

PROTOCOL

Open Access



# Predicting adverse outcomes in adults with a community-acquired lower respiratory tract infection: a protocol for the development and validation of two prediction models for (i) all-cause hospitalisation and mortality and (ii) cardiovascular outcomes

Merijn H. Rijk<sup>1\*</sup> , Tamara N. Platteel<sup>1</sup>, Geert-Jan Geersing<sup>1</sup>, Monika Hollander<sup>1</sup>, Bert L. G. P. Dalmolen<sup>2</sup>, Paul Little<sup>3</sup>, Frans H. Rutten<sup>1</sup>, Maarten van Smeden<sup>4†</sup> and Roderick P. Venekamp<sup>1†</sup>

## Abstract

**Background** Community-acquired lower respiratory tract infections (LRTI) are common in primary care and patients at particular risk of adverse outcomes, e.g., hospitalisation and mortality, are challenging to identify. LRTIs are also linked to an increased incidence of cardiovascular diseases (CVD) following the initial infection, whereas concurrent CVD might negatively impact overall prognosis in LRTI patients. Accurate risk prediction of adverse outcomes in LRTI patients, while considering the interplay with CVD, can aid general practitioners (GP) in the clinical decision-making process, and may allow for early detection of deterioration. This paper therefore presents the design of the development and external validation of two models for predicting individual risk of all-cause hospitalisation or mortality (model 1) and short-term incidence of CVD (model 2) in adults presenting to primary care with LRTI.

**Methods** Both models will be developed using linked routine electronic health records (EHR) data from Dutch primary and secondary care, and the mortality registry. Adults aged  $\geq 40$  years with a GP-diagnosis of LRTI between 2016 and 2019 are eligible for inclusion. Relevant patient demographics, medical history, medication use, presenting signs and symptoms, and vital and laboratory measurements will be considered as candidate predictors. Outcomes of interest include 30-day all-cause hospitalisation or mortality (model 1) and 90-day CVD (model 2). Multivariable elastic net regression techniques will be used for model development. During the modelling process, the incremental predictive value of CVD for hospitalisation or all-cause mortality (model 1) will also be assessed. The models will be validated through internal-external cross-validation and external validation in an equivalent cohort of primary care LRTI patients.

<sup>†</sup>Maarten van Smeden and Roderick P. Venekamp contributed equally to this work.

\*Correspondence:

Merijn H. Rijk

m.h.rijk@umcutrecht.nl

Full list of author information is available at the end of the article



**Discussion** Implementation of currently available prediction models for primary care LRTI patients is hampered by limited assessment of model performance. While considering the role of CVD in LRTI prognosis, we aim to develop and externally validate two models that predict clinically relevant outcomes to aid GPs in clinical decision-making. Challenges that we anticipate include the possibility of low event rates and common problems related to the use of EHR data, such as candidate predictor measurement and missingness, how best to retrieve information from free text fields, and potential misclassification of outcome events.

**Keywords** Lower respiratory tract infection, Cardiovascular disease, Primary care, Electronic Health Record, Prognosis, Prediction model, Hospitalisation, Mortality

## Background

Community-acquired lower respiratory tract infections (LRTI), such as acute bronchitis and pneumonia, are common reasons for primary care consultations. Prognosis is generally favourable, allowing the majority of patients to be managed in primary care with or without antibiotic treatment depending on disease severity and suspected pathogen [1–3]. Adverse outcomes such as hospitalisation or mortality occur in less than 1% of patients with uncomplicated LRTI, i.e. not suggestive of pneumonia, and antibiotics seem not to reduce the occurrence of these outcomes [4]. The risk of complications is far more pronounced in patients with community-acquired pneumonia (CAP) [5], but identifying these patients in primary care can be challenging [6, 7]. In addition, concurrent cardiovascular diseases (CVD) have also been linked to poor prognosis in patients with LRTI, which is supported by recent literature on coronavirus disease 2019 (COVID-19) [8–10].

