

REVIEW ARTICLE

Bridging Clinical Prediction and Workplace Exposure: A Data-Driven Risk Stratification and Prediction Model for Occupational Respiratory Diseases.

Muhamad Syazni Asraff, Siti Munira Yasin*, Mohamad Rodi Isa, Nurhuda Ismail, Ely Zarina Zainudin, Mohamad Irfan Mohamad Sallehuddin

Department of Public Health, Faculty of Medicine, UiTM Sungai Buloh, Selangor

Corresponding Author

Siti Munira Yasin 

Department of Public Health, Faculty of Medicine, UiTM Sungai Buloh, Selangor, Malaysia.

Email: smunira@uitm.edu.my

DOI: <https://doi.org/10.70672/n73azh75>

Received: 18/06/2026. Revised: 03/09/2026. Accepted: 22/09/2026. Published online: 01/11/2026.

Abstract

Occupational respiratory disease (ORD) surveillance remains largely reactive, with diagnosis often occurring only after clinically significant symptoms develop. Existing risk stratification tools are primarily clinically derived and rarely incorporate occupational exposure variables such as exposure intensity, duration, and workplace hazards, limiting early identification of at-risk workers. This review synthesizes current evidence on predictive modelling and risk stratification tools for respiratory diseases, evaluates their applicability to occupational health, and proposes a data-driven framework for proactive risk stratification. A structured literature search was conducted in PubMed, Scopus, and Web of Science for studies published between 2021 and 2025. Medical Subject Headings (MeSH) and free-text terms related to occupational respiratory diseases, predictive modelling, prognostic tools, and occupational exposure assessment were combined using Boolean operators. Extracted data included study characteristics, modelling approaches, predictors, and model performance metrics, which were narratively synthesized using thematic mapping. Of 618 records identified, six representative evidence approaches were retained for detailed comparative synthesis. Clinical models such as BODE and DOSE provided useful prognostic information but lacked occupational exposure parameters. Exposure-based approaches showed associations between cumulative exposure and respiratory outcomes, while a meta-analysis of machine-learning models for acute respiratory distress syndrome reported a pooled AUC of 0.7407. Key predictors included age, smoking status, lung function, and exposure-related factors. Current reactive and clinically focused approaches limit effective ORD prevention. Integrating occupational exposure variables into predictive, data-driven risk stratification frameworks is essential for early identification of high-risk workers. The proposed ORDIRS framework represents a step toward proactive occupational health surveillance in industrial settings.

Keywords: *Data-driven framework, occupational exposure, occupational health surveillance, occupational respiratory diseases, predictive modelling, risk stratification.*



This work is licensed under a [Creative Commons Attribution-Non-Commercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Introduction

Occupational respiratory diseases (ORDs) encompass a spectrum of preventable lung disorders resulting from prolonged inhalation of hazardous agents in the workplace, including dusts, fumes, vapours, and gases [1]. Common conditions such as occupational asthma, pneumoconiosis, silicosis, and work-related chronic obstructive pulmonary disease (COPD) continue to represent a major contributor to morbidity and premature mortality worldwide. The World Health Organization (WHO) estimates that more than 500 million workers are exposed to respiratory hazards annually, accounting for approximately 17% of all work-related diseases and over 400,000 deaths each year. Despite progress in industrial hygiene and regulatory control, the burden of ORDs remains substantial, particularly in low- and middle-income countries where surveillance and exposure monitoring are often inconsistent [2,3].

Regional evidence further underscores this burden. In a hospital-based study in Egypt, 1.49% of occupationally exposed railway workers were diagnosed with interstitial lung disease (ILD), compared to 1.1% in non-exposed individuals, highlighting the measurable impact of workplace exposures [4]. Similarly, in Thailand, 16.7% of asthma patients were identified as having work-related asthma, particularly among occupations involving exposure to dust, smoke, and industrial fumes [5]. In Malaysia, rapid industrialization and expansion of high-risk sectors including quarrying, construction, manufacturing, and agriculture have increased exposure to respirable crystalline silica, metal fumes, asbestos fibres, and organic dusts, all of which are established determinants of chronic respiratory impairment [6,7].

Despite these risks, a key limitation in current occupational health practice is the predominantly reactive nature of disease surveillance systems. In Malaysia, although legislation under the Department of Occupational Safety and Health (DOSH) mandates periodic medical surveillance, detection of occupational respiratory disease largely de-

pends on the identification of symptomatic individuals. This approach often results in delayed diagnosis, under-reporting, and intervention at stages when lung damage is already irreversible [8]. Consequently, opportunities for early risk identification and prevention remain limited.

