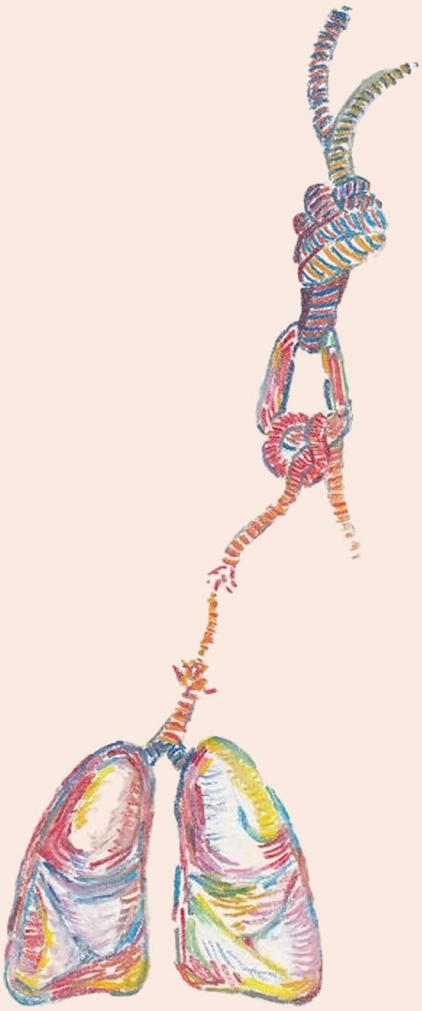
Improving the prognostication of lower respiratory tract infections in general practice

Merijn Huub Rijk



Improving the prognostication of lower respiratory tract infections in general practice

Merijn Huub Rijk

ISBN 978-94-6534-395-2

DOI 10.33540/3642.

Cover design: André Rijk (2026)

Printing: ProefschriftMaken || www.proefschriftmaken.nl

Financial support by SBOH, employer of GP trainees, for the publication of this thesis is gratefully acknowledged.

Improving the prognostication of lower respiratory tract infections in general practice

Verbeterde inschatting van de prognose van lage luchtweginfecties in de huisartspraktijk

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. ir. W. Hazeleger,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op

dinsdag 15 september 2026 des middags te 2.15 uur

door

Merijn Huub Rijk

geboren op 25 september 1993
te Breda

Promotoren:
Prof. dr. R.P. Venekamp
Prof. dr. F.H. Rutten

Copromotor:
Dr. T.N. Platteel

Beoordelingscommissie:
Prof. dr. J. Bont
Prof. dr. R.A.M.J. Damoiseaux
Prof. dr. M.H. Emmelot-Vonk (voorzitter)
Prof. dr. ir. H.M. den Ruijter
Prof. dr. E.W. Steyerberg

TABLE OF CONTENTS

1 Introduction

INTRODUCTION

Lower respiratory tract infections in general practice

Respiratory tract infection symptoms are the most common reason for patients to consult their general practitioner (GP).[1] These infections can be divided into those arising from the upper respiratory tract – including 'common cold', acute rhinosinusitis, and pharyngotonsillitis – and those arising from the lower respiratory tract, including acute bronchitis and pneumonia.

Around three percent of the Dutch population is diagnosed with a lower respiratory tract infection (LRTI) annually.[1] Signs and symptoms suggestive of LRTI include acute cough, fever, shortness of breath, tachypnoea, and (unilateral) abnormalities upon lung auscultation such as crackles.[2] Upper respiratory tract infection symptoms (e.g. rhinorrhoea, sore throat, sneezing) are often present, and LRTI patients may also experience systemic symptoms such as malaise, fatigue, muscle aches, and chills.

Community-acquired LRTI is a clinical diagnosis and pathogen testing is not routinely recommended in general practice. Nevertheless, studies that performed systematic pathogen testing using serum, sputum, and nasopharyngeal swab samples have shown that most LRTI are of viral origin, with a bacterium present in around 20% of patients.[3]

Community-acquired LRTIs usually run a favourable course. Antibiotic treatment in patients with uncomplicated LRTI, i.e. those presenting with acute cough as main symptom judged to be infectious and without complicating conditions such as pulmonary comorbidity or a compromised immune system, and nearly always viral, does not provide benefit in terms of symptom duration or severity, and is therefore not routinely recommended.[2,4–6] Nevertheless, adverse outcomes – such as hospitalisation or even mortality – do occur in around one percent of adult patients presenting to general practice with (initially) uncomplicated LRTI up to nine percent in older patients (i.e., ≥65 years) with both complicated and uncomplicated LRTI.[7,8] Adverse outcome rates are substantially higher in patients diagnosed with community-acquired pneumonia (CAP) by their GP, with reported hospitalisation rates up to 50% in Germany and the United Kingdom.[9–11] Therefore, the guideline on 'acute cough' issued by the Dutch College of General Practitioners (NHG) recommends GPs to try to substantiate patients with acute cough, in whom close monitoring and/or antibiotics might be considered based on additional 'risk factors' for a complicated disease trajectory, from those with suspected CAP, in whom antibiotic treatment is recommended.[2] A point-of-care (PoC) C-reactive protein (CRP) test can be used to rule out CAP in intermediate risk patients according to the GP's judgement: low CRP levels (<20 mg/L) are likely in patients with viral LRTI whereas high levels (>100 mg/L) support the diagnosis of CAP. However, distinguishing CAP or presence of a bacterial pathogen from viral LRTI based on signs, symptoms, and a

PoC CRP test remains difficult in general practice.[12,13] This leaves GPs with uncertainty on how to best manage patients with LRTI in everyday practice.

Prognostication of patients presenting to general practice with a community-acquired lower respiratory tract infection

The 'acute cough' guideline issued by the NHG lists the following patient characteristics as prognostic factors for a complicated trajectory in adult patients: higher age (arbitrary >75 years); a history of severe cardiac disease, pulmonary disease, diabetes, or neurological disorders; severe liver or kidney impairment; and a severely impaired immune system.[2] It is emphasized that this list of prognostic factors is based on consensus due to a lack of evidence on prognostic factors and prediction models on this subject. Equally important, the list of individual factors predictive of a complicated trajectory mentioned in the guideline does not translate into absolute individual risks, hampering unequivocal clinical interpretation and communication with patients. Thus, irrespective of the high burden of LRTI in the population, GPs are left without evidence-based guidance on how to identify patients at highest risk of adverse outcomes to support their clinical assessment and the decision-making process.

Prediction models incorporating a predefined set of predictors are tools that can assist clinicians by estimating individual risk of an outcome. Existing prediction models for adverse outcomes in LRTI patients that have been adopted in routine hospital care are – most notably – the Pneumonia Severity Index (PSI) and CURB-65 (confusion, blood urea nitrogen, respiratory rate, blood pressure, and age ≥65 years).[14,15] Both prediction models have been developed to predict individual risk of 30-day mortality in hospitalised CAP patients which limits applicability to community LRTI patients in whom both CAP and mortality are rare.[16] The lower a priori risk of the outcome in community patients would likely affect the predictive performance of a prediction model.[17] In addition, both PSI and CURB-65 include radiographic or laboratory assessments as predictors that are not readily or immediately (i.e. blood nitrogen urea) available to GPs (Table 1).

Table 1. Predictors of PSI and CURB-65

PSI		CURB-65	
Predictor	**Points**	**Predictor**	**Points**
Age			
Men	Age (yr)	Confusion	1
Women	Age (yr) - 10		
Nursing home resident	+10	Blood urea nitrogen >7 mmol/litre	1

Table 1. Continued

PSI		CURB-65	
Coexisting illness		Respiratory rate ≥30/min	1
Neoplastic disease	+30		
Liver disease	+20		
Congestive heart failure	+10	SBP <90 or DBP ≤60 mmHg	1
Cerebrovascular disease	+10		
Renal disease	+10		
Physical examination		Age ≥65 years	1
Altered mental status	+20		
Respiratory rate ≥30/min	+20	**Total**	**Mortality rate (%)**
SBP <90 mmHg	+20	0-1 points:	1.5
Temperature <35 or ≥40°C	+15	2 points:	9.2
Pulse ≥125/min	+10	≥3 points:	22.0
Laboratory and radiographic			
Arterial pH <7.35	+30		
Blood urea nitrogen ≥11 mmol/litre	+20		
Sodium <130 mmol/litre	+20		
Glucose ≥14 mmol/litre	+10		
Haematocrit <30%	+10		
Partial pressure of arterial Oxygen <60 mmHg	+10		
Pleural effusion	+10		

Total	**Mortality rate (%)**
0 points:	0.1
≤70 points:	0.6
71-90 points:	0.9
91-130 points:	9.3
>130 points:	27.0

Abbreviations: PSI, pneumonia severity index; CURB-65, confusion, blood urea nitrogen, respiratory rate, blood pressure, and age ≥65 years; SBP, systolic blood pressure; DBP, diastolic blood pressure.

Mortality rates based on model validation cohorts as presented in the primary model development and validation study.[14,15]

In their publication on the development of the CURB-65, the authors also present an alternative for use in the community setting: the CRB-65, consisting of the CURB-65 predictors excluding blood urea nitrogen.[14] This CRB-65 scoring rule – with predictors readily available also in general practice – is incorporated in the guideline on pneumonia in adults of the National Institute for Health and Care Excellence (NICE) from the United Kingdom.[18] Following this guideline, GPs are encouraged to use CRB-65 in addition to clinical judgement, and it is recommended to consider hospital admission in CAP patients with a score of 2 or more. This recommendation, however, is based on the 30-day mortality risk of hospitalised CAP patients observed in those with a score of 2 or more – almost 10% – whereas it is likely that this risk is substantially lower in CAP patients presenting

to their GP. When applying a prediction model in a population that differs from that of model development in, for example, outcome prevalence, it should be validated in the population of intended use.[17] Moreover, from a community patient and GP perspective it would be relevant to include hospitalisation as an endpoint.

The attempts of current practice to guide GPs in estimating the prognosis of LRTI patients and the lack of consensus across guidelines imply a need for a thorough synthesis of the available prognostic literature on this topic. This may identify prognostic factors or existing prediction models or provide guidance for derivation of a novel prediction model.