On the other hand, LRTIs may increase the risk or trigger the occurrence of CVD—and thromboembolic events in particular—such as acute myocardial infarction (AMI), stroke, pulmonary embolism, and deep venous thrombosis for several months after the acute phase of the disease [11, 12]. Activation of the immune system by acute infection is thought to trigger the interaction between inflammatory and prothrombotic pathways (i.e. immunothrombosis) [13]. Infections with respiratory pathogens such as influenza and SARS-CoV-2 have particularly been associated with increased CVD incidence, even in mildly affected patients managed in primary care [14, 15]. For example, the incidence of AMI was found to increase sixfold during the first week after influenza infection [16].

Accurate prediction of the risk of adverse outcomes can aid general practitioners (GP) in identifying LRTI patients in whom close follow-up or (antibiotic) treatment is warranted. Well-known prediction models—such as the Pneumonia Severity Index (PSI) and CURB-65—have been developed in hospitalised patients and primary care validation of these models is hampered by the inclusion of advanced laboratory and radiographic variables [17, 18]. The CRB-65, including confusion, respiratory

rate, blood pressure, and age, is proposed as a primary care alternative to predict mortality but has been incompletely validated hampering implementation in primary care [19–22]. Another primary care-derived model including diagnosis, age, heart failure, diabetes, use of oral glucocorticoids, number of hospitalisations in the previous year, and antibiotic use in the previous month to predict hospitalisation and mortality also suffers from limited validation and prediction modelling methods have advanced since the development of both models [8]. A model suitable for routine use in primary care is therefore currently lacking. Here we present the design of the development and external validation of two prediction models aimed at estimating the individual risk of all-cause hospitalisation or mortality (model 1), and short-term incidence of CVD (model 2) in adult primary care LRTI patients using linked routine electronic health records (EHR) data from Dutch primary and secondary care, and the mortality registry.

## Method

### Study design and setting

This prognostic model development and validation study will make use of pseudonymised routine EHR data. The model development cohort will be derived from the Julius General Practitioners' Network (JGPN) [23], which covers approximately 450,000 Dutch inhabitants, representative of the Dutch population, enlisted in both urban and rural practices in the region of Utrecht. It contains data on patient demographics, consultations (free text including anamnesis, physical examination, and prescribed treatments), coded disease episodes and medical history (using the International Classification of Primary Care (ICPC) [24]), coded prescriptions (using Anatomical Therapeutic Chemical (ATC) codes [25]), data on influenza and pneumococcal vaccinations, coded measurements, and laboratory results. The primary care EHR data will be enriched with linked data on emergency department visits and hospital admissions (from Dutch Hospital Data (DHD) [26]) and mortality (from the National Mortality Registry of Statistics Netherlands (CBS)), resulting in a comprehensive database that covers

**Table 1** Details on data sources that will be used for model development and validation

Data source	Characteristics			Type of data				Outcomes				Intended use	
	Approximate number of participants	Setting	Demographics	Patient history	Medication use	Measurements (vital and laboratory)	Free text	Hospitalisation	Mortality	Cardiovascular events	D/V		
JGPN	0.45M	Primary care	+	+	+	+	+			+		D	
DHD	17.5M	Hospital		+				+		+		D and V	
National Mortality Registry (CBS)	17.5M	General population	+						+	+		D and V	
ANHA	0.6M	Primary care	+	+	+	+	+			+		V	

Abbreviations: ICN Intercity network, DHD Dutch Hospital Data, CBS Statistics Netherlands, ANHA Academic Network of General Practitioners Amsterdam, M Million, D Development, V Validation

relevant individual disease trajectories. Details on the various data sources and their coverage are presented in Table 1.

### Participants

All patients aged 40 years and older who presented to a GP affiliated to JGPN with an LRTI between 1 January 2016 and 31 December 2019 will be included in the model development cohort. A GP-diagnosed LRTI is defined as the registration of an ICPC code for either pneumonia (R81) or acute bronchitis (R78). Only the first episode of individual patients within the study period will be included. An LRTI-related consultation after a period of 28 days without such consultations is considered a new episode. For external validation, a similarly defined but more recent (i.e. 2022-2023) cohort of primary care LRTI patients will be derived from the Academic Network of General Practitioners from the region of Amsterdam (ANHA) which has a data structure and coverage of healthcare domains similar to JGPN [27].

### Outcomes of interest

We will develop and validate two prediction models that estimate individual risk of 30-day all-cause hospitalisation or mortality (yes/no; model 1) and CVD within 90 days (yes/no; model 2) (Fig. 1).