This reactive paradigm is further compounded by the lack of predictive models tailored to occupational settings. Existing risk stratification tools in respiratory medicine, such as the BODE and DOSE indices for COPD, have demonstrated strong prognostic performance in clinical populations [9]. However, these models are primarily designed to assess disease severity and progression rather than to predict disease onset. Importantly, they do not incorporate critical occupational determinants, including exposure type, intensity, duration, and workplace environmental conditions, thereby limiting their applicability in industrial contexts.

While exposure-response models in occupational epidemiology have provided valuable insights into the relationship between hazardous exposures and respiratory outcomes, these approaches are typically applied at the population level and are rarely translated into practical tools for individual risk stratification [3]. At the same time, advances in predictive modelling and data-driven approaches, including machine learning, have demonstrated the capacity to integrate multidimensional variables and improve predictive accuracy [10,11]. However, their application in occupational health remains limited, particularly in settings where data integration and standardization are still evolving.

Collectively, these limitations highlight a critical disconnect between clinically driven prediction models and occupational health needs, as well as a broader failure to transition from reactive disease detection to proactive risk identification. Addressing this gap requires the development of integrated, data-driven predictive frameworks capable of combining clinical, occupational, and exposure-related determinants to enable early identification of high-risk individuals.

This review therefore aims to synthesize current evidence on predictive modelling and risk stratification tools relevant to respiratory diseases, critically evaluate their applicability to occupational health, and outline a conceptual foundation for a data-driven Occupational Respiratory Disease Risk Stratification (ORDiRS) framework.

Materials and methods

This narrative review was supported by a structured literature search conducted across three major scientific databases: PubMed, Scopus, and Web of Science, covering publications from January 2021 to September 2025.

A comprehensive literature search was performed across three major scientific databases: PubMed, Scopus, and Web of Science, covering publications from January 2021 to September 2025.

The primary search terms included combinations of keywords such as “*respiratory disease*”, “*occupational respiratory disease*”, “*predictive model*”, “*risk stratification*”, “*prognostic scoring tool*”, “*exposure assessment*”, and “*occupational health surveillance*.” Search terms were combined using Boolean operators (AND/OR) and adapted to the indexing and search functions of each database. Reference lists of key papers were also manually screened to identify additional relevant publications, including seminal studies describing established prediction tools that predated the primary search period. Reference lists of key papers were also manually screened to identify additional relevant publications not indexed in the main databases.

Inclusion criteria were: (1) peer-reviewed articles describing or evaluating predictive, prognostic, or risk-stratification approaches relevant to respiratory or occupational health outcomes; (2) studies presenting model development, validation, application, or occupational exposure–risk assessment; and (3) reviews or policy-oriented publications providing relevant methodological or occupa-

tional health context for risk prediction and prevention. Exclusion criteria comprised conference abstracts, non-peer-reviewed reports, studies focused solely on treatment outcomes without predictive or risk-assessment relevance, and publications without sufficient relevance to respiratory risk stratification or occupational health. The retrieved studies were initially screened by title and abstract, followed by full-text review.

Retrieved records were initially screened by title and abstract for relevance to the review objectives, followed by full-text assessment of potentially relevant publications. A thematic mapping approach was used to organize the evidence into four core domains: (1) clinical predictive tools for respiratory disease, (2) exposure-based or occupational risk approaches, (3) statistical and computational modelling techniques, and (4) validation and implementation practices. Findings were synthesised narratively to compare methodological approaches, key predictors, predictive performance where reported, occupational applicability, and important evidence gaps.

Results

As shown in Figure 1, the structured database search identified 618 records. After application of the free full-text filter, 471 records remained for screening. Title and abstract screening identified 141 potentially relevant publications for full-text assessment. Six representative evidence approaches relevant to predictive modelling, risk stratification, occupational exposure–risk assessment, and occupational respiratory health were retained for detailed comparative synthesis in Table 1. Additional seminal publications identified through reference screening provided methodological and historical context where appropriate.

Given the narrative nature of this review and the heterogeneity of the included evidence, a formal quality appraisal using standardized tools was not

conducted. Instead, methodological rigor, relevance, and applicability were considered during evidence selection and synthesis, with attention to study design, population predictors, outcomes, validation methods, and reported model performance.

Clinical prognostic models

Clinical indices such as the BODE Index [12], evaluated by [9], and the DOSE Index [13], validated by [14], were consistently reported as well-validated tools for chronic obstructive pulmonary disease (COPD). The BODE index demonstrated moderate discriminatory performance in a recent clinical evaluation, with an area under the receiver operating characteristic curve (AUC) of 0.716 for identifying severe COPD; a BODE score >3 showed a sensitivity of 62.1% and specificity of 62.6% for hospitalization [9]. Similarly, the DOSE index has demonstrated associations with subsequent exacerbations and adverse clinical outcomes across different COPD cohorts [13,14].

