:61-71. doi:10.2147/COPD.S93322

13. Bricard D, Or Z. Impact of early primary care follow-up after discharge on hospital readmissions. *Eur J Health Econ*. 2019;20(4):611-623. doi:10.1007/s10198-018-1022-y

14. Min X, Yu B, Wang F. Predictive Modeling of the Hospital Readmission Risk from Patients' Claims Data Using Machine Learning: A Case Study on COPD. *Sci Rep*. 2019;9(1):2362. doi:10.1038/s41598-019-39071-y

15. Li M, Cheng K, Ku K, Li J, Hu H, Ung COL. Modelling 30-day hospital readmission after discharge for COPD patients based on electronic health records. *npj Prim Care Respir Med*. 2023;33(1):1-8. doi:10.1038/s41533-023-00339-6

16. Shmueli G. To Explain or to Predict? *Statist Sci*. 2010;25(3). doi:10.1214/10-STS330

17. Bonomo M, Hermsen MG, Kaskovich S, et al. Using Machine Learning to Predict Likelihood and Cause of Readmission After Hospitalization for Chronic Obstructive Pulmonary Disease Exacerbation. *Int J Chron Obstruct Pulmon Dis*. 2022;17:2701-2709. doi:10.2147/COPD.S379700

18. Echevarria C, Steer J, Heslop-Marshall K, et al. The PEARL score predicts 90-day readmission or death after hospitalisation for acute exacerbation of COPD. *Thorax*. 2017;72(8):686-693. doi:10.1136/thoraxjnl-2016-209298

19. van Walraven C, Dhalla IA, Bell C, et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. *CMAJ*. 2010;182(6):551-557. doi:10.1503/cmaj.091117

20. Donzé JD, Williams MV, Robinson EJ, et al. International Validity of the HOSPITAL Score to Predict 30-Day Potentially Avoidable Hospital Readmissions. *JAMA Intern Med*. 2016;176(4):496-502. doi:10.1001/jamainternmed.2015.8462

21. Moore BJ, White S, Washington R, Coenen N, Elixhauser A. Identifying Increased Risk of Readmission and In-hospital Mortality Using Hospital Administrative Data: The AHRQ Elixhauser Comorbidity Index. *Medical Care*. 2017;55(7):698-705. doi:10.1097/MLR.0000000000000735

22. Higgins JPT, Cochrane Collaboration, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Second edition. Wiley-Blackwell; 2019.

23. Higgins JP, Green S. Guide to the Contents of a Cochrane Protocol and Review. In: *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons, Ltd; 2008:51-79. doi:10.1002/9780470712184.ch4

24. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71

25. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*. 2015;349:g7647. doi:10.1136/bmj.g7647

26. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Services Research*. 2014;14(1):579. doi:10.1186/s12913-014-0579-0

27. Kansagara D, Englander H, Salanitro A, et al. Risk Prediction Models for Hospital

Readmission: A Systematic Review. *JAMA*. 2011;306(15):1688-1698. doi:10.1001/jama.2011.1515

28. Bellou V, Belbasis L, Konstantinidis AK, Tzoulaki I, Evangelou E. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. *BMJ*. 2019;367:l5358. doi:10.1136/bmj.l5358

29. Mahmoudi E, Kamdar N, Kim N, Gonzales G, Singh K, Waljee AK. Use of electronic medical records in development and validation of risk prediction models of hospital readmission: systematic review. *BMJ*. 2020;369:m958. doi:10.1136/bmj.m958

30. Covidence - Better systematic review management. Covidence. Accessed September 5, 2024. https://www.covidence.org/

31. Kellermeyer L, Harnke B, Knight S. Covidence and Rayyan. *J Med Libr Assoc*. 2018;106(4):580-583. doi:10.5195/jmla.2018.513

32. Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. *PLoS Med*. 2014;11(10):e1001744. doi:10.1371/journal.pmed.1001744

33. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med*. 2019;170(1):W1-W33. doi:10.7326/M18-1377

34. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med*. 2019;170(1):51-58. doi:10.7326/M18-1376

35. Debray TPA, Damen JAAG, Snell KIE, et al. A guide to systematic review and meta-analysis of prediction model performance. Published online January 5, 2017.

doi:10.1136/bmj.i6460

36. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol*. 2014;14:25. doi:10.1186/1471-2288-14-25

37. Thompson SG. Why and how Sources of Heterogeneity Should be Investigated. In: *Systematic Reviews in Health Care*. John Wiley & Sons, Ltd; 2001:157-175. doi:10.1002/9780470693926.ch9

38. Deeks JJ, Higgins JP, Altman DG, Group on behalf of the CSM. Analysing data and undertaking meta-analyses. In: *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons, Ltd; 2019:241-284. doi:10.1002/9781119536604.ch10