For model 1, registration of hospitalisation (DHD) and mortality (CBS) within 30 days after LRTI diagnosis will be extracted, irrespective of diagnosis or cause. For model 2, the composite outcome consists of CVD-related mortality and acute (arterial and venous) thromboembolic events within 90 days after LRTI diagnosis: data on CVD-related mortality, acute coronary syndrome, cerebrovascular accident and pulmonary embolism will be extracted from CBS and DHD data using ICD-10 diagnosis codes [28]. Additionally, data on transient ischaemic attack and deep venous thrombosis will be retrieved from DHD (using ICD-10 coding) and JGPN (based on ICPC coding and free text fields of consultations), as these events can be managed in both primary and secondary care. We will additionally explore the option of including exacerbations of heart failure in the outcome for model 2, depending on the feasibility and validity of retrieving such events from primary care free text data. Since outcome events will be collected from a national registry and follow-up time will be short, the number of missing outcome events will likely be minimal and a non-survival model will be used in our analysis.

### Candidate predictors

Initial candidate predictor selection will be based on a review of the literature and clinical expertise. Candidate predictors will be measured at GP diagnosis of LRTI (i.e.

moment of prediction). We will consider the following categories as candidate predictors: demographics (e.g. age, sex), patient history (e.g. smoking status, comorbidities) and chronic medication use (e.g. immunosuppressants, inhalation medication). In addition, we will explore the feasibility of retrieving and—where relevant—the added predictive value of candidate predictor data on signs and symptoms (e.g. shortness of breath, fever), measurements (e.g. oxygen saturation, respiratory rate), and laboratory tests (e.g. point-of-care C-reactive protein (CRP) measured at diagnosis) from free text fields. An overview of all variables that will be considered as candidate predictors can be found in Additional file 1.

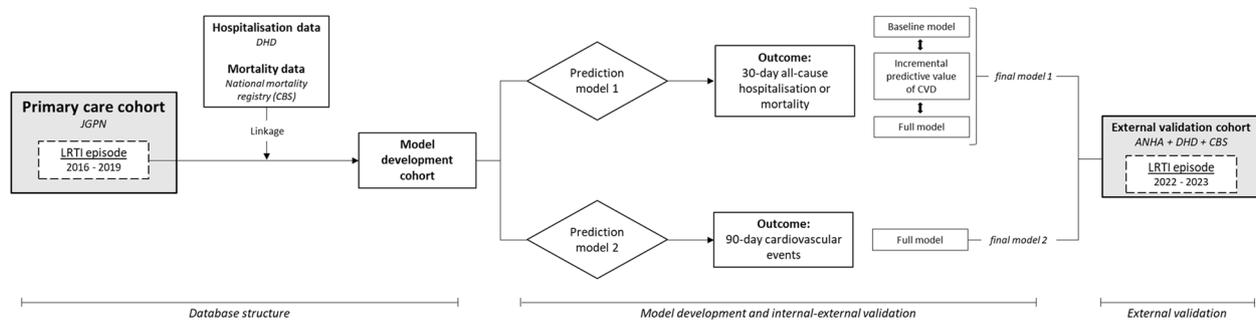
### Sample size

Calculations of the required sample sizes for model development are based on estimated event fractions of the outcomes of both prediction models. A total of 15 candidate predictors, interaction terms included, is anticipated. Sample size calculations were performed using the 'pmsampsize' package [29] in R version 4.2.2 [30], targeting a maximum shrinkage of 10% to minimize potential overfitting. In the absence of reported R-squared values of previously developed models, we aim at developing models with a minimal anticipated c-statistic (area under the ROC curve) of 0.70. For model 1, the minimal required sample size is 8,635 LRTI episodes assuming a conservative event rate of 3% [1, 22, 31]. For model 2, estimation of the expected cardiovascular event rate is more difficult due to varying definitions and follow-up periods in previous reports [32]. Assuming an event rate of 2.5%, the minimal required sample size is 10,231 LRTI episodes.

The incidence of LRTI in Dutch primary care is estimated at 27.8 episodes per 1000 person-years [33]. With approximately 50% of JGPN participants aged 40 years or older [23] and assuming a similar incidence rate in this population, we anticipate around 25,000 LRTI episodes in our development cohort which would be more than sufficient given our sample size calculations.