Despite their established utility, both models share a critical limitation: the absence of occupational exposure variables. As summarized in Table 1, neither model incorporates exposure intensity, duration, or workplace environmental conditions, restricting their applicability for early risk identification in occupational settings [15].

Asthma and work-related prediction tools

Asthma-related predictive tools, including the Asthma Predictive Index (API) [16], demonstrate the feasibility of identifying individuals at increased risk of subsequent asthma development. The stringent API demonstrated high specificity for subsequent active asthma, although its sensitivity was comparatively low [16]. The API exhibited high specificity (>80%) [17], while occupational asthma tools achieved sensitivities of approximately 70–80% when serial PEF measurements were incorporated.

However, these approaches are limited by their reliance on self-reported data and symptom-based

assessment, which introduces recall bias and reduces sensitivity for detecting subclinical or low-level occupational exposures. In addition, the API is primarily designed for paediatric populations, limiting its applicability to adult industrial workers.

Exposure-based and occupational models

Exposure–response models demonstrated strong associations between cumulative exposure to dust, fumes, and chemical agents and respiratory outcomes. Studies by [18] and [3] highlighted the importance of exposure intensity, duration, and workplace conditions in determining disease risk. At the policy and surveillance level, the Healthy China 2030 pneumoconiosis strategy described by Wang et al. [3] emphasizes occupational exposure control, workplace prevention, health surveillance, diagnosis, and treatment as complementary components of pneumoconiosis prevention. However, this strategy represents a prevention and policy framework rather than a validated individual-level prediction model. Despite these strengths, exposure-based models are typically not translated into practical risk scoring systems and often lack integration with clinical and demographic variables, limiting their application in routine occupational health surveillance [18].

Data-driven and machine learning approaches

Emerging data-driven approaches, including machine learning models, demonstrated improved predictive performance by integrating multidimensional variables such as clinical parameters, exposure data, and comorbidities. In the systematic review and meta-analysis by Yang et al. [10], which included 17 studies, 148,384 patients, and 64 machine-learning models for acute respiratory distress syndrome prediction, the pooled AUC was 0.7407, with pooled sensitivity and specificity of approximately 0.67 and 0.68, respectively. However, substantial heterogeneity in populations, predictors, modelling approaches, and validation procedures limits direct extrapolation of these findings to occupational respiratory disease prediction.

These approaches offer significant advantages in capturing complex and nonlinear relationships between predictors. However, their application in occupational health remains limited by data availability, lack of standardization, and challenges related to interpretability, particularly in low- and middle-income settings [19].

Synthesis of findings

Across all model categories, several consistent predictors of respiratory disease risk were identified, including age, smoking status, lung function parameters, and exposure-related factors such as duration and intensity of hazardous exposure.

A key finding from this review is the fragmentation of existing predictive approaches. Clinical models demonstrate strong prognostic accuracy but lack occupational relevance, while exposure-based models provide contextual insights without practical risk stratification capability. Data-driven models show high predictive potential but remain underutilized in occupational health due to infrastructural and validation limitations [3,15].

Direct comparison of predictive performance across these approaches should nevertheless be interpreted cautiously because the models were developed for different populations, outcomes, and purposes and employed different validation strategies. BODE and DOSE primarily assess severity or prognosis among patients with established COPD, whereas the API focuses on subsequent asthma development in children. Occupational exposure-response approaches estimate population-level disease risk rather than individual predictive probabilities, while machine-learning studies use heterogeneous clinical populations and outcomes. Accordingly, differences in reported AUCs or other performance measures should not be interpreted as evidence that one approach is inherently superior to another.

Collectively, these findings highlight the absence of a comprehensive framework integrating clinical, occupational, environmental, and behavioural determinants and provide the conceptual basis for the proposed Occupational Respiratory Disease Risk Stratification (ORDiRS) framework.

Discussion

The findings of this review highlight a persistent limitation in occupational respiratory disease (ORD) prevention: the continued reliance on reactive, symptom-based surveillance systems. Across the included studies, predictive tools demonstrated strong performance within clinical contexts but showed limited applicability in occupational settings. This reflects a fundamental mismatch between clinically derived models and the multidimensional nature of workplace exposure. Clinical indices such as the BODE and DOSE scores remain well-validated for prognostic assessment in chronic obstructive pulmonary disease (COPD) [9,14]. However, their design is inherently oriented toward disease progression rather than preclinical risk prediction, and they do not incorporate occupational exposure variables [15]. As a result, these tools are insufficient for identifying at-risk workers before disease onset, reinforcing a reactive approach to surveillance.