The role of concurrent cardiovascular disease in the prognosis of patients with a lower respiratory tract infection

When predicting future health outcomes, age and sex are usually strong predictors. This also applies to the prognosis of community patients with LRTI, with age and sex reported as prognostic factors for short-term hospitalisation or mortality in several studies.[7,19,20] Apart from these 'usual suspects', concurrent cardiovascular diseases (CVD) have also been linked to poor prognosis of community-acquired LRTI. For example, heart failure, prior stroke, and diabetes have been linked to increased risk of hospitalisation and mortality within a month from LRTI diagnosis among community patients.[19,21,22] Evidence obtained during the coronavirus disease 2019 (COVID-19) pandemic showed that CVD negatively impacted the prognosis of SARS-CoV-2 infected patients presenting to general practice. Apart from age and sex, a wide range of CVD – including heart failure, coronary artery disease, peripheral arterial disease, cerebrovascular disease, venous thromboembolism, atrial fibrillation, and diabetes – were predictive for hospitalisation within 90 days.[23] Despite consistent evidence on the prognostic value of concurrent CVD in community patients with LRTI, the incremental value when added to other strong predictors such as age and sex – indicating its incremental predictive value – remains to be quantified.

Lower respiratory tract infections as a trigger for cardiovascular diseases

The interplay between CVD and LRTI also works in the opposite direction: LRTI have been associated with an increased incidence of acute cardiovascular events. An association between seasonal influenza and cardiovascular mortality was reported as early as 1932.[24] The highest risk of acute cardiovascular events seems to be shortly following the initial infection. Within 14 days following LRTI in community patients, incidence rates of arterial events (acute myocardial infarction and stroke) increased fourfold whereas incidence rates of venous thromboembolism (deep venous thrombosis and pulmonary embolism) were found to have almost doubled.[25,26] Specific respiratory pathogens, such as influenza virus and SARS-CoV-2, were particularly studied for cardiovascular sequelae.[27–30] A proposed mechanism underlying the increase in cardiovascular events following LRTI is immunothrombosis – i.e. activation of the immune system by acute infection, trigger-

ing an interaction between inflammatory and prothrombotic pathways.[31,32] In addition, LRTI patients can experience systemic changes such as increased heart rate and blood pressure instability, which result in increased cardiac demand and therefore pose a 'stress test' for the cardiovascular system.[33] Although evidence is accumulating that acute cardiovascular events can be triggered by LRTI, the absolute excess burden of cardiovascular events associated with LRTI – and therefore the population impact – remains unknown since previous studies were unable to translate relative increases in incidence rates to absolute excess risks. In addition, from a clinical perspective, it would be valuable to gain more insight into specific patients or disease characteristics contributing to highest cardiovascular risk following LRTI, since such approach holds potential to identify patients that might benefit most from novel interventions to prevent cardiovascular events following a LRTI episode.[33] These might, for example, include close monitoring strategies for symptoms of unfolding CVD, early detection and management of 'provoked' atrial fibrillation, or treatment aimed at prevention of thrombo-embolic events.

Utilising the full potential of electronic health records data in prognostic research

Prognostic research usually requires a relatively large number of study participants, especially when outcome events – such as hospitalisation, mortality, and cardiovascular events following LRTI in general practice – are rare. Electronic health records (EHR) hold routine health care data on large numbers of unselected individual patients and is increasingly recognised as a valuable source of large longitudinal cohorts that can be used for such purposes.[34] For example, EHR data have been used to develop well-known prediction models such as the QRisk scores, and to calculate individual scores of existing prediction models such as the Framingham Heart score, both predicting 10-year risk of CVD in the general population.[35,36]

EHR contain structured data on domains including patient demographics, medical history, medication use, consultations, diagnoses, and treatments. The use of standardised coding systems such as the International Classification of Primary Care (ICPC)[37] or International Classification of Diseases (ICD)[38] for medical history and diagnoses and the Anatomical Therapeutic Chemical (ATC)[39] codes for medication prescriptions allow for relatively straightforward extraction of structured data. In addition to structured data, EHR also contain unstructured data captured in clinical notes, with information on domains including presenting signs and symptoms, physical examination, and measurements. The potential of such information from unstructured clinical notes, for example as candidate predictors in prediction model development studies, is evident.[40] However, manual extraction of information from clinical notes on large amounts of patients is both time- and resource-consuming, hence the focus on structured data in EHR-based research to date.[41]

Recent advances in the field of natural language processing (NLP) increasingly allow for automated retrieval of information from unstructured EHR data.[42] Especially the development of large language models (LLM), an NLP technique which can process text while taking context into account, is promising.[43] LLMs have already been used to extract data from clinical notes on reported symptoms and chronic diseases.[44,45]

Since EHR data are not primarily collected for research purposes, its use in (prognostic) research also poses challenges. For example, working with structured, code-based EHR data could introduce information bias through misclassification of the study population, candidate predictors, or outcome measures when codes are not uniformly used or even incorrectly registered.[46] Moreover, it is likely that a proportion of the information on extracted variables will not be recorded, and values will therefore be missing. Missing data on candidate predictors should be properly handled based on the (assumed) underlying mechanisms and intended use of the prediction model.[47] The issue of missing data and determination of the underlying mechanisms is probably even more challenging when using candidate predictor information from free text clinical notes. Clinicians are likely to selectively record positive or abnormal findings, which may result in the assumption that missing values can be regarded as negative or normal. On the other hand, when a certain sign or symptom is not recorded, it can be debated whether it is valid to assume the sign or symptom was absent.

Despite the obvious challenges of working with unstructured EHR-data from clinical notes, its potential for prognostic research cannot be ignored. However, before incorporating candidate predictors derived from unstructured clinical notes in a prediction model, it is instrumental to gain more insight into the quality and completeness of clinical notes and the performance of LLMs for the automatic extraction of candidate predictor information from such data. This can support decisions regarding specific LLMs to be used and how to handle non-recorded values. Ultimately, this will allow for the assessment of the incremental predictive value of candidate predictors derived from clinical notes when added to a prediction model based on structured data.

Aims of this thesis

In this thesis, we aim to improve prognostication of adults presenting to their GP with LRTI while considering the role of CVD as well as exploring and utilising the full potential of unstructured EHR-data in this context. To this end, large cohorts of LRTI patients presenting to general practice were derived from routine EHR databases, and the opportunities, challenges, and incremental value of including unstructured data on presenting signs, symptoms, and measurements in prediction research was explored.

The work presented in this thesis was inspired by the following clinical and methodological questions:

In patients presenting to their GP with LRTI,

1. what are prognostic factors for 90-day all-cause hospitalisation or mortality?

2. what prediction models have been developed to estimate individual risk of 90-day all-cause hospitalisation or mortality, and how well do these perform?

3. in the absence of an existing prediction model, which combination of predictors can most accurately estimate individual risk of 30-day hospitalisation or mortality?

4. what is the incremental predictive value of a history of CVD for 30-day all-cause hospitalisation or mortality?

In prognostic research,

5. what is the quality of unstructured EHR-data from clinical notes in terms of completeness, and what are opportunities and challenges of including candidate predictors derived from clinical notes?

6. what is the performance of LLMs when used for automatic extraction of candidate predictor information from clinical notes in Dutch EHR?

Based on findings of the clinical and methodological questions listed above, in patients presenting to their GP with LRTI,

7. what is the incremental predictive value of presenting signs and symptoms automatically extracted from unstructured EHR-data by LLMs when predicting 30-day all-cause hospitalisation or mortality?

8. what is the 90-day incidence of acute cardiovascular events following LRTI, and which subset of patients are at highest risk of such events?

9. what is the absolute excess burden of acute cardiovascular events attributable to LRTI?

Thesis outline

Based on our aims, this thesis consists of the following chapters:

In **Chapter 2**, we report the results of a systematic review of available literature on prognostic factors and existing prediction models for all-cause hospitalisation and mortality in patients presenting to general practice with LRTI, with the aim to identify established prediction models – that need external validation or updating – and prognostic factors that need to be considered when developing a new prediction model.

In **Chapter 3**, we present the rationale and design for the development of two models that predict individual risk of (i) 30-day all-cause hospitalisation or mortality and (ii) acute cardiovascular events in patients presenting to general practice with LRTI.

In **Chapter 4**, we explore the quality and completeness of unstructured EHR data from clinical notes on presenting signs, symptoms, and measurements of patients presenting to general practice with LRTI and explore challenges of working with unstructured EHR data in the context of prognostic research.

In **Chapter 5**, we present the development, internal and external validation of an EHR-based model that predicts individual risk of 30-day all-cause hospitalisation or mortality in patients presenting to general practice with LRTI, while using a stepwise model development approach to assess the incremental predictive value of a history of CVD.

In **Chapter 6**, we establish and compare the performance of several LLMs fine-tuned for the automatic extraction of candidate predictor information on presenting signs and symptoms from clinical notes in Dutch EHR of patients presenting to general practice with LRTI.

In **Chapter 7**, we assess the incremental predictive value of presenting signs and symptoms extracted from unstructured EHR-data by LLMs for predicting individual risk of 30-day all-cause hospitalisation or mortality in patients presenting to general practice with LRTI.

In **Chapter 8**, we map the short-term incidence of acute cardiovascular events following LRTI in patients presenting to general practice with LRTI to identify subgroups of patients with highest incidence of these events following LRTI. Subsequently, we present the results of a population-based nested self-controlled case-series assessing the association between LRTI and 90-day incidence of acute cardiovascular events, which is ultimately used to determine the excess burden of these events attributable to LRTI in general practice.

In the **general discussion (Chapter 9)**, I will reflect on the overarching implications of this thesis in the context of the consultation room, discuss opportunities and challenges of implementing prediction models in clinical practice, and consider potential directions for interventions aimed at mitigating LRTI-related cardiovascular risk in light of emerging societal issues including 'diagnosis-expansion', continuously increasing healthcare expenditures, and limited healthcare resources.