### Missing data

Inherent to the use of EHR data, missing data is anticipated. We will consider the absence of registered comorbidities and prescriptions as the absence of the condition or medication use. For all other candidate predictors, we will assess the proportion of cases with missing data and its assumed mechanism. Where appropriate, if the missing at-random assumption is met, missing data will be handled using appropriate techniques, such as multiple imputations with chained equations [34].



**Fig. 1** Graphical representation of database structure and anticipated prediction model development process. Abbreviations: JGPN, Julius General Practitioners' Network; LRTI, lower respiratory tract infection; DHD, Dutch Hospital Data; CBS, Statistics Netherlands; ANHA, Academic Network of General Practitioners Amsterdam

### Statistical analysis

Candidate predictor selection will be based on existing literature, clinical expertise, data availability, and a maximum number of candidate predictors according to the sample size calculation. Restricted splines will be considered for continuous predictors, such as age. Both prediction models will be developed using multivariable elastic net regression, accounting for predictor selection during the modelling process.

Model 1 will be developed using an incremental predictive value approach. First, a baseline model (i.e. including age, sex, and an interaction term) will be fitted. Next, the incremental predictive information captured by CVD comorbidity is explored by forcing CVD comorbidities into the model. Finally, a model will be built using all candidate predictors. Model performance of these subsequent models will be compared based on the difference in c-statistics calculated by bootstrapping, change in the distribution of risks, and change in R-squared (pseudo-R-squared), allowing us to choose a final model based on model performance and suitability for use in clinical practice (e.g. a complex model with only a slight increase in performance might not be preferable). For model 2, the final model will directly be developed by forcing all selected candidate predictors into the model. Model performance of the final models will be assessed using quantitative measures of discrimination (c-statistic) and calibration (intercept, slope, and flexible calibration plot), the Cox-Snell R-squared, and decision curve analysis.

The final models will initially be validated through internal-external cross-validation—both geographical and temporal—to assess the heterogeneity of predictor effects by place and time. Subsequently, both final models will be externally validated in the ANHA cohort. Measures of discrimination and calibration will be assessed and, if necessary, the models will be updated using the validation cohort, which would require additional validation [35]. All analyses will be performed in R version

4.2.2 [30], and while reporting we will adhere to the TRIPOD statement (Additional file 2) [36].

### Discussion

In this paper, we present the rationale and design for the development and external validation of two prediction models that can aid GPs in identifying primary care LRTI patients with an increased risk of adverse outcomes. By assessing both the predictive value of CVD for adverse outcomes and the occurrence of cardiovascular outcome events these models specifically address the role of CVD in the prognosis of LRTI.

Currently, available models that stratify LRTI patients based on the risk of poor prognosis are either designed for use in hospitalised patients or suffer from limited validation for use in primary care [8, 19–22, 37, 38], and predicted outcomes are limited to hospitalisation and mortality. In addition, advances in the field of prediction research have led to more sophisticated model development and validation methods. Rather than updating existing models, we therefore aim to include the predictors of these models as candidate predictors while developing new prediction models using state-of-the-art development and validation techniques.

The burden of CVD following an LRTI episode has received considerable attention in hospitalised patients with CAP, and the most frequently observed events include exacerbation of heart failure, atrial fibrillation, and acute coronary syndrome [32, 39]. Some specific respiratory pathogens, such as influenza and SARS-CoV-2, are particularly associated with cardiovascular complications [14, 15, 40]. These observations are not limited to hospitalised patients alone, since several primary care-based studies on patients with RTI also revealed an increased incidence of AMI and stroke up to 90 days after initial infection [41]. Regarding individual prognostic factors, a history of hypertension and a QRISK2 score—a 10-year cardiovascular risk tool—of > 10% were

found to be associated with an increased risk of cardiovascular events following an RTI [42]. However, tools for individual risk-stratification of CVD following an LRTI episode in primary care are currently lacking.