Exposure-response models, on the other hand, provide robust evidence linking cumulative exposure to respiratory outcomes [3,18]. Nevertheless, their application is largely confined to population-level analysis and lacks translation into actionable, individual-level risk stratification tools. This limits their utility in routine occupational health practice, where early intervention depends on timely identification of high-risk individuals. Data-driven and machine-learning approaches offer clear methodological advantages by enabling integration of multidimensional predictors and capturing nonlinear relationships [10,11]. However, their adoption in occupational health remains constrained by data fragmentation, limited standardization, and concerns regarding interpretability, particularly in low- and middle-income settings [19].

Taken together, these findings reveal a fragmented predictive landscape, in which clinical, exposure-based, and data-driven models operate in isolation. This fragmentation contributes to delayed detection and missed opportunities for early intervention. A shift toward integrated, data-

driven predictive frameworks is therefore necessary to support proactive occupational health surveillance and prevention.

Towards the development of ORDiRS

The findings of this review highlight a clear need for an integrated, predictive framework capable of addressing the limitations of existing clinical and exposure-based models. In response, the Occupational Respiratory Disease Risk Stratification (ORDiRS) framework is proposed as a data-driven approach to enable early identification of at-risk workers before disease onset.

ORDiRS is conceptualized to integrate five key domains: sociodemographic characteristics, occupational exposure, workplace and organizational factors, clinical indicators, and behavioural risk factors. This multidomain structure addresses the fundamental limitation of current models, which largely exclude occupational determinants and therefore fail to reflect real-world risk profiles.

Methodologically, ORDiRS adopts a hybrid modelling approach, combining traditional statistical methods with data-driven techniques to balance interpretability and predictive performance. This is critical, as purely clinical models lack contextual relevance, while data-driven models without domain integration risk poor generalizability. By enabling preclinical risk stratification, ORDiRS supports a shift from reactive, symptom-based detection to proactive surveillance and targeted prevention. However, its implementation will depend on the availability of integrated datasets, standardization of exposure metrics, and rigorous validation across diverse industrial settings [20,21].

As illustrated in Figure 2, the proposed ORDiRS framework differs conceptually from existing respiratory risk approaches by integrating clinical, occupational exposure, workplace, sociodemographic, and behavioural determinants within a single risk-stratification framework. Existing clinical tools such as BODE and DOSE predominantly address prognosis in patients with established respiratory disease, whereas occupational

exposure-based approaches primarily characterize associations between workplace hazards and disease outcomes. ORDiRS is intended to bridge these approaches and support future individual-level risk stratification before clinically apparent occupational respiratory disease develops. At this stage, however, ORDiRS should be regarded as a conceptual framework that requires empirical development and validation rather than as an established prediction model.

Beyond its methodological contribution, ORDiRS has the potential to improve occupational safety and health by enabling early identification of high-risk workers and timely intervention. This may reduce disease burden, healthcare costs, and productivity losses, while supporting workforce sustainability and regulatory compliance. At a broader level, it can inform evidence-based policy, resource allocation, and industrial risk management, advancing a more preventive and data-driven approach to occupational health. Several implementation challenges should be considered in the future development of ORDiRS. Occupational exposure data may be incomplete, inconsistently recorded, or unavailable in routine occupational health systems, while integration of clinical, workplace, and exposure information may require interoperable data infrastructures. More complex data-driven models may also create challenges related to interpretability, computational requirements, and implementation costs, particularly in resource-limited occupational health settings. Where machine-learning approaches are incorporated, explainable modelling strategies will be important to support clinical interpretation and worker communication. Future implementation should therefore prioritize feasible predictor collection, transparent model outputs, compatibility with existing surveillance systems, and evaluation of resource requirements before routine adoption.

Ethical and governance considerations are also important because occupational risk prediction may involve sensitive clinical and employment-

related information. Appropriate data protection, confidentiality, transparent use of predicted risk, and safeguards against discriminatory employment decisions should therefore be incorporated into future development and implementation. Risk estimates should support preventive occupational health interventions rather than be used in isolation to determine employment suitability.

External validation will be essential before ORDiRS can be considered for clinical or occupational health implementation. Model development should ideally be followed by validation across different industries, occupational exposures, geographic settings, and worker populations to assess discrimination, calibration, and generalizability. Particular attention should be given to settings with different exposure profiles and healthcare resources, as predictor availability and baseline respiratory risk may vary substantially between populations.

Overall, the proposed ORDiRS framework provides a conceptual approach for integrating clinical and workplace determinants relevant to occupational respiratory risk stratification. Its potential value lies in shifting the focus from predominantly reactive disease detection towards earlier identification of workers who may benefit from enhanced surveillance or preventive interventions. However, its predictive utility, feasibility, and clinical value remain to be established through model development, internal and external validation, and prospective evaluation in occupational populations.