REFERENCES

1. Linden M van der, Westert G, Bakker D de, Schellevis F. *Tweede nationale studie naar ziekten en verrichtingen in de huisartspraktijk: klachten en aandoeningen in de bevolking en in de huisartspraktijk*. Utrecht: NIVEL, 2004.
2. Verheij T, Hopstaken R, Prins J, et al. *NHG-Standaard Acuut hoesten* [internet]. Utrecht: Nederlands Huisartsen Genootschap; 2020. Accessed May 6, 2025. Available at: https://richtlijnen.nhg.org.
3. Ieven M, Coenen S, Loens K, et al. Aetiology of lower respiratory tract infection in adults in primary care: a prospective study in 11 European countries. *Clin Microbiol Infect*. 2018;**24**(11):1158–1163.
4. Little P, Stuart B, Smith S, et al. Antibiotic prescription strategies and adverse outcome for uncomplicated lower respiratory tract infections: prospective cough complication cohort (3C) study. *BMJ*. 2017;**357**:j2148.
5. National Institute for Health and Care Excellence. *Suspected acute respiratory infection in over 16s: assessment at first presentation and initial management (NG237)* [internet]. London: National Institute for Health and Care Excellence; 2023. Accessed May 6, 2025. Available at: https://www.nice.org.uk/guidance/ng237.
6. Smith SM, Fahey T, Smucny J, Becker LA. Antibiotics for acute bronchitis. *Cochrane Database Syst Rev*; 2017;**6**(6):CD000245.
7. Moore M, Stuart B, Lown M, et al. Predictors of adverse outcomes in uncomplicated lower respiratory tract infections. *Ann Fam Med*. 2019;**17**(3):231–238.
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):969–975.
9. Theilacker C, Sprenger R, Leverkus F, et al. Population-based incidence and mortality of community-acquired pneumonia in Germany. *PLoS One*. 2021;**16**(6):e0253118.
10. Kolditz M, Tesch F, Mocke L, Höffken G, Ewig S, Schmitt J. Burden and risk factors of ambulatory or hospitalized CAP: a population based cohort study. *Respir Med*. 2016;**121**:32–38.
11. Millett ERC, Quint JK, Smeeth L, Daniel RM, Thomas SL. Incidence of community-acquired lower respiratory tract infections and pneumonia among older adults in the United Kingdom: a population-based study. *PLoS One*. 2013;**8**(9).
12. Teepe J, Broekhuizen BDL, Loens K, et al. Predicting the presence of bacterial pathogens in the airways of primary care patients with acute cough. *CMAJ*. 2017;**189**(2):E50–E55.
13. Minnaard MC, De Groot JAH, Hopstaken RM, 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–E63.
14. Lim WS, Van Der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax*. 2003;**58**(5):377–382.
15. Fine M, Auble T, Yealy D, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J of Med*. 1997;**336**(4):243–250.
16. Francis NA, Cals JW, Butler CC, 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.
17. Damen JAAG, Debray TPA, Pajouheshnia R, et al. Empirical evidence of the impact of study characteristics on the performance of prediction models: a meta-epidemiological study. *BMJ Open*. 2019;**9**(4).
18. National Institute for Health and Care Excellence. *Pneumonia in adults: diagnosis and management (CG191)* [internet]. London: National Institute for Health and Care Excellence; 2014. Accessed May 7, 2025. Available at: https://www.nice.org.uk/guidance/cg191.

19. 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):375–380.
20. Winchester CC, Macfarlane T V., Thomas M, Price D. Antibiotic prescribing and outcomes of lower respiratory tract infection in UK primary care. *Chest*. 2009;**135**(5):1163–1172.
21. Millett ERC, De Stavola BL, Quint JK, Smeeth L, Thomas SL. Risk factors for hospital admission in the 28 days following a community-acquired pneumonia diagnosis in older adults, and their contribution to increasing hospitalisation rates over time: a cohort study. *BMJ Open*. 2015;**5**(12).
22. van de Nadort C, Smeets HM, Bont J, Zuithoff NPA, Hak E, Verheij TJM. Prognosis of primary care patients aged 80 years and older with lower respiratory tract infection. *Br J Gen Pract*. 2009;**59**(561):261–265.
23. van Royen FS, Joosten LPT, van Smeden M, 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.
24. Collins SD. Excess mortality from causes other than influenza and pneumonia during influenza epidemics. *Public Health Reports*. 1932;**47**(46):2159–2179.
25. 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):2611–2618.
26. Liam Smeeth, Claire Cook, Sara Thomas, et al. Risk of deep vein thrombosis and pulmonary embolism after acute infection in a community setting. *Lancet*. 2006;**367**(9516):1075–1079.
27. de Boer AR, Riezebos-Brilman A, van Hout D, et al. Influenza infection and acute Myocardial Infarction. *NEJM Evid*. 2024;**3**(7).
28. Macias AE, McElhaney JE, Chaves SS, et al. The disease burden of influenza beyond respiratory illness. *Vaccine*. 2021;**39**:A6–A14.
29. Xie Y, Xu E, Bowe B, Al-Aly Z. Long-term cardiovascular outcomes of COVID-19. *Nat Med*. 2022;**28**(3):583–590.
30. la Roi-Teeuw HM, van Smeden M, Geersing GJ, et al. Incidence and individual risk prediction of post-COVID-19 cardiovascular disease in the general population: a multivariable prediction model development and validation study. *Eur Heart J Open*. 2023;**3**(6).
31. Stark K, Massberg S. Interplay between inflammation and thrombosis in cardiovascular pathology. *Nat Rev Cardiol*. 2021;**18**(9):666–682.
32. Violi F, Cangemi R, Calvieri C. Pneumonia, thrombosis and vascular disease. *J Thromb and Haemost*. 2014;**12**(9):1391–1400.
33. Lee JJ, Koshiaris C, Hobbs FDR, Sheppard JP. Beyond COVID-19: respiratory infection and cardiovascular events. *Br J Gen Pract*. 2021;**71**(709):342–343.
34. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using electronic health records for population health research: a review of methods and applications. *Annu Rev Public Health*. 2016;**37**:61–81.
35. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ*. 2017;**357**:j2099.
36. Green BB, Anderson ML, Cook AJ, et al. Using body mass index data in the electronic health record to calculate cardiovascular risk. *Am J Prev Med*. 2012;**42**(4):342–347.
37. World Organization of Family Doctors (WONCA). I*CPC-2: International Classification of Primary Care*. Oxford: Oxford University Press; 1998.
38. World Health Organization (WHO). *ICD-10: International Statistical Classification of Diseases and Related Health Problems: tenth revision, 2nd ed*. Geneva: World Health Organization; 2004.

39. World Health Organization (WHO) Collaborating Centre for Drug Statistics Methodology. *ATC Classification Index with DDDs*. Oslo: WHO Collaborating Centre for Drug Statistics Methodology; 2022.
40. Seinen TM, Fridgeirsson EA, Ioannou S, et al. Use of unstructured text in prognostic clinical prediction models: a systematic review. *J Am Med Inform Assoc*. 2022;**29**(7):1292–1302.
41. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012;**13**(6):395–405.
42. Fu S, Wang L, Moon S, et al. Recommended practices and ethical considerations for natural language processing-assisted observational research: a scoping review. *Clin Transl Sci*. 2023;**16**(3):398–411.
43. Muhammad Usman Hadi, Qasem Al Tashi, Rizwan Qureshi, et al. Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects. *TechRxiv* (preprint). February 10, 2025. Available from: https://www.techrxiv.org/doi/full/10.36227/techrxiv.23589741.
44. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *J Am Med Inform Assoc*. 2019;**26**(4):364–379.
45. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform*. 2019;**7**(2).
46. Bots SH, Groenwold RHH, Dekkers OM. Using electronic health record data for clinical research: a quick guide. *Eur J Endocrinol*. 2022;**186**(4):E1–E6.
47. Nijman SWJ, Leeuwenberg AM, Beekers I, et al. Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review. *J Clin Epidemiol*. 2022;**142**:218–229.

2

Prognostic factors and prediction models for hospitalisation and all-cause mortality in adults presenting to primary care with a lower respiratory tract infection: a systematic review

Merijn H. Rijk
Tamara N. Platteel
Teun M.C. van den Berg
Geert-Jan Geersing
Paul Little
Frans H. Rutten
Maarten van Smeden
Roderick P. Venekamp

BMJ Open. 2024;**14**(3):e075475.

Abstract

Objective To identify and synthesize relevant existing prognostic factors (PF) and prediction models (PM) for hospitalisation and all-cause mortality within 90 days in primary care patients with acute lower respiratory tract infections (LRTI).

Design Systematic review.

Methods Systematic searches of MEDLINE, Embase, and the Cochrane Library were performed. All PF and PM studies on risk of hospitalisation or all-cause mortality within 90 days in adult primary care LRTI patients were included. Risk of bias was assessed using the QUIPS and PROBAST tools for PF and PM studies, respectively. Results of included PF and PM studies were descriptively summarised.

Results Of 2,799 unique records identified, 16 were included: nine PF studies, six PM studies, and one combination of both. Risk of bias was judged high for all studies, mainly due to limitations in the analysis domain. Based on reported multivariable associations in PF studies, increasing age, sex, current smoking, diabetes, a history of stroke, cancer, or heart failure, previous hospitalisation, influenza vaccination (negative association), current use of systemic corticosteroids, recent antibiotic use, respiratory rate ≥25/minute, and diagnosis of pneumonia were identified as most promising candidate predictors. One newly developed PM was externally validated (c-statistic 0.74, 95% confidence interval (CI) 0.71 to 0.78) whereas the previously hospital-derived CRB-65 was externally validated in primary care in five studies (c-statistic ranging from 0.72 (95% CI 0.63 to 0.81) to 0.79 (95% CI 0.65 to 0.92)). None of the PM studies reported measures of model calibration.

Conclusions Implementation of existing models for individualised risk prediction of 90-day hospitalisation or mortality in primary care LRTI patients in everyday practice is hampered by incomplete assessment of model performance. The identified candidate predictors provide useful information for clinicians and warrant consideration when developing or updating PMs using state-of-the-art development and validation techniques.