Our models will be developed in a period prior to the COVID-19 pandemic. During the first years of the pandemic, aetiology of RTIs was largely reduced to SARS-CoV-2 infection, and circulation of this virus in a naïve population resulted in a wave of severely ill patients [43]. Consequently, hospitalisation and mortality rates among LRTI patients have been relatively high during this period, whereas cardiovascular complications are also more frequent with increasing COVID-19 disease severity [15]. The burden of COVID-19 in the Netherlands has declined ever since as a result of vaccination strategies, natural immunisation after infection and the shift towards the predominance of the less severe Omicron variant. We therefore expect that the epidemiology of primary care LRTIs in the current (post-pandemic) era largely resembles that of the pre-pandemic years in which our models will be developed, which would lead to a stable model performance. This, however, warrants confirmation by our intended external validation in a post-pandemic cohort of primary care LRTI patients.

### Strengths, challenges and limitations

Strengths of the design of our study include the large cohorts used for model development and external validation that derive from a population that is representative of Dutch primary care, the enrichment with linked data on hospitalisation and mortality, and the state-of-the-art assessment of model performance on internal-external and external validation.

Nevertheless, we do anticipate several challenges during the study process. First, the use of routine EHR data requires proper handling of missing data, which in turn depends on the assumed mechanism of missingness [44]. JGPN-affiliated practices receive education on the proper coding of medical history and prescriptions using the ICPC and ATC coding systems, resulting in a rich database. It seems therefore appropriate to consider the absence of such registrations as negative values. Handling of missing data might however be more complex for candidate predictors that will potentially be retrieved from free text fields of index consultations, such as signs, symptoms, measurements, and laboratory test results. We aim to explore the registration of free text-derived parameters and the challenges this introduces in a random sample of the study population prior to embarking on the process of model development with candidate predictors retrieved from free text fields.

A second possible challenge that we consider is a low outcome event rate for model 2 (predicting

cardiovascular outcomes), for which we estimated an event rate of 2.5%. If the actual occurrence of CVD following an LRTI episode proves to be lower we may refrain from developing a prediction model, since it is challenging to maintain a high predictive performance in case of rare outcome events [45]. In such an event, we will consider an alternative approach to identify patient and disease characteristics that are associated with an increased risk of CVD following an LRTI by comparing CVD incidence rates among subgroups based on various characteristics, such as patient demographics, comorbidities, and medication use.

Lastly, the use of routine EHR data potentially introduces misclassification on the level of candidate predictors (e.g. medical history), study population (i.e. definition of LRTI episode), and outcome events (e.g. CVD-related mortality). To mitigate potential validity problems due to misclassification, the models should ideally be implemented in a context with similarly structured input data from EHRs. If the models prove to be a safe and valuable addition to the clinical decision-making process, this ultimately results in real-time predicted risks of adverse outcomes in primary care LRTI patients. The potential for developing and implementing such prediction models for LRTI patients is also addressed in the pneumonia guidelines of the American Thoracic Society and Infectious Diseases Society of America [46].

### Conclusion

Community-acquired LRTIs are common in primary care, and patients with increased risk of adverse outcomes are challenging to identify. Currently, existing prediction models for adverse outcomes only focus on hospitalisation and mortality and suffer from incomplete model validation, hampering implementation in clinical practice. The importance of CVD for the prognosis of LRTIs is proposed by both its association with overall poor prognosis and the observed increased incidence of CVD following the initial infection. While considering the interplay between LRTI and CVD, we aim to develop and externally validate two prediction models that predict clinically relevant outcomes such as cardiovascular events, hospitalisation, and mortality. These models can aid GPs in stratifying the risk of poor prognosis in primary care LRTI patients, which may ultimately allow for early detection and prevention of deterioration.

### Abbreviations

LRTI	Lower respiratory tract infections
CAP	Community-acquired pneumonia
CVD	Cardiovascular disease
COVID-19	Coronavirus disease 2019
AMI	Acute myocardial infarction
GP	General practitioner

PSI	Pneumonia Severity Index
EHR	Electronic health record
JGPN	Julius General Practitioners' Network
ICPC	International Classification of Primary Care
ATC	Anatomical Therapeutic Chemical
DHD	Dutch Hospital Data
CBS	Statistics Netherlands ( <i>Dutch: Centraal Bureau voor de Statistiek</i> )
CRP	C-reactive protein
ANHA	Academic Network of General Practitioners Amsterdam

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41512-023-00161-1>.