Conclusion

This narrative review highlights an important gap between established clinical prediction tools and occupational exposure-based approaches for respiratory risk assessment. Existing clinical models provide useful prognostic information but have limited integration of workplace exposure determinants, whereas occupational approaches capture relevant exposure-related risks but are not

commonly translated into practical individual-level risk-stratification tools. Emerging data-driven methods offer additional opportunities for integrating multidimensional predictors, although challenges related to model validation, interpretability, data availability, and implementation remain.

The proposed Occupational Respiratory Disease Risk Stratification (ORDiRS) framework provides a conceptual foundation for integrating clinical, occupational exposure, workplace, socio-demographic, and behavioural determinants within a unified risk-stratification approach. Its potential utility in supporting earlier identification of workers at increased respiratory risk requires further investigation through empirical model development, internal and external validation, and prospective evaluation across diverse occupational populations. Future research should therefore focus on establishing the predictive performance, feasibility, and practical value of ORDiRS before its implementation in routine occupational health surveillance.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this review paper.

Acknowledgements

The authors would like to acknowledge the Faculty of Medicine, Universiti Teknologi MARA (UiTM) Sungai Buloh, for the institutional support and facilities provided throughout this study. The authors also thank colleagues and occupational health professionals for their technical guidance and academic support during the preparation of this manuscript.

Authors' contributions

MSA conceptualised the study, conducted the literature review, synthesised the findings, and wrote the manuscript. SMY contributed to study design, methodology development, and critical revision. MRI provided expertise in occupational

health interpretation and contributed to manuscript review and validation of the conceptual framework. NI assisted in data interpretation, methodological refinement, and manuscript editing. EZZ contributed to critical appraisal of the literature and supervised the academic direction of the study. MIMS assisted in literature screening, data organisation, and manuscript preparation. All authors reviewed and approved the final version of the manuscript.