INTRODUCTION

Acute lower respiratory tract infections (LRTI) are common in primary care.[1,2] Uncomplicated LRTI episodes generally have a favourable natural course in otherwise healthy adults, and antibiotic treatment with amoxicillin confers only little benefit in terms of earlier symptom resolution, both overall and in higher-risk subgroups of patients.[3–5] The clinical spectrum of LRTI patients, however, is heterogenous in terms of patient and disease specific characteristics, and thus the risk of hospitalisation and death varies substantially. For example, observational evidence suggests that adverse outcomes such as hospitalisation and death occur in fewer than 1% of patients with uncomplicated LRTI[5], whereas this risk is far more pronounced among those with community-acquired pneumonia (CAP).[6,7] However, diagnosing CAP in a primary care setting is notoriously challenging[8,9], leaving general practitioners (GP) with uncertainties how to identify higher-risk groups among those presenting with LRTI symptoms.

Nevertheless, identification of LRTI patients at highest risk of complications is pivotal as this might help GPs identify those in whom close follow-up or early (antibiotic) treatment is warranted. Currently, several scores exist to assist physicians in estimating risk of adverse outcomes in LRTI patients, such as the Pneumonia Severity Index (PSI) and CURB-65 (a prediction rule including confusion, blood urea nitrogen, respiratory rate, blood pressure, and age ≥65 years).[10,11] However, these prediction rules have been developed in secondary care and the inclusion of laboratory and/or radiographic features hampers applicability to outpatients. A modified version of the CURB-65 (CRB-65, not including blood urea nitrogen) has been proposed as primary care alternative, but its development in hospitalised patients raises uncertainty about its performance in outpatients.[10] An overview of existing primary care-specific prognostic factors (PF) and prediction models (PM) for adverse outcomes in LRTI patients could provide insight on (i) existing PFs or PMs suitable for use in clinical practice and (ii) relevant PFs to include when developing a new or updating an existing model. We therefore aimed to synthesize existing knowledge on PFs and PMs for hospitalisation and mortality within 90 days in adult patients presenting to primary care with LRTI.

METHODS

The protocol for this review was registered in PROSPERO (CRD42022341233) and the review was reported according to the PRISMA reporting guideline for systematic reviews.[12]

Searches and study selection

The electronic databases of MEDLINE, Embase, and the Cochrane Library were searched on 27 June 2022 for eligible studies. Keywords for 'respiratory tract infection', 'hospitalisa-

tion', 'death', 'adults' and 'primary care' were combined with a validated search string to identify relevant prognostic studies (Table S1).[13] Eligibility criteria were based on PICOTS guidance; we included all studies on adults with LRTI (patients) in which prognostic factors or prediction models (index) for 90-day hospitalisation or all-cause mortality (outcome) were assessed on the day of diagnosis (timing) in primary care (setting).[14,15] Non-English articles or those without full-text availability were excluded.

After de-duplication, two review authors (MHR, TMCB) independently reviewed titles/abstracts of unique records retrieved from the electronic databases against the eligibility criteria. Next, the same authors independently reviewed the full text of potentially relevant articles for eligibility. Any disagreements were resolved by discussion and, where necessary, by consulting a third reviewer (TNP). To increase the yield of potentially eligible articles, reference lists of the included articles were reviewed for eligibility. Studies assessing the prognostic value of PFs, developing a new PM, or externally validating a previously developed PM were eligible for inclusion.

Data collection

Data from eligible PF and PM studies were extracted by MHR using standardised forms based on the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of PF studies (CHARMS-PF) and PM studies (CHARMS), respectively.[14,16] Main outcomes of interest were hospitalisation or death within 90 days after initial LRTI consultation. For PF studies we extracted odds ratios, hazard ratios, or risk ratios with accompanying 95% confidence intervals (CI) from both univariable and multivariable analyses. For those studies that did not report such estimates, these associations were calculated from crude data using Rothman's Episheet version 15 November 2021 wherever possible. For PM studies, model performance estimates in terms of discrimination (c-statistic with accompanying 95% CI) and calibration (intercept and slope) were extracted. If calibration measures were absent, the total observed over expected (O/E) ratio with accompanying 95% CI of patients with the outcome was calculated wherever possible as a proxy of overall model calibration.[15] To this end, we used data on the distribution of subjects and outcomes across different PM risk strata and the corresponding predicted risks of the strata as reported in the original model development study. These calculations were performed in R version 4.2.2 using the *metamisc* package.

Other data extraction items included study population, candidate predictors, sample size, (handling of) missing data, analysis (modelling method and/or model development and validation method), and interpretation of results. Where needed, corresponding authors of individual studies were contacted to request additional data.

Assessment of risk of bias and applicability

Risk of bias assessment was performed by two authors (MHR, TNP) independently using the Quality in Prognostic Studies tool (QUIPS, assessing study participation, study attrition, PF measurement, outcome measurement, and statistical analysis) and the Prediction Model Risk Of Bias Assessment Tool (PROBAST, covering the domains of participants, predictors, outcome, and analysis) for PF and PM studies, respectively.[17,18] We excluded the 'study confounding' item from the QUIPS, since confounding is not applicable to prognostic research. Applicability of included articles was assessed based on the PICOTS criteria.

Synthesis methods

Due to substantial heterogeneity in domain, predictor, and outcome definition, combining PF effect estimates or PM performance measures in meta-analyses was considered inappropriate and results were therefore descriptively summarised, stratified according to diagnosis (i.e. LRTI versus pneumonia). Forest plots were created to visually summarise (i) univariable and multivariable effect estimates of PFs, and (ii) discrimination and calibration performance measures of PMs. PFs were categorised into either demographics, patient history, healthcare use, medication use, signs and symptoms, laboratory tests, or diagnosis. Promising candidate predictors were identified based on reported results of multivariable analysis in PF studies. For each promising PF and PM, we applied the grading of recommendations, assessment, development, and evaluations (GRADE) framework using prognostic research-specific guidance to rate the quality of evidence.[19,20]

Patient and public involvement

There was no patient or public involvement in this study.

RESULTS

Study selection

The search strategy yielded a total of 3,191 records. Removing duplicates left 2,799 unique records (Figure 1). After screening title and abstract, 34 studies remained for full-text review. Of these, 19 were excluded (Figure 1). One additional publication was retrieved from screening reference lists[21], leaving 16 publications for inclusion: nine PF studies[21–29], six PM studies[30–35], and one combination of PM and PF[36].

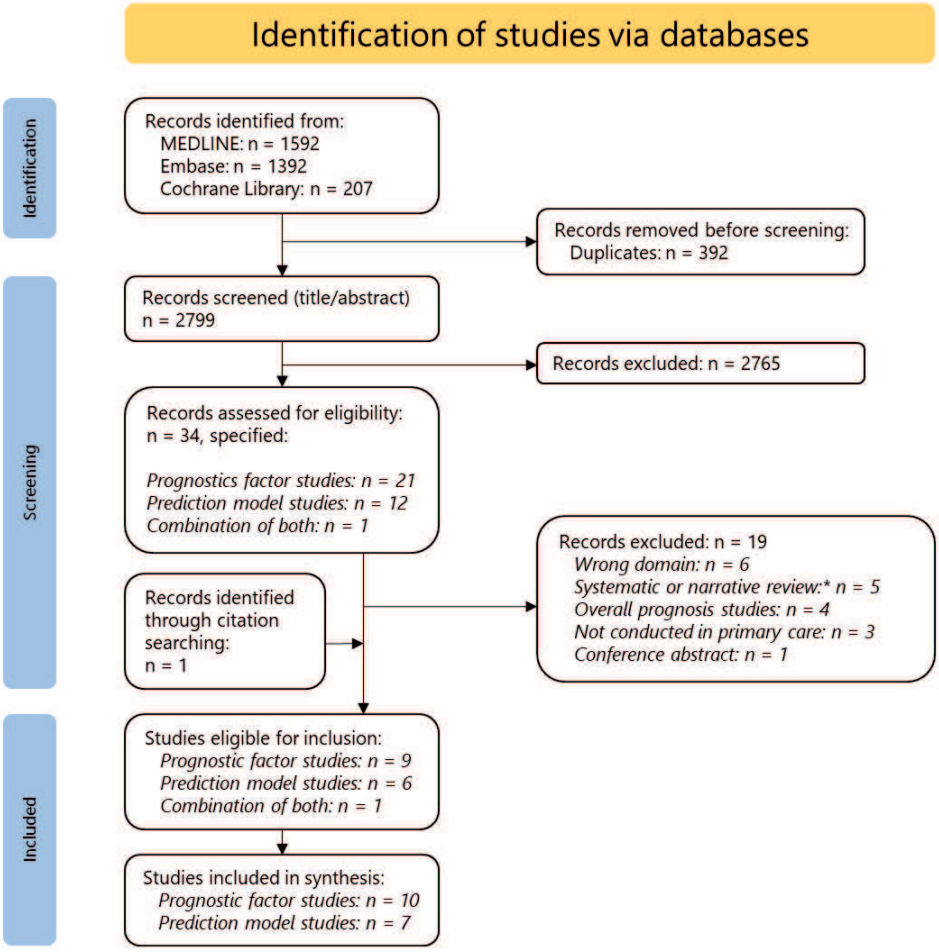


Figure 1. Flow of articles through study

* Systematic reviews have been screened for eligible primary studies.

Study characteristics

The PF studies were conducted between 1997-2021 of which most (7/10) used a retrospective cohort design (Table 1). The study population in terms of age (minimum age ranging from 1-80 years) and LRTI definition (ranging from suspected LRTI to diagnosis of CAP), as well as the hospitalisation rate (from 0.7% in LRTI patients to 76.5% CAP patients) varied substantially across studies. Mortality was used as outcome in four studies, hospitalisation in three studies, and a composite of both in one study. Two studies also included deterioration of existing illness and late-onset pneumonia in addition to hospitalisation and mortality in their outcome[23,36], while the timing of outcome assessment was unknown in one study[29].