**Additional file 1.** Overview of variables considered as candidate predictors.

**Additional file 2.** TRIPOD Checklist: Prediction Model Development and Validation.

## Acknowledgements

Not applicable.

## Authors' contributions

All authors were involved in the conceptualisation and design of the study. MHR wrote the first draft of the manuscript, which was initially reviewed by TNP, MvS, and RPV. All authors read and approved the final manuscript.

## Funding

This work was supported by ZonMw (grant number 08391052110003). ZonMw had no involvement in the design of the study or the writing of the manuscript.

## Availability of data and materials

The data that will be used for this study is available from the Dutch primary care registries (JGPN and ANHA), hospital registry (DHD), and the national mortality registry. Access to these data is restricted and has been granted under license of the current study. Data are therefore only available from the authors upon reasonable request and after formal permission of the respective registries.

## Declarations

### Ethics approval and consent to participate

Due to the design of the study and the use of pseudonymised routine electronic health records data, the need for a formal ethical review of the protocol was waived by the Medical Ethical Review Committee NedMec, the Netherlands (protocol number 22/623).

### Consent for publication

Not applicable.

### Competing interests

Maarten van Smeden is an associate editor at Diagnostic and Prognostic Research.

### Author details

<sup>1</sup>Department of General Practice & Nursing Science, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands. <sup>2</sup>Patient and Public Involvement member, Utrecht, the Netherlands. <sup>3</sup>Primary Care Research Center, Primary Care Population Sciences and Medical Education Unit, University of Southampton, Southampton, United Kingdom. <sup>4</sup>Department of Epidemiology & Health Economics, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands.

Received: 8 June 2023 Accepted: 10 October 2023  
Published online: 07 December 2023

## References

1. Snijders BEP, van der Hoek W, Stirbu I, van der Sande MAB, van Gageldonk-Lafeber AB. General practitioners' contribution to the management of community-acquired pneumonia in the Netherlands: a retrospective analysis of primary care, hospital, and national mortality databases with individual data linkage. *Prim Care Respir J*. 2013;22(4):400–5.
2. Little P, Stuart B, Moore M, Coenen S, Butler CC, Godycki-Cwirko M, et al. Amoxicillin for acute lower-respiratory-tract infection in primary care when pneumonia is not suspected: a 12-country, randomised, placebo-controlled trial. *Lancet Infect Dis*. 2013;13(2):123–9.
3. Moore M, Stuart B, Coenen S, Butler CC, Goossens H, Verheij TJM, et al. Amoxicillin for acute lower respiratory tract infection in primary care: subgroup analysis of potential high-risk groups. *Br J Gen Pract*. 2014;64(619):75–80.
4. Little P, Stuart B, Smith S, Thompson MJ, Knox K, Van Den Bruel A, et al. Antibiotic prescription strategies and adverse outcome for uncomplicated lower respiratory tract infections: Prospective cough complication cohort (3C) study. *BMJ (Online)*. 2017;357:j2148.