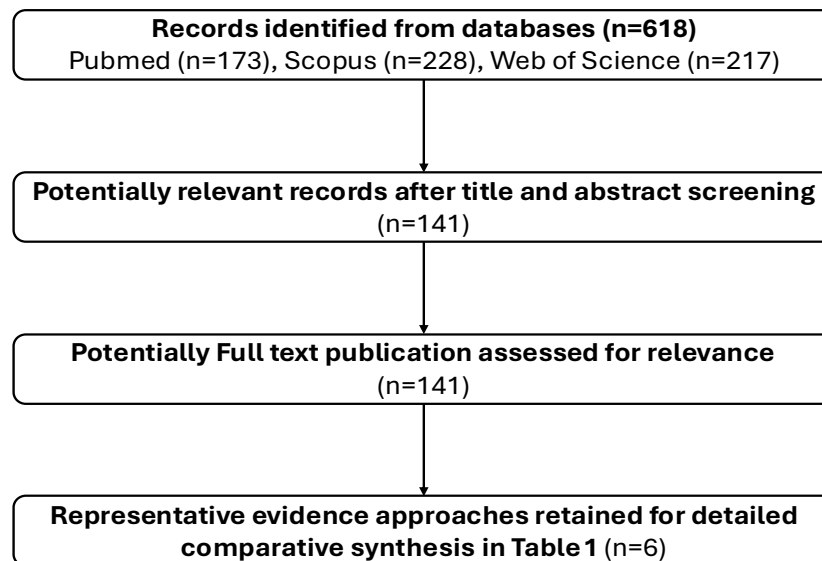


Figure 1. Literature Search and Evidence Selection Flow Diagram

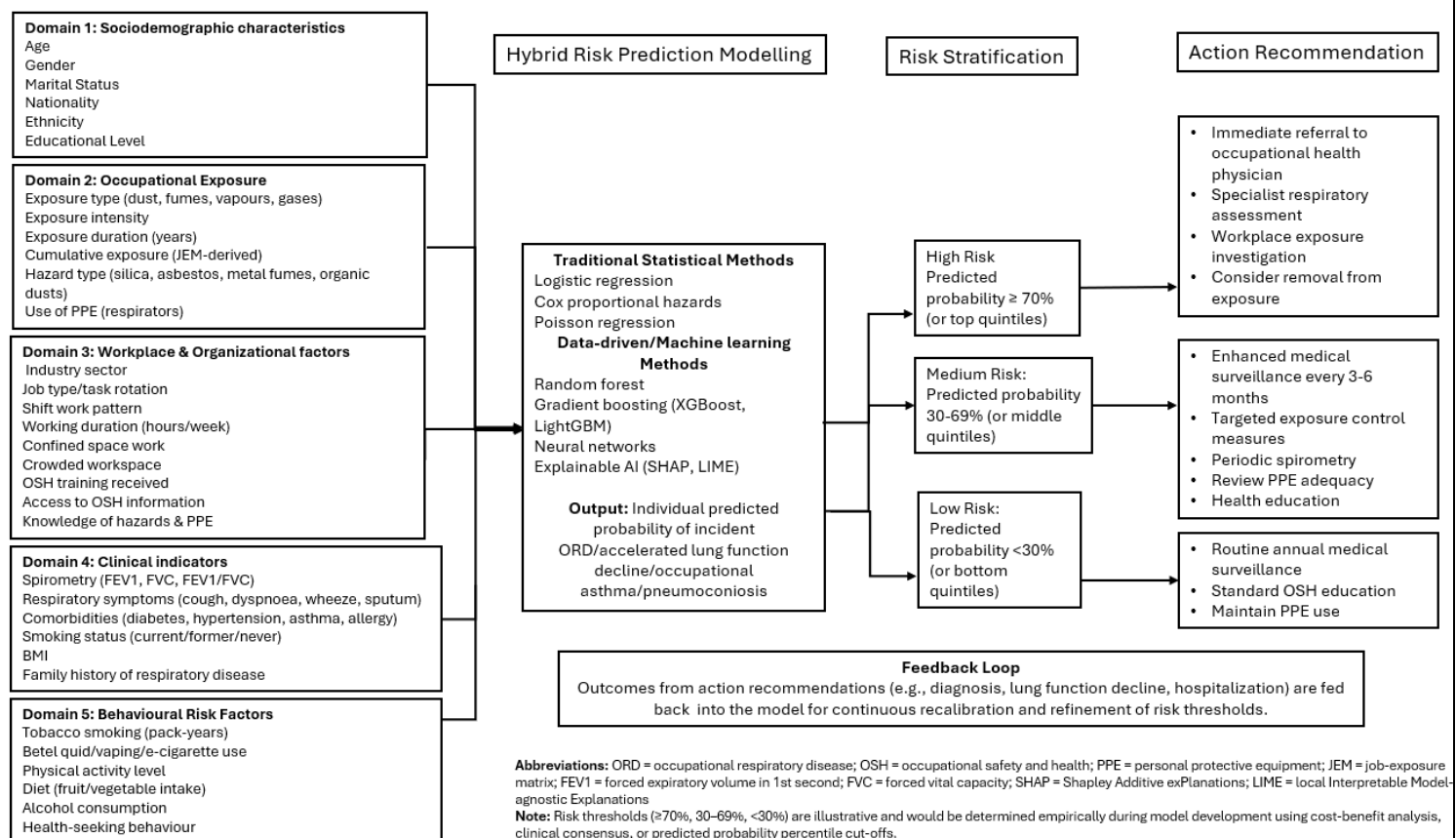


Figure 2. Conceptual framework of the Occupational Respiratory Disease Risk Stratification (ORDiRS)

Table 1. Summary of Existing Predictive and Risk Stratification Tools in Respiratory Medicine and Their Applicability to Occupational Contexts