Table 1. Study characteristics of included prognostic factor studies

Study	Source of data	Country	Setting	Study population	Participants, n	Included episodes	Age, mean (SD) or median (IQR)	Male sex, %	Number of candidate predictors	Outcome[a]	Hospitalisation, n (%)	Mortality, n (%)	Total outcome events, n (%)	EPP
Houston, 1997	Cohort, retrospective	US	Community-based	Elderly (≥65 years) LRTI patients	413	413	NR	38.0	19	30-day all-cause mortality	NR	44 (10.7)	44 (10.7)	2.3
Seppä, 2001	Cohort, prospective	FIN	Primary care	Elderly (≥65 years) patients with LRTI suggestive of pneumonia	950	950	Survival: 76 (IQR 71-81), death: 80 (IQR 73-86)	52	20	LRTI-related mortality within 30 days	NR	38 (4.0)	38 (4.0)	1.9
Hak, 2005	Cohort, retrospective	NL	Primary care	Patients ≥60 years with an LRTI	455	455	75 (SD 8.6)	45	18	Composite of: i) Dysregulation of diabetes, stroke, heart failure, MI, ii) hospitalisation, iii) all-cause mortality, all within 30 days	21 (4.6)	24 (5.3)	65 (14.3)	3.6

Table 1. Continued

Study	Source of data	Country	Setting	Study population	Participants, n	Included episodes	Age, mean (SD) or median (IQR)	Male sex, %	Number of candidate predictors	Outcome[a]	Hospitalisation, n (%)	Mortality, n (%)	Total outcome events, n (%)	EPP
Winchester, 2009	Cohort, retrospective	UK	Primary care	Patients ≥1 year with an LRTI	151,088	151,088	54 (IQR 37)	42.9	17	Respiratory infection-related admission or mortality within 3 months	1,147 (0.8)	2,126 (1.4)	NR	67[b]
Van de Nadort, 2009	Cohort, retrospective	NL	Primary care	Elderly (≥80 years) LRTI patients	509	860	85.2 (no SD)	37.7	20	Hospitalisation or death within 30 days	NR	51 (5.9)	109 (12.7)	5.5
Myles, 2009	Cohort, retrospective	UK	Primary care	Patients ≥40 years with an LRTI	NR	3,681	NR	NR	4	All-cause mortality within 30 days	NR	905 (24.6)	905 (24.6)	226
Millett, 2015	Cohort, retrospective	UK	Primary care	Elderly (≥65 years) CAP patients	39,211	43,576	81 (IQR 75-87)	47	40	Hospital admission within 28 days	33,321 (76.5)	NR	33,321 (76.5)	833
Moore, 2019	Cohort, prospective	UK	Primary care	Patients ≥16 years with acute cough attributed to LRTI	28,846	28,846	NR	40.7	29	LRTI-related hospitalisation, mortality or late-onset pneumonia within 2-30 days	196 (0.7)	29 (0.1)	325 (1.1)	11.2

Table 1. Continued

Study	Source of data	Country	Setting	Study population	Participants, n	Included episodes	Age, mean (SD) or median (IQR)	Male sex, %	Number of candidate predictors	Outcome[a]	Hospitalisation, n (%)	Mortality, n (%)	Total outcome events, n (%)	EPP
Hamilton, 2021	Cohort, retrospective	UK	Primary care	Patients ≥50 years with pneumonia and history of lymphocyte test	28,556	28,556	76 (IQR 66-85)	49.5	14	28-day mortality	NR	2,544 (8.9)	2,544 (8.9)	182
Martínez-Redondo, 2021	Cohort, prospective	ESP	Primary care	Adults (≥18 years) with a COVID-19 pneumonia	161	161	53.0 (no SD)	50.9	4	Hospital admission, no time window reported	37 (23.0)	NR	37 (23.0)	9.3

Abbreviations: US, United States; FIN, Finland; NL, the Netherlands; UK, Unites Kingdom; ESP, Spain; LRTI, lower respiratory tract infection; CAP, community-acquired pneumonia; COVID-19, coronavirus disease 2019; SD, standard deviation; IQR, interquartile range; NR, not reported; MI, myocardial infarction; EPP, events per predictor.

[a] Time window of outcome is defined as days from first consultation within LRTI episode.

Table 2. Study characteristics of included prediction model studies

Study	Type of study	Model(s)	Source of data	Country	Setting	Study population	Partici-pants, n	Age, mean (SD)	Male sex, %	Outcome[c]	Hospitalisa-tion, n (%)	Mortality, n (%)	Total outcome events, n (%)	EPP[d]	Method of validation
Bont, 2007	D+V	New	Cohort, retrospec-tive	NL	Primary care	Elderly (≥65) LRTI patients	D: 3,166, V: 2,465	75.5 (no SD)	D: 45, V: NR	Hospitali-sation or all-cause mortality within 30 days	NR	D: 76 (2.4), V: 59 (2.4)	D: 274 (8.7), V: 178 (7.2)	13.7	Internal (random split-sample) and external
Bont, 2008	V	CRB-65	Cohort, prospec-tive	NL	Primary care	Elderly (≥65) CAP patients	314	77.3 (no SD)	46	Mortality within 30 days	47 (15)	11 (3.5)	11 (3.5)	-	External
Ochoa-Gondar, 2011	V	PSI, CURB-65, CRB-65	Cohort, retrospec-tive	ESP	Popu-lation-based	Elderly (≥65) patients with radio-graphically confirmed CAP	590	77.4 (7.6)	63.2	Mortality within 30 days	513 (86.9)	80 (13.6)	80 (13.6)	-	External
Francis, 2012	V	CRB-65	Cohort, prospec-tive	13 coun-tries[a]	Primary care	Adults (≥18) presenting with acute cough or suspected LRTI	339	49.3 (16.5)	NR	Hospi-talisation within 28 days	10 (2.9)	NR	10 (2.9)	-	External
Ochoa-Gondar, 2013	D+V	CORB-75 (new), CRB-65	Cohort, retrospec-tive	ESP	Popu-lation-based	Elderly (≥65) patients with radio-graphically confirmed CAP	D: 350, V: 260	78.1 (7.7)	D: 62.6, V: 59.2	Mortality within 30 days	D: 282 (80.6), V: 206 (79.2)	D: 46 (13.1), V: 34 (13.1)	D: 46 (13.1), V: 34 (13.1)	3.3	Internal-external (temporal, CORB-75), external (CRB-65)

Table 2. Continued

Study	Type of study	Model(s)	Source of data	Country	Setting	Study population	Partici-pants, n	Age, mean (SD)	Male sex, %	Outcome[c]	Hospitalisa-tion, n (%)	Mortality, n (%)	Total outcome events, n (%)	EPP[d]	Method of validation
Bruyn-donckx, 2018	D (RISCC85) + V (all models)	RISCC85 (new), PSI (stage I), CRB, CRB-65, CURB, CURB-65	Cohort, prospec-tive	6 coun-tries[b]	Primary care	Adults (≥18) with symptoms suggestive of LRTI	2,604	50 (17)	40	Re-con-sultation with new/ worsened symptoms or hospi-talisation within 28 days	11 (0.5)	0 (0)	521 (20.0)	6.4	Internal (random split sample, RISCC85), external (other models)
Moore, 2019	D+V	New	Cohort, prospec-tive	UK	Primary care	Patients ≥16 years with acute cough attributed to LRTI	28,846	NR	40.7	Hospi-talisation, mortality, or clinical diagnosis of pneu-monia within 2-30 days	196 (0.7)	29 (0.1)	325 (1.1)	11.2	Internal (bootstrap-ping)

Abbreviations: D, development; V, validation; NL, the Netherlands; ESP, Spain; UK, United Kingdom; LRTI, lower respiratory tract infection; CAP, community acquired pneumonia; SD, standard deviation; NR, not reported; EPP, events per predictor. a Belgium, Finland, Germany, Hungary, Italy, the Netherlands, Norway, Poland, Slovakia, Spain, Sweden, the United Kingdom. b Belgium, Germany, the Netherlands, Poland, Spain, the United Kingdom. c Time window of outcome is defined as days from first consultation. d Only provided for newly developed prediction models and based on derivation cohort.

The PM studies were conducted between 2000-2013 and the majority (4/7) used a prospective cohort design. In total, nine different PMs could be included in the synthesis: CRB, CRB-65, CURB, CURB-65, PSI, a model developed by Bont et al., RISSC85, CORB-75, and a model developed by Moore et al. (Table 2). Four studies developed a new prediction model[30,32,35,36] and five externally validated previously developed models[31–35], of which the CRB-65 was validated in all five studies. Clinical heterogeneity was mostly attributed to age (minimum age ranging from 16-65 years) and LRTI definition (ranging from suspected LRTI to radiographically confirmed CAP). Hospitalisation rates varied between 0.5% in LRTI patients and 86.9% in CAP patients. All PM studies used binary outcomes. Mortality was used as outcome in three studies, hospitalisation in one study, and a composite of both in one study. One study also included re-consultation with new or worsened symptoms in addition to hospitalisation in their outcome[35], while another study included late-onset pneumonia in addition to hospitalisation and mortality.[36]

Risk of bias assessment

Overall risk of bias was judged high for all studies (Figures S2 and S3). For PF studies, this was mostly due to issues in the statistical analysis domains. For instance, five studies selected predictors for multivariable analysis based on univariable associations[21–25], and two included multiple episodes per patient but did not perform recurrent event analysis.[25,27] Furthermore, in some studies an adequate definition of LRTI or PFs was lacking, and methods to account for missing data were mostly judged inadequate or not described.

Similarly, risk of bias due to the issues in the analysis domain was judged high for all PM studies. In all model development studies continuous predictors were dichotomised, and handling of missing data was judged inadequate or not reported. None of the PM studies reported any evaluation of model calibration.

Prognostic factors

Three PF studies did not report estimates of univariable associations of PFs, and these could also not be calculated from the reported data.[24,27,28] The remaining PF studies reported univariable analysis of 69 PFs, of which sex, age, and history of pulmonary disease were analysed most frequently (Figure S4). If reported in the primary study, absolute risks according to the presence and absence of individual PFs per study are presented in Table S5.