5. Welte T, Torres A, Nathwani D. Clinical and economic burden of community-acquired pneumonia among adults in Europe. *Thorax*. 2012;67(1):71–9.
6. Teepe J, Broekhuizen BDL, Loens K, Lammens C, Ieven M, Goossens H, et al. Predicting the presence of bacterial pathogens in the airways of primary care patients with acute cough. *CMAJ*. 2017;189(2):E50–5.
7. Minnaard MC, De Groot JAH, Hopstaken RM, Schierenberg A, De Wit NJ, Reitsma JB, et al. The added value of C-reactive protein measurement in diagnosing pneumonia in primary care: a meta-analysis of individual patient data. *CMAJ*. 2017;189(2):E56–63.
8. Bont J, Hak E, Hoes AW, Schipper M, Schellevis FG, Verheij TJM. A prediction rule for elderly primary-care patients with lower respiratory tract infections. *Eur Respir J*. 2007;29(5):969–75.
9. Van Doorn S, Tavenier A, Rutten FH, Hoes AW, Moons KGM, Geersing GJ. Risk of cardiac and non-cardiac adverse events in community-dwelling older patients with atrial fibrillation: a prospective cohort study in the Netherlands. *BMJ Open*. 2018;8(8):1–7.
10. van Royen FS, Joosten LPT, van Smeden M, Slottje P, Rutten FH, Geersing GJ, et al. Cardiovascular vulnerability predicts hospitalisation in primary care clinically suspected and confirmed COVID-19 patients: a model development and validation study. *PLoS One*. 2022;17(4):e0266750. <https://doi.org/10.1371/journal.pone.0266750>.
11. Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of myocardial infarction and stroke after acute infection or vaccination. *N Engl J Med*. 2004;351(25):2611–8.
12. Violi F, Cangemi R, Calviere C. Pneumonia, thrombosis and vascular disease. *J Thromb Haemost*. 2014;12(9):1391–400.
13. Stark K, Massberg S. Interplay between inflammation and thrombosis in cardiovascular pathology. *Nat Rev Cardiol*. 2021;18:666–82.
14. Macias AE, McElhaney JE, Chaves SS, Nealon J, Nunes MC, Samson SI, et al. The disease burden of influenza beyond respiratory illness. *Vaccine*. 2021;15(39):A6–14.
15. Xie Y, Xu E, Bowe B, Al-Aly Z. Long-term cardiovascular outcomes of COVID-19. *Nat Med*. 2022;28(3):583–90.
16. Kwong JC, Schwartz KL, Campitelli MA, Chung H, Crowcroft NS, Karnau-chow T, et al. Acute myocardial infarction after laboratory-confirmed influenza infection. *N Engl J Med*. 2018;378(4):345–53.
17. Fine M, Auble T, Yealy D, Hanusa B, Weissfeld L, Singer D, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med*. 1997;336(4):243–50.
18. Lim WS, Van Der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, et al. Defining community acquired pneumonia severity on presentation to hospital: AN international derivation and validation study. *Thorax*. 2003;58(5):377–82.
19. Bont J, Hak E, Hoes AW, Macfarlane JT, Verheij TJ. Predicting death in elderly patients with community-acquired pneumonia: a prospective validation study reevaluating the CRP-65 severity assessment tool. *Arch Intern Med*. 2008;168(13):1465–8.
20. Ochoa-Gondar O, Vila-Corcoles A, Rodriguez-Blanco T, Ramos F, De Diego C, Salsench E, et al. Comparison of three predictive rules for assessing severity in elderly patients with CAP. *Int J Clin Pract*. 2011;65(11):1165–72.
21. Francis NA, Cals JW, Butler CC, Hood K, Verheij T, Little P, et al. Severity assessment for lower respiratory tract infections: potential