Author	Tool / Approach	Study Design	Country	Sample Size	Key Predictors	Outcome /	Validation Method	Performance	Occupational Relevance	Key Limitation
Celli et al., 2022[12]; Ajay et al., 2025 [9]	BODE Index	Celli: narrative clinical review; Ajay: observational clinical evaluation	India; patients with COPD	n=120 in Ajay et al.	BMI, FEV ₁ /air-flow obstruction, dyspnoea, 6-min walk distance	COPD severity and hospitalization risk	ROC-based clinical evaluation against GOLD severity and hospitalization; not an external validation of a newly developed model	AUC 0.716 for severe COPD; BODE >3: sensitivity 62.1%, specificity 62.6% for hospitalization	Demonstrates value of multi-dimensional clinical assessment but contains no occupational exposure variables	Prognostic/severity-oriented rather than preclinical disease-onset prediction; limited direct occupational applicability
Jones et al., 2009; [13], Li et al., 2024 [14]	DOSE Index	Model derivation and multicohort validation; subsequent retrospective cross-sectional evaluation	UK derivation; validation in UK, Netherlands and Japan; Li study in Taiwan; patients with COPD	n=375 derivation cohort; Li: n=396	Dyspnoea, air-flow obstruction, smoking status, exacerbation frequency	COPD severity, future exacerbations, healthcare utilization/costs	Cross-sectional and longitudinal validation across multiple healthcare settings; subsequent real-world clinical evaluation	DOSE ≥4 associated with hospital admission OR 8.3 (4.1–17.0) and respiratory failure OR 7.8 (3.4–18.3); predicted subsequent-year exacerbations (P≤0.014). Li found higher DOSE quartiles associated with hospitalization and costs	Smoking is relevant to respiratory risk, but workplace exposure intensity, duration, and hazard type are absent	Primarily assesses established COPD severity/prognosis rather than occupational disease onset
Castro-Rodríguez et al., 2000 [16]	Asthma Predictive Index (API)	Longitudinal birth-cohort prediction-index development	USA; Tucson Children's Respiratory Study; children followed from infancy to school age	n=1,246 enrolled; complete data: stringent API n=1,002; loose API n=986	Early/recurrent wheeze, parental asthma, eczema, eosinophilia, wheezing without colds, allergic rhinitis	Prediction of active asthma at ages 6–13 years	Longitudinal assessment of prediction at multiple follow-up ages within the cohort	Stringent API specificity approximately 96–97%, sensitivity 15–28% across ages; positive stringent API associated with 4.3–9.8-fold greater likelihood of later active asthma	Illustrates pre-clinical risk stratification, relevant conceptually to ORDiRS	Paediatric-specific; predictors are largely clinical/atopic; no workplace or occupational exposure determinants
Mehta et al., 2012 [18]	Occupational exposure	Prospective population-based cohort	Switzerland; working-age	n=4,267	Biological dust, mineral dust, gases/fumes and combined	Incident spirometry-defined	Longitudinal exposure–outcome analysis using	High occupational exposure levels were associated	Strong direct evidence that occupational exposure	Provides population-level exposure–risk estimation rather than an

	posure–response approach	study (SAPALDIA)	non-asthmatic adults aged 18–62		VGDF exposure assigned by job-exposure matrix; smoking and other co-variables	COPD over longitudinal follow-up	mixed Poisson regression; not an individual prediction-model validation study	with approximately 2–5-fold higher risk of stage II+ COPD; associations persisted in analyses restricted to non-smokers	posure type/intensity contributes independently to respiratory disease risk	actionable individual risk score; limited integration with multidomain clinical prediction
Wang et al., 2023 [3]	Healthy China 2030 pneumoconiosis prevention strategy	Commentary / occupational prevention and policy framework	China; occupational populations at risk of pneumoconiosis	N/A	Dust exposure control, occupational health surveillance, diagnosis, treatment, workplace prevention and policy measures	Pneumoconiosis prevention and control	Not applicable — no predictive model was developed or validated	No discrimination/calibration statistics reported; article describes national prevention and control strategies rather than model performance	Highly relevant contextual evidence for integrating workplace prevention and surveillance factors into occupational respiratory risk frameworks	Not a risk-prediction model and should not be interpreted as a nationally validated predictive scoring system
Yang et al., 2025 [10]	Machine-learning prediction models for ARDS	Systematic review and meta-analysis of prediction-model studies	Multiple countries; ICU and hospitalized clinical populations	17 studies; 148,384 patients; 64 models	Heterogeneous clinical, laboratory and physiological predictors across ML models	Prediction of acute respiratory distress syndrome	Internal/external validation methods varied across included studies; risk of bias assessed using PROBAST	Pooled AU-ROC 0.7407; sensitivity 0.67 (95% CI 0.66–0.67); specificity 0.68 (95% CI 0.67–0.68). Substantial heterogeneity was present	Demonstrates capacity and limitations of data-driven multidimensional respiratory prediction	ARDS is predominantly an acute clinical outcome rather than an occupational respiratory disease; model heterogeneity, high/unclear risk of bias, and limited occupational generalizability