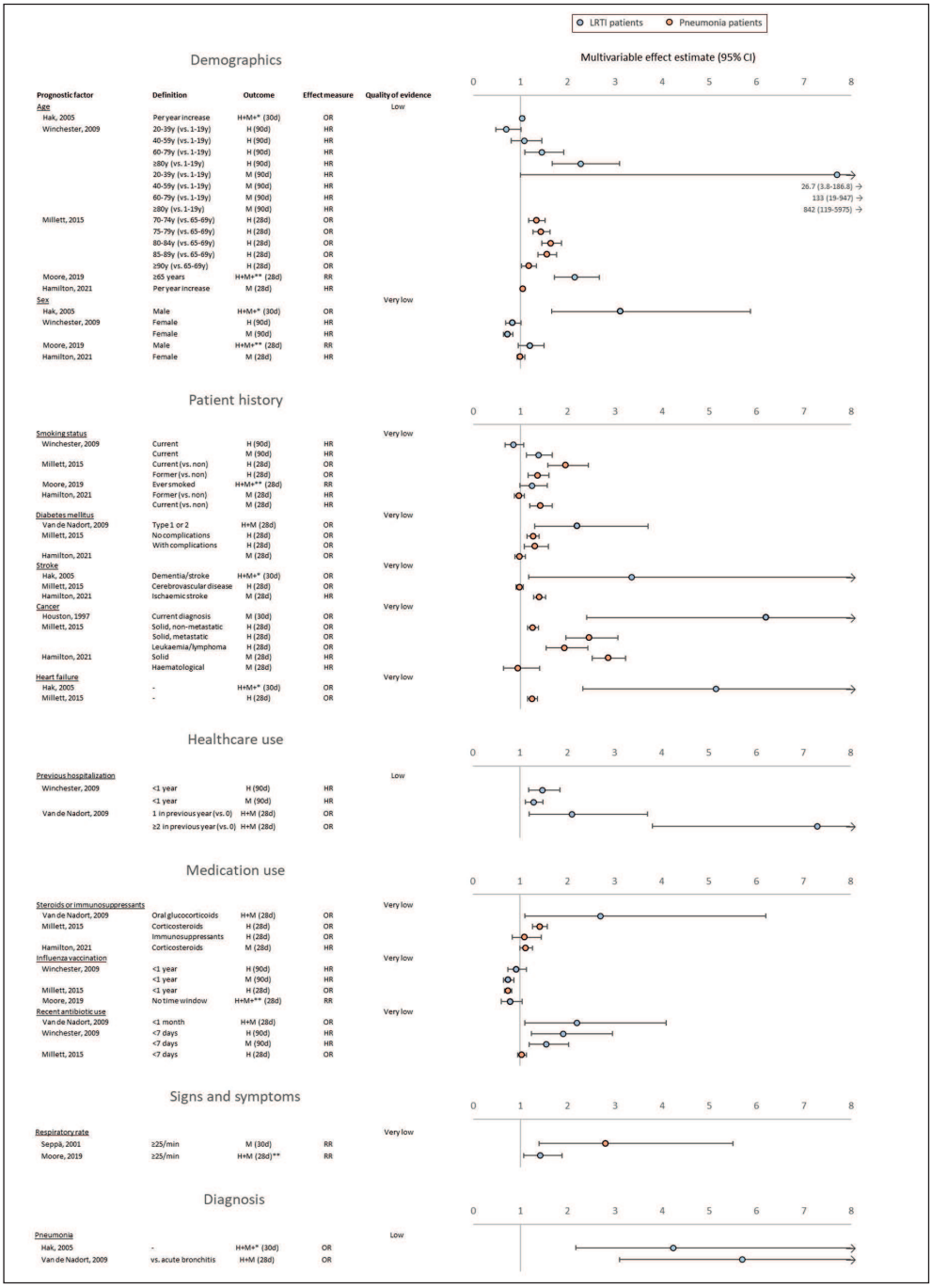
Estimates of multivariable associations were reported for 74 PFs (Figure S6), of which 21 were analysed in at least two studies. The number of covariates per multivariable analysis ranged from 3 to 29, and a common set of covariates across studies was lacking. Age was analysed in five studies, and all found increasing age to be associated with poor outcome.[23,24,27,28,36] Apart from increasing age, PFs that were identified as promising based on multiple multivariable analyses were sex (significant association in two/five

studies), current smoking (three/four studies), diabetes (two/three studies), a history of stroke (two/three studies), cancer (three/three studies), or heart failure (two/two studies), hospitalisation in the previous year (two/two studies), current use of systemic corticosteroids (two/three studies), antibiotic use in the previous month (two/three studies), a respiratory rate ≥25/minute (two/two studies), and diagnosis of pneumonia (two/two studies) (Figure 2). Having received an influenza vaccination in the previous year was the only promising candidate predictor estimated to have a negative association, based on a significant multivariable association in two/three studies. The quality of evidence was judged low for age, hospitalisation in the previous year, and diagnosis of pneumonia and very low for the remaining promising PFs (Table S7).

Prediction models

Several PMs included predictors that were also identified as promising candidate predictors based on PF studies (Table S8). Sex, current smoking, and influenza vaccination were the only promising candidate predictors that were not incorporated in any of the PMs. Predictors that were not identified as promising PFs but were incorporated in PMs include history of lung, renal or liver disease, extent to which complaints interfere with daily activities, sputum or coryza, chest pain, confusion, presence of crackles, heart rate, blood pressure, oxygen saturation, body temperature, blood urea nitrogen, and an overall GP severity assessment score.

Four studies developed a new PM (three for use in LRTI and one for CAP patients), in which the number of events per predictor ranged from 3.3-13.7.[30,34–36] C-statistics ranged from 0.63 (95% CI 0.61-0.67) to 0.82 (95% CI 0.75-0.88) after (internal or external) model validation (Figure 3). One study performed external validation[30], whereas other studies only conducted internal(-external) validation.[34–36] None reported measures of model calibration, while a total O/E ratio could not be calculated for any model.



Prognostic factor	Definition	Outcome	Effect measure	Quality of evidence
Demographics				
Age				Low
Hak, 2005	Per year increase	H+M+* (30d)	OR	
Winchester, 2009	20-39y (vs. 1-19y)	H (90d)	HR	
	40-59y (vs. 1-19y)	H (90d)	HR	
	60-79y (vs. 1-19y)	H (90d)	HR	
	≥80y (vs. 1-19y)	H (90d)	HR	
	20-39y (vs. 1-19y)	M (90d)	HR	
	40-59y (vs. 1-19y)	M (90d)	HR	
	60-79y (vs. 1-19y)	M (90d)	HR	
	≥80y (vs. 1-19y)	M (90d)	HR	
Millett, 2015	70-74y (vs. 65-69y)	H (28d)	OR	
	75-79y (vs. 65-69y)	H (28d)	OR	
	80-84y (vs. 65-69y)	H (28d)	OR	
	85-89y (vs. 65-69y)	H (28d)	OR	
	≥90y (vs. 65-69y)	H (28d)	OR	
Moore, 2019	≥65 years	H+M+** (28d)	RR	
Hamilton, 2021	Per year increase	M (28d)	HR	
Sex				Very low
Hak, 2005	Male	H+M+* (30d)	OR	
Winchester, 2009	Female	H (90d)	HR	
	Female	M (90d)	HR	
Moore, 2019	Male	H+M+** (28d)	RR	
Hamilton, 2021	Female	M (28d)	HR	
Patient history				
Smoking status				Very low
Winchester, 2009	Current	H (90d)	HR	
	Current	M (90d)	HR	
Millett, 2015	Current (vs. non)	H (28d)	OR	
	Former (vs. non)	H (28d)	OR	
Moore, 2019	Ever smoked	H+M+** (28d)	RR	
Hamilton, 2021	Former (vs. non)	M (28d)	HR	
	Current (vs. non)	M (28d)	HR	
Diabetes mellitus				Very low
Van de Nadort, 2009	Type 1 or 2	H+M (28d)	OR	
Millett, 2015	No complications	H (28d)	OR	
	With complications	H (28d)	OR	
Hamilton, 2021		M (28d)	OR	
Stroke				Very low
Hak, 2005	Dementia/stroke	H+M+* (30d)	OR	
Millett, 2015	Cerebrovascular disease	H (28d)	OR	
Hamilton, 2021	Ischaemic stroke	M (28d)	HR	
Cancer				Very low
Houston, 1997	Current diagnosis	M (30d)	OR	
Millett, 2015	Solid, non-metastatic	H (28d)	OR	
	Solid, metastatic	H (28d)	OR	
	Leukaemia/lymphoma	H (28d)	OR	
Hamilton, 2021	Solid	M (28d)	HR	
	Haematological	M (28d)	HR	
Heart failure				Very low
Hak, 2005	-	H+M+* (30d)	OR	
Millett, 2015	-	H (28d)	OR	
Healthcare use				
Previous hospitalization				Low
Winchester, 2009	<1 year	H (90d)	HR	
	<1 year	M (90d)	HR	
Van de Nadort, 2009	1 in previous year (vs. 0)	H+M (28d)	OR	
	≥2 in previous year (vs. 0)	H+M (28d)	OR	
Medication use				
Steroids or immunosuppressants				Very low
Van de Nadort, 2009	Oral glucocorticoids	H+M (28d)	OR	
Millett, 2015	Corticosteroids	H (28d)	OR	
	Immunosuppressants	H (28d)	OR	
Hamilton, 2021	Corticosteroids	M (28d)	HR	
Influenza vaccination				Very low
Winchester, 2009	<1 year	H (90d)	HR	
	<1 year	M (90d)	HR	
Millett, 2015	<1 year	H (28d)	OR	
Moore, 2019	No time window	H+M+** (28d)	RR	
Recent antibiotic use				Very low
Van de Nadort, 2009	<1 month	H+M (28d)	OR	
Winchester, 2009	<7 days	H (90d)	HR	
	<7 days	M (90d)	HR	
Millett, 2015	<7 days	H (28d)	OR	
Signs and symptoms				
Respiratory rate				Very low
Seppä, 2001	≥25/min	M (30d)	RR	
Moore, 2019	≥25/min	H+M (28d)**	RR	
Diagnosis				
Pneumonia				Low
Hak, 2005	-	H+M+* (30d)	OR	
Van de Nadort, 2009	vs. acute bronchitis	H+M (28d)	OR	

Figure 2. Overview of most promising candidate predictors for hospitalisation or mortality within 90 days based on multivariable analysis

Blue coloured points: studies on LRTI patients. Orange coloured points: studies on patients with pneumonia. Quality of evidence: as judged based on the Grading of Recommendations, Assessment,

Development and Evaluations framework. *Composite outcome also includes dysregulation of diabetes, stroke, heart failure, myocardial infarction. **Composite outcome also includes late-onset pneumonia. CAP , community-acquired pneumonia; H, hospitalisation; LRTI, lower respiratory tract infection; M, mortality; min, minute; RR, risk ratio.