39. Shah T, Press VG, Huisingh-Scheetz M, White SR. COPD Readmissions: Addressing COPD in the Era of Value-based Health Care. *Chest*. 2016;150(4):916-926. doi:10.1016/j.chest.2016.05.002

40. Wang H, Robinson RD, Johnson C, et al. Using the LACE index to predict hospital readmissions in congestive heart failure patients. *BMC Cardiovasc Disord*. 2014;14:97. doi:10.1186/1471-2261-14-97

41. Sharma V, Kulkarni V, McAlister F, et al. Predicting 30-Day Readmissions in Patients With Heart Failure Using Administrative Data: A Machine Learning Approach. *J Card Fail*. 2022;28(5):710-722. doi:10.1016/j.cardfail.2021.12.004

42. Jamei M, Nisnevich A, Wetchler E, Sudat S, Liu E. Predicting all-cause risk of 30-day hospital readmission using artificial neural networks. *PLoS One*. 2017;12(7):e0181173. doi:10.1371/journal.pone.0181173

43. Subramanian M, Wojtusciszyn A, Favre L, et al. Precision medicine in the era of artificial intelligence: implications in chronic disease management. *J Transl Med*. 2020;18(1):472. doi:10.1186/s12967-020-02658-5

44. Singh H, Mhasawade V, Chunara R. Generalizability challenges of mortality risk prediction models: A retrospective analysis on a multi-center database. *PLOS Digital Health*. 2022;1(4):e0000023. doi:10.1371/journal.pdig.0000023

45. Schwab P, DuMont Schütte A, Dietz B, Bauer S. Clinical Predictive Models for COVID-19: Systematic Study. *J Med Internet Res*. 2020;22(10):e21439. doi:10.2196/21439

46. Mao Q, Jay M, Hoffman JL, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ Open*. 2018;8(1):e017833. doi:10.1136/bmjopen-2017-017833

47. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25(1):24-29. doi:10.1038/s41591-018-0316-z

48. Hossain E, Rana R, Higgins N, et al. Natural Language Processing in Electronic Health Records in relation to healthcare decision-making: A systematic review. *Comput Biol Med*. 2023;155:106649. doi:10.1016/j.compbiomed.2023.106649

49. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Published online August 10, 2015:1721-1730. doi:10.1145/2783258.2788613

50. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56-67.

doi:10.1038/s42256-019-0138-9

51. Chen IY, Joshi S, Ghassemi M. Treating health disparities with artificial intelligence. *Nat Med*. 2020;26(1):16-17. doi:10.1038/s41591-019-0649-2

52. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-453. doi:10.1126/science.aax2342

53. Van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. *BMC Medicine*. 2019;17(1):230. doi:10.1186/s12916-019-1466-7

54. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Published online 2014. https://dx.doi.org/10.1093/eurheartj/ehu207

55. Molnar C. *Interpretable Machine Learning : A Guide For Making Black Box Models Explainable*. Lulu.com; 2020.

Annex

Appendix 1 PubMed search strategy

Search number	Query
13	#3 AND #6 AND #9 AND #12
12	#10 OR #11
11	"Predictive variabl*"[TIAB] OR "Risk factors"[TIAB] OR "Util*"[TIAB] OR "value*"[tiab] OR "tool*"[tiab] OR "index"[tiab] OR "indices"[tiab] OR "assess*"[tiab] OR "risk prediction"[TIAB] OR "risk score"[TIAB] OR "risk calculation"[TIAB] OR "risk assessment"[TIAB]
10	"Risk"[Mesh]
9	#7 OR #8
8	"Data mining*"[TIAB] OR "Forecasting*"[TIAB] OR "Explanatory modelling"[TIAB] OR "Neural Networks*"[TIAB] OR "Support Vector Machine"[TIAB] OR "Naïve Bayesian classifier"[TIAB] OR "Logistic regression"[TIAB] OR "heuristic approach*"[TIAB] OR "Statistical approach*"[TIAB] OR "c-statistic*"[tiab] OR "ROC"[tiab] OR "nomogram"[tiab] OR "indicat*"[tiab] OR "calibration"[tiab] OR "area under the curve"[tiab] OR "area under the receiver operator characteristic curve"[tiab] OR "Predictive Value of Tests"[TIAB]

7	"Regression Analysis"[Mesh] OR "Algorithms"[Mesh] OR "Nomograms"[mesh] OR "Prognosis"[Mesh:NoExp] OR "Support Vector Machine"[Mesh] OR "Area Under Curve"[Mesh]
6	#4 OR #5
5	"Rehosp*"[TIAB] OR "Discharge"[TIAB] OR "Unplanned readmission*"[TIAB] OR "Readmission*"[tiab] OR "Re-admission*"[tiab] OR "Readmit*"[tiab] OR "Repeat*"[tiab] OR "Admission*"[tiab]
4	"Patient Readmission"[Mesh] OR "Hospitalization*"[Mesh] OR Hospital*"[Mesh]
3	#1 OR #2
2	"Pulmonary Emphysema*"[TIAB] OR "Chronic Pulmonary disease*"[TIAB] OR "COPD*"[TIAB]
1	"Pulmonary Disease, Chronic Obstructive"[Mesh]

Appendix 2 CHARMS 2014 Relevant items to extract from individual studies in a systematic review of prediction models[32].