References

- [1]. Minov J. Occupational chronic obstructive pulmonary disorder: prevalence and prevention. *Expert Rev Respir Med.* 2022;16(4):429-436. <https://doi.org/10.1080/17476348.2021.2011722>
- [2]. Nishida C, Yatera K. The impact of ambient environmental and occupational pollution on respiratory diseases. *Int J Environ Res Public Health.* 2022;19(5):2788. <https://doi.org/10.3390/ijerph19052788>
- [3]. Wang H, Ye Q, Zhang H, et al. Prevention and treatment of pneumoconiosis in the context of Healthy China 2030. *China CDC Wkly.* 2023;5(41):927-933. <https://doi.org/10.46234/ccdcw2023.162>
- [4]. Mostafa IM, Gamal K, Kader MAA, Ruby D. Evaluation of occupational and nonoccupational interstitial lung disease in railway workers. *Egypt J Chest Dis Tuberc.* 2024. https://doi.org/10.4103/ecdt.ecdt_37_23
- [5]. Jongkumchok W, Chaiear N, Boonsawat W, Chaisuksant S. Occupations and causative agents among work-related asthma patients in Easy Asthma Clinic, Srinagarind. *Am J Public Health Res.* 2018;6:111-116. <https://doi.org/10.12691/ajphr-6-2-14>
- [6]. Asraff MS, Azzani M, Anuar MR, et al. Prevalence of occupational respiratory disease and its determinants among workers in major industrial sectors in Malaysia in 2023. *Sci Rep.* 2025;15:25807. <https://doi.org/10.1038/s41598-025-10365-8>
- [7]. Abdullah Hair AF, Mohd Roze A, Syed Jamaludin MS, Nik Mohamed NMN, Samsuddin N, Razali A, et al. Prevalence of occupational diseases among small and medium industry workers in Malaysia: a systematic review. *J Clin Health Sci.* 2019;4(2):4-30. doi:10.24191/jchs.v4i2.7508
- [8]. Awaluddin SM, Mahjom M, Lim KK, et al. Occupational disease and injury in Malaysia: a thematic review of literature from 2016 to 2021. *J Environ Public Health.* 2023;2023:1798434. <https://doi.org/10.1155/2023/1798434>
- [9]. Ajay D, Bhatnagar M, Chandra N. BODE index: a predictor of hospitalization and severity in chronic obstructive pulmonary disease patients. *Bioinformation.* 2025;21(7):18311835. <https://doi.org/10.6026/973206300211831>
- [10]. Yang J, et al. Predictive modeling of acute respiratory distress syndrome using machine learning: systematic review and meta-analysis. *J Med Internet Res.* 2025;27:e66615. doi:10.2196/66615
- [11]. Shen X, Liu H. Using machine learning for early detection of chronic obstructive pulmonary disease: a narrative review. *Respir Res.* 2024;25(1):336. <https://doi.org/10.1186/s12931-024-02960-6>
- [12]. Celli BR, Singh D, Vogelmeier C, Agusti A. New perspectives on chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis.* 2022;17:2127-2138. <https://doi.org/10.2147/COPD.S365771>
- [13]. Jones RC, Donaldson GC, Chavannes NH, et al. Derivation and validation of a composite index of severity in chronic obstructive pulmonary disease: the DOSE index. *Am J Respir Crit Care Med.* 2009;180(12):1189-1195. <https://doi.org/10.1164/rccm.200902-0271OC>
- [14]. Li CL, Chang HC, Tseng CW, et al. The DOSE index in chronic obstructive pulmonary disease: evaluating healthcare costs. *BMC Pulm Med.* 2024;24(1):560. <https://doi.org/10.1186/s12890-024-03368-0>
- [15]. Espantoso-Romero M, Roman Rodriguez M, Duarte-Perez A, et al. External validation of multidimensional prognostic indices (ADO, BODEx and DOSE) in a primary care international cohort (PRO-EPOC/COPD cohort). *BMC Pulm Med.* 2016;16(1):143. <https://doi.org/10.1186/s12890-016-0314-1>

- [16]. Castro-Rodriguez JA, Holberg CJ, Wright AL, Martinez FD. A clinical index to define risk of asthma in young children with recurrent wheezing. *Am J Respir Crit Care Med*. 2000;162(4):1403-1406. <https://doi.org/10.1164/ajrccm.162.4.9912111>
- [17]. Romero-Tapia SJ, Guzman Priego CG, Del-Rio-Navarro BE, Sanchez-Solis M. Advances in the relationship between respiratory viruses and asthma. *J Clin Med*. 2023;12(17):5679. <https://doi.org/10.3390/jcm12175679>
- [18]. Mehta AJ, Miedinger D, Keidel D, et al. Occupational exposure to dusts, gases, and fumes and incidence of chronic obstructive pulmonary disease in the Swiss Cohort Study on Air Pollution and Lung and Heart Diseases in Adults. *Am J Respir Crit Care Med*. 2012;185(12):1292-1300. <https://doi.org/10.1164/rccm.201110-1917OC>
- [19]. Holt, K. L. (2025). Predictive Modeling of Occupational Exposure Using Machine Learning and Environmental Sensor Data. *Journal of Exceptional Multidisciplinary Research*, 2(1), 82-89. <https://doi.org/10.69739/jemr.v2i1.617>
- [20]. Riley RD. *Prognosis Research in Healthcare: Concepts, Methods, and Impact*. Oxford University Press; 2020. <https://doi.org/10.1093/med/9780198766610.001.0001>
- [21]. Collins GS, Reitsma JB, Altman DG, Moons KGM; TRIPOD Group. Transparent reporting of a multi-variable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Circulation*. 2015;131(2):211-219. <https://doi.org/10.1161/CIRCULATIONAHA.114.014508>.