All five PM validation studies externally validated the CRB-65.[31–35] C-statistics for 30-day mortality in CAP patients (range outcome events: 11 to 80) ranged from 0.72 (95% CI 0.66-0.78) to 0.79 (95% CI 0.65-0.92)[31,32,34] (Figure 3). The c-statistic for 28-day re-consultation or hospitalisation in LRTI patients was substantially lower: 0.53 (95% CI 0.51-0.56).(35) None reported measures of model calibration. A total O/E ratio could be calculated for the CRB-65 in four studies, ranging from 0.67 (95% CI 0.26-1.08) to 1.42 (95% CI 1.13-1.71).

Other validated PMs were PSI, CURB-65, CURB, and CRB. C-statistics for 28-day re-consultation or hospitalisation in LRTI patients ranged from 0.51 (95% CI 0.50-0.54) to 0.53 (95% CI 0.50-0.55).[35] Calibration measures were not reported and total O/E ratios could not be calculated. C-statistics of PSI and CURB-65 for 30-day mortality in CAP patients were 0.73 (95% CI 0.67-0.79) and 0.67 (95% CI 0.61-0.74), respectively.[32] Model calibration measures were not reported, while the total O/E ratios were estimated at 1.46 (95% CI 1.16-1.76) and 1.92 (95% CI 1.53-2.32), respectively. The quality of evidence was judged low for CRB-65 and very low for all other PMs included in this systematic review (Table S9).

DISCUSSION

Summary of main findings

This systematic review summarised evidence on PFs and PMs for hospitalisation and mortality within 90 days in patients presenting to primary care with LRTI. The risk of bias of included studies was judged high, mainly due to issues in the analysis domains. Of all PFs analysed, thirteen were identified as most promising; increasing age, sex, current smoking, diabetes, a history of stroke, cancer, or heart failure, hospitalisation in the previous year, current use of systemic corticosteroids, antibiotic use in the previous month, a respiratory rate ≥25/minute, and diagnosis of pneumonia were associated with poor outcome whereas seasonal influenza vaccination was estimated to have a negative association. Most of these promising candidate predictors have also been incorporated in the identified PMs. The secondary care-derived CRB-65, predicting mortality in CAP patients, and a newly developed PM by Bont et al.[30] predicting hospitalisation and mortality in elderly LRTI patients both showed promising results after external validation. However, none assessed model calibration, leaving the precision of predicted risks unknown.

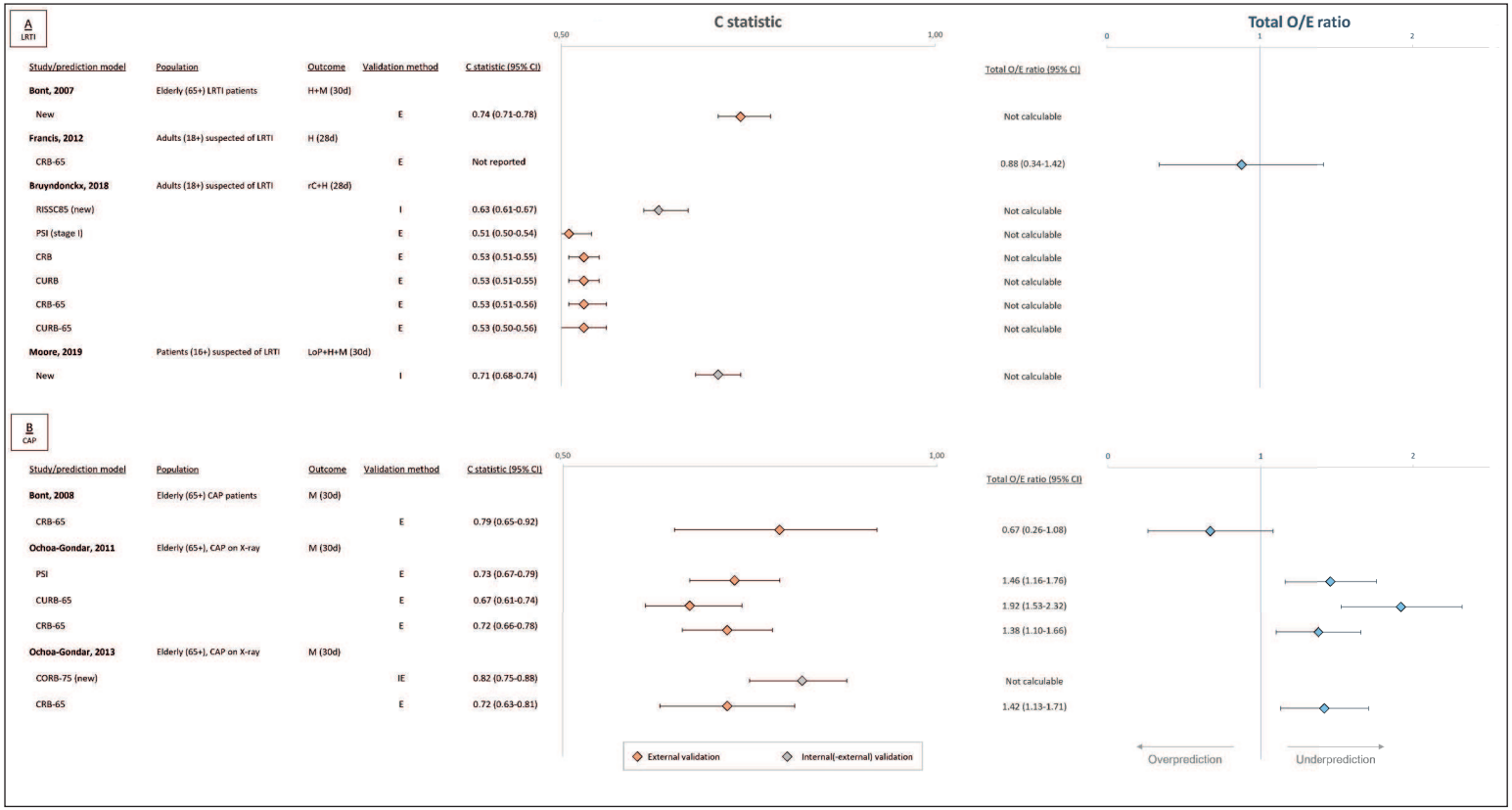


Figure 3. Overview of discrimination and calibration performance measures of the included prediction models

(**A**) Performance of prediction models evaluated in LRTI patients. (**B**) Performance of prediction models evaluated in community-acquired pneumonia patients. CAP, community-acquired pneumonia; CORB-75, confusion, oxygen saturation ≤90%, respiratory rate ≥30/minute, and systolic blood pressure ≤90mmHg or diastolic blood pressure ≤60mmHg, and age ≥75 years; CRB(-65), confusion, respiratory rate ≥30/minute, systiolic blood pressure ≤90mmHg

or diastolic blood pressure ≤60mmHg, and age ≥65 years (and age ≥65 years); CURB(-65), confusion, blood urea nitrogen >7 millimole/liter, respiratory rate ≥30/minute, systiolic blood pressure ≤90mmHg or diastolic blood pressure ≤60mmHg (and age ≥65 years); E, external; H, hospitalisation; I, internal; IE, internal external; LoP, late-onset pneumonia; LRTI, lower respiratory tract infection; M, mortality; O/E, observed over expected; PSI, Pneumonia Severity Index; rC, reconsultation; RISSC85, a priori risk of poor outcome by country, inference in daily activities, number of years stopped smoking, severe sputum, presence of crackles, and diastolic blood pressure <85mmHg.

Comparison with existing literature

During the coronavirus disease 2019 (COVID-19) pandemic, many studies have focused on PFs associated with poor outcomes in COVID-19 patients, and some were similar to those identified in our review of all-cause LRTI patients. A recent, field-wide systematic review and meta-analysis of PFs for adverse outcomes in COVID-19 patients found cancer, insulin use, and smoking – among other factors – to be associated with mortality, and age with both hospitalisation and mortality.[37] Another community-based systematic review of European COVID-19 patients found male sex, heart failure, and diabetes to be associated with hospitalisation, whereas stroke, heart failure, and neurologic disease were associated with increased mortality risk.[38] Although we did not identify studies that assessed the prognostic value of specific viral respiratory pathogens such as SARS-CoV-2, respiratory syncytial virus and influenza, it is likely that the specific pathogen is a relevant prognostic factor, and point-of-care tests for respiratory pathogens are increasingly available in the post-pandemic era. The potential usefulness of such point-of-care tests is supported by evidence of patients with a viral LRTI admitted to the emergency room in whom specific viral pathogens were associated with increased risk of hospitalisation.[39]

Some of the identified PMs have been incorporated in LRTI guidelines – mainly those focusing on CAP – as additional tool to aid patient management decisions. The CAP guideline issued by the American Thoracic Society (ATS) and the Infectious Diseases Society of America (IDSA) supports the use of PSI or CURB-65 – in addition to clinical judgement – to determine the appropriate level of care, i.e. inpatient versus outpatient management.[40,41] Although this guideline applies to all levels of health care, the value of the suggested models for GPs is limited since models' laboratory and radiographic features are not routinely available in primary care. The UK National Institute for Health and Care Excellence (NICE) guideline on pneumonia advocates the use of CRB-65; it recommends GPs to consider hospital admission in patients with a score of ≥1.[42] Guidelines vary with regard to the incorporation of individual factors that are associated with poor outcome. The ATS/IDSA guideline and the Dutch primary care guideline both list tachypnoea and confusion as indicators of disease severity, while the Dutch guideline also emphasised the importance of PFs such as age, history of heart failure, and diabetes.[40,43] The NICE guideline does not explicitly highlight individual factors associated with poor outcome in adult patients.[42]

Similar to our findings, a review of the performance of CRB-65 for predicting 30-day mortality among CAP patients in community and hospital settings concluded that this model has been insufficiently validated for use in primary care.[44] Although the population of this review slightly differed from ours by also including patients who visited the emergency department, a similarly low number of outcome events was found. Contrary to the slight overprediction reported in this review, our synthesis of more recent studies suggested

underprediction of risks based on total O/E ratios. However, in absence of proper assessment of model calibration the true precision of predicted risks remains to be elucidated.