Domain	Key items	Reported
Source of data	Source of data (e.g., cohort, case-control, randomized trial participants, or registry data)	
Participants	Participant eligibility and recruitment method (e.g., consecutive participants, location, number of centers, setting, inclusion and exclusion criteria) Participant description Details of treatments received, if relevant Study dates	
Outcome(s) to be predicted	Definition and method for measurement of outcome Was the same outcome definition (and method for measurement) used in all patients? Type of outcome (e.g., single or combined endpoints) Was the outcome assessed without knowledge of the candidate predictors (i.e., blinded)? Were candidate predictors part of the outcome (e.g., in panel or consensus diagnosis)? Time of outcome occurrence or summary of duration of follow-up	
Candidate predictors (or index tests)	Number and type of predictors (e.g., demographics, patient history, physical examination, additional testing, disease characteristics) Definition and method for measurement of candidate predictors	

	Timing of predictor measurement (e.g., at patient presentation, at diagnosis, at treatment initiation)
	Were predictors assessed blinded for outcome, and for each other (if relevant)?
	Handling of predictors in the modelling (e.g., continuous, linear, non-linear transformations or categorised)
Sample size	Number of participants and number of outcomes/events
	Number of outcomes/events in relation to the number of candidate predictors (Events Per Variable)
Missing data	Number of participants with any missing value (include predictors and outcomes)
	Number of participants with missing data for each predictor
	Handling of missing data (e.g., complete-case analysis, imputation, or other methods)
Model development	Modelling method (e.g., logistic, survival, neural network, or machine learning techniques)
	Modelling assumptions satisfied
	Method for selection of predictors for inclusion in multivariable modelling (e.g., all candidate predictors, pre-selection based on unadjusted association with the outcome)
	Method for selection of predictors during multivariable modelling (e.g., full model approach, backward or forward selection) and criteria used (e.g., p-value, Akaike Information Criterion)
	Shrinkage of predictor weights or regression coefficients (e.g., no shrinkage, uniform shrinkage, penalized

Model performance	Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination Classification measures (e.g., sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used
Model Evaluation	Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators) In case of poor validation, whether model was adjusted or updated (e.g., intercept recalibrated, predictor effects adjusted, or new predictors added)
Results	Final and other multivariable models (e.g., basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals) Any alternative presentation of the final prediction models, e.g., sum score, nomogram, score chart, predictions for specific risk subgroups with performance Comparison of the distribution of predictors (including
Interpretation and discussion	Interpretation of presented models (confirmatory, i.e., model useful for practice versus exploratory, i.e., more research needed)

	Comparison with other studies, discussion of generalizability, strengths and limitations.

Appendix 3 PROBAST: Assessment of Risk of Bias and Concerns Regarding Applicability[34]

1. Participants	2. Predictors	3. Outcome	4. Analysis
Signaling questions			
1.1. Were appropriate data sources used, e.g., cohort, RCT, or nested case-control study data?	**2.1.** Were predictors defined and assessed in a similar way for all participants?	**3.1.** Was the outcome determined appropriately?	**4.1.** Were there a reasonable number of participants with the outcome?
1.2. Were all inclusions and exclusions of participants appropriate?	**2.2.** Were predictor assessments made without knowledge of outcome data?	**3.2.** Was a prespecified or standard outcome definition used?	**4.2.** Were continuous and categorical predictors handled appropriately?
	2.3. Are all predictors available at the time the model is intended to be used?	**3.3.** Were predictors excluded from the outcome definition?	**4.3.** Were all enrolled participants included in the analysis?
		3.4. Was the outcome defined and determined in a	**4.4.** Were participants with missing data handled appropriately?

		similar way for all participants? **3.5.** Was the outcome determined without knowledge of predictor information? **3.6.** Was the time interval between predictor assessment and outcome determination appropriate?	**4.5.** Was selection of predictors based on univariable analysis avoided? **4.6.** Were complexities in the data (e.g., censoring, competing risks, sampling of control participants) accounted for appropriately? **4.7.** Were relevant model performance measures evaluated appropriately? **4.8.** Were model overfitting, underfitting, and optimism in model performance accounted for?

			4.9. Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?
ROB Selection of participants	Predictors or their assessment	Outcome or its determination	Analysis
Applicability Included participants or setting does not match the review question	Definition, assessment, or timing of predictors does not match the review question	Its definition, timing, or determination does not match the review question	

RCT = randomized controlled trial; ROB = risk of bias. Signaling questions are answered as yes, probably yes, probably no, no, or no information. ROB and concerns for applicability are rated as low, high, or unclear.