- use and validity of the CRB-65 in primary care. *Prim Care Respir J*. 2012;21(1):65–70.
22. Bruyndonckx R, Hens N, Verheij TJM, Aerts M, Ieven M, Butler CC, et al. Development of a prediction tool for patients presenting with acute cough in primary care: a prognostic study spanning six European countries. *Br J Gen Pract*. 2018;68(670):e342–50.
  23. Smeets H, Kortekaas M, Rutten F, Bots M, van der Kraan W, Daggelders G, et al. Routine primary care data for scientific research, quality of care programs and educational purposes: the Julius General Practitioners' Network (JGPN). *BMC Health Serv Res*. 2018;18(1):735.
  24. World Organization of National Colleges, Academies and AA of GPP. *ICPC-2: International Classification of Primary Care*. Oxford: Oxford University Press; 1998.
  25. WHO Collaborating Centre for Drug Statistics Methodology. *ATC classification index with DDDs*. Oslo, Norway: Norwegian Institute of Public Health; 2022.
  26. Dutch Hospital Data (DHD). Landelijke Basisregistratie Ziekenhuiszorg (LBZ) - years: 2016–2021. Available from: <https://www.dhd.nl/producten-diensten/lbz/paginas/dataverzameling-lbz.aspx>.
  27. Amsterdam UMC. Academisch Netwerk Huisartsgeneeskunde Amsterdam (ANHA). [Cited 31 Mar 2023]. Available from: <https://www.vumc.nl/research/overzicht/academisch-netwerk-huisartsgeneeskunde-amsterdam-umc-anha.htm>.
  28. World Health Organization (WHO). *ICD-10: international statistical classification of diseases and related health problems: tenth revision, 2nd ed*. Geneva: World Health Organization; 2004.
  29. Ensor J, Martin EC, Riley RD. R package version 1.1.2. 2022 [Cited 16 Feb 2023]. *pmsampsize*: Calculates the Minimum Sample Size Required for Developing a Multivariable Prediction Model. Available from: <https://CRAN.R-project.org/package=pmsampsize>.
  30. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna; 2022. Available from: <https://www.R-project.org/>.
  31. Hak E, Bont J, Hoes AW, Verheij TJM. Prognostic factors for serious morbidity and mortality from community-acquired lower respiratory tract infections among the elderly in primary care. *Fam Pract*. 2005;22(4):375–80.
  32. Tralhão A, Póvoa P. Cardiovascular events after community-acquired pneumonia: a global perspective with systematic review and meta-analysis of observational studies. *J Clin Med*. 2020;9:414 MDPI.
  33. van der Linden M, Westert G, de Bakker D, Schellevis F. Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk. Utrecht/Bilthoven: NIVEL/RIVM: Klachten en aandoeningen in de bevolking en in de huisartspraktijk; 2004.
  34. Sisk R, Sperrin M, Peek N, van Smeden M, Martin GP. Imputation and missing indicators for handling missing data in the development and deployment of clinical prediction models: a simulation study. *Stat Methods Med Res*. 2023;096228022311650. Available from: <https://doi.org/10.1177/09622802231165001>.
  35. Van Calster B, Steyerberg EW, Wynants L, van Smeden M. There is no such thing as a validated prediction model. *BMC Med*. 2023;21(1):70.
  36. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Ann Intern Med*. 2015;162(1):55–63.
  37. Ochoa-Gondar O, Vila-Corcoles A, Rodriguez-Blanco T, Hospital I, Salsench E, Ansa X, et al. Validation of the CORB75 (confusion, oxygen saturation, respiratory rate, blood pressure, and age  $\geq 75$  years) as a simpler pneumonia severity rule. *Infection*. 2013;42(2):371–8.
  38. Moore M, Stuart B, Lown M, Van Den Bruel A, Smith S, Knox K, et al. Predictors of adverse outcomes in uncomplicated lower respiratory tract infections. *Ann Fam Med*. 2019;17(3):231–8.
  39. Pieralli F, Vannucchi V, Nozzoli C, Augello G, Dentali F, De Marzi G, et al. Acute cardiovascular events in patients with community acquired pneumonia: results from the observational prospective FADOI-ICECAP study. *BMC Infect Dis*. 2021;21(1):116.
  40. Raisi-Estabragh Z, Cooper J, Salih A, Raman B, Lee AM, Neubauer S, et al. Cardiovascular disease and mortality sequelae of COVID-19 in the UK Biobank. *Heart*. 2022;109(2):119–26.
  41. Singanayagam A, Singanayagam A, Elder DHJ, Chalmers JD. Is community-acquired pneumonia an independent risk factor for cardiovascular disease? *Eur Respir J*. 2012;39:187–96.
  42. Davidson JA, Banerjee A, Smeeth L, McDonald HI, Grint D, Herrett E, et al. Risk of acute respiratory infection and acute cardiovascular events following acute respiratory infection among adults with increased cardiovascular risk in England between 2008 and 2018: a retrospective, population-based cohort study. *Lancet Digit Health*. 2021;3(12):e773–83.
  43. Chow EJ, Uyeki TM, Chu HY. The effects of the COVID-19 pandemic on community respiratory virus activity. *Nat Rev Microbiol*. 2023;21:195–210.
  44. Wells BJ, Nowacki AS, Chagin K, Kattan MW. Strategies for handling missing data in electronic health record derived data. *eGEMs (Generating Evidence & Methods to improve patient outcomes)*. 2013;1(3):7.
  45. Cartus AR, Samuels EA, Cerdá M, Marshall BDL. Outcome class imbalance and rare events: an underappreciated complication for overdose risk prediction modeling. *Addiction*. 2023;118(6):1167–76.
  46. Metlay JP, Waterer GW, Long AC, Anzueto A, Brozek J, Crothers K, et al. Diagnosis and treatment of adults with community-acquired pneumonia. *Am J Respir Crit Care Med*. 2019;200(7):E45–67.

## Publisher's Note

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

### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