Strengths and limitations

By covering both PFs and PMs in a single review, we have provided a comprehensive overview of available prognostic literature of primary care LRTI patients. We prospectively registered our review protocol, conducted a broad search with incorporation of a validated prognostic search string, and thoroughly reviewed citations of eligible studies to assure all relevant literature was included in the synthesis. By including various prognostic study designs, our synthesis provides a comprehensive overview of all prognostic evidence on primary care LRTI patients. Studies assessing the prognostic value of individual factors, developing a new or validating an existing model all provide insight on important PFs that could be used in future prognostic research, whereas studies validating a new or previously developed model such as CRB-65 provide insight on the suitability for its use in clinical practice.

Some limitations should, however, be acknowledged. First, our findings should be interpreted with caution since risk of bias was judged high for all included studies and the quality of the evidence was judged low to very low for the promising PFs and PMs. Second, due to substantial heterogeneity in domain, predictor, and outcome definitions between studies, we refrained from conducting meta-analyses of PF effect estimates or PM performance measures, which makes direct comparison of study results difficult. It is plausible that PF effect estimates for poor outcome vary according to age, diagnosis, and disease severity (a priori risk of event). For example, half of the participants in Hak et al.[23] had either pneumonia or an exacerbation of chronic obstructive pulmonary disease, whereas Moore et al.[36] had fewer than 2% of such patients. Furthermore, one study included patients aged >1 year (median age of included participants: 54 years (interquartile range: 37)) but did not specifically report separate results for age strata.[24] Given the median age, we considered including this study on prognostic factors in our review justified. Consistency in effect estimates would have further facilitated direct comparison of PFs, but converting the various reported estimates to odds ratios appeared not possible. Albeit our synthesis of PFs provides some useful information for clinicians, it merely informs future research aimed at developing or updating PMs. Third, we excluded most studies on patients with a specific respiratory pathogen (e.g. SARS-CoV-2 or influenza) since these included patients with both upper and lower respiratory tract infections and did not report stratified results for LRTI. As such, we were unable to present any respiratory pathogen-specific results regarding prognostic factors for poor outcome in patients presenting to primary care with LRTI. However, since GPs are typically unaware of the specific pathogen in LRTI patients, such results would have limited applicability to current clinical practice. Lastly, PM building techniques were generally suboptimal which increased risk of bias, and none of the identified PM studies assessed model calibration. In contrast to

discrimination, calibration generally receives little attention when assessing PM performance, whereas well discriminating but poorly calibrated models can lead to imprecise, misleading predictions.[45] Total O/E ratios calculations provided some insight into mean model calibration – the most basic level in the hierarchy of calibration assessment – but this is insufficient to provide guidance on the safe use of a PM in clinical practice.[46]

Implications for research and clinical practice

Based on our synthesis, implementation of existing PMs for individualised risk prediction of 90-day hospitalisation or all-cause mortality in primary care LRTI patients in everyday practice is hampered by incomplete assessment of model performance. The identified candidate predictors provide useful information for clinicians about factors associated with poor outcome in primary care LRTI patients and warrant consideration when developing or updating PMs using state-of-the-art development and appropriate validation techniques.

REFERENCES

1. Bergmann M, Haasenritter J, Beidatsch D, et al. Prevalence, aetiologies and prognosis of the symptom cough in primary care: a systematic review and meta-analysis. *BMC Fam Pract*. 2021;**22**(1):1–19.
2. Hak E, Rovers MM, Kuyvenhoven MM, Schellevis FG, Verheij TJM. Incidence of GP-diagnosed respiratory tract infections according to age, gender and high-risk co-morbidity: the second Dutch national survey of general practice. *Fam Pract*. 2006;**23**(3):291–4.
3. Little P, Stuart B, Moore 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.
4. Moore M, Stuart B, Coenen S, 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.
5. Little P, Stuart B, Smith S, et al. Antibiotic prescription strategies and adverse outcome for uncomplicated lower respiratory tract infections: prospective cough complication cohort (3C) study. *BMJ*. 2017;**357**:j2148.
6. Welte T, Torres A, Nathwani D. Clinical and economic burden of community-acquired pneumonia among adults in Europe. *Thorax*. 2012;**67**(1):71–9.
7. McLaughlin JM, Khan FL, Thoburn EA, Isturiz RE, Swerdlow DL. Rates of hospitalization for community-acquired pneumonia among US adults: a systematic review. *Vaccine*. 2020;**38**(4):741–51.
8. Teepe J, Broekhuizen BDL, Loens K, et al. Predicting the presence of bacterial pathogens in the airways of primary care patients with acute cough. *CMAJ*. 2017;**189**(2):E50–5.
9. Minnaard MC, De Groot JAH, Hopstaken RM, 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.
10. Lim WS, Van Der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax*. 2003;**58**(5):377–82.
11. Fine M, Auble T, Yealy D, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med*. 1997;**336**(4):243–50.
12. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;**372**.
13. Geersing GJ, Bouwmeester W, Zuithoff P, et al. Search filters for finding prognostic and diagnostic prediction studies in medline to enhance systematic reviews. *PLoS One*. 2012;**7**(2):3–8.
14. 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).
15. Debray TPA, Damen JAAG, Snell KIE, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ*. 2017;**356**:i6460.
16. Riley RD, Moons KGM, Snell KIE, et al. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ*. 2019;**364**:k4597.
17. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med*. 2013;**158**(4):280–6.
18. 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):51–8.
19. Foroutan F, Guyatt G, Zuk V, et al. GRADE Guidelines 28: Use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks. *J Clin Epidemiol*. 2020;**121**:62–70.

20. Brozek JL, Canelo-Aybar C, Akl EA, et al. GRADE Guidelines 30: the GRADE approach to assessing the certainty of modeled evidence—An overview in the context of health decision-making. *J Clin Epidemiol*. 2021;**129**:138–50.
21. Houston MS, Silverstein MD, Suman VJ. Risk factors for 30-day mortality in elderly patients with lower respiratory tract infection: community-based study. *Arch Intern Med*. 1997;**157**(19):2190–5.
22. Seppä Y, Bloigu A, Honkanen PO, Miettinen L, Syrjälä H. Severity assessment of lower respiratory tract infection in elderly patients in primary care. *Arch Intern Med*. 2001;**161**(22):2709–13.
23. 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.
24. Winchester CC, Macfarlane T V., Thomas M, Price D. Antibiotic prescribing and outcomes of lower respiratory tract infection in UK primary care. *Chest*. 2009;**135**(5):1163–72.
25. van de Nadort C, Smeets HM, Bont J, et al. Prognosis of primary care patients aged 80 years and older with lower respiratory tract infection. *Br J Gen Pract*. 2009;**59**(561):261–5.
26. Myles P, Hubbard R, Gibson J, et al. The impact of statins, ace inhibitors and gastric acid suppressants on pneumonia mortality in a UK general practice population cohort. *Pharmacoepidemiol Drug Saf*. 2009;**18**(8):697–703.
27. Millett ERC, De Stavola BL, Quint JK, Smeeth L, Thomas SL. Risk factors for hospital admission in the 28 days following a community-acquired pneumonia diagnosis in older adults, and their contribution to increasing hospitalisation rates over time: a cohort study. *BMJ Open*. 2015;**5**(12):e008737.
28. Hamilton F, Arnold D, Payne R. Association of prior lymphopenia with mortality in pneumonia: a cohort study in UK primary care. *Br J Gen Pract*. 2021;**71**(703):E148–56.
29. Martínez-Redondo J, Comas C, Salud JP, et al. The risk of hospitalization in COVID-19 patients can be predicted by lung ultrasound in primary care. *Int J Environ Res Public Health*. 2021;**18**(11).
30. Bont J, Hak E, Hoes AW, et al. A prediction rule for elderly primary-care patients with lower respiratory tract infections. *Eur Resp J*. 2007;**29**(5):969–75.
31. 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 CRB-65 severity assessment tool. *Arch Intern Med*. 2008;**168**(13):1465–8.
32. Ochoa-Gondar O, Vila-Corcoles A, Rodriguez-Blanco T, et al. Comparison of three predictive rules for assessing severity in elderly patients with CAP. *Int J Clin Pract*. 2011;**65**(11):1165–72.
33. Francis NA, Cals JW, Butler CC, 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.
34. Ochoa-Gondar O, Vila-Corcoles A, Rodriguez-Blanco T, et al. Validation of the CORB75 (confusion, oxygen saturation, respiratory rate, blood pressure, and age ≥75 years) as a simpler pneumonia severity rule. *Infection*. 2013;**42**(2):371–8.
35. Bruyndonckx R, Hens N, Verheij TJM, 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.
36. Moore M, Stuart B, Lown M, et al. Predictors of adverse outcomes in uncomplicated lower respiratory tract infections. *Ann Fam Med*. 2019;**17**(3):231–8.
37. Bellou V, Tzoulaki I, van Smeden M, et al. Prognostic factors for adverse outcomes in patients with COVID-19: a field-wide systematic review and meta-analysis. *Eur Respir J*. 2022;**59**(2):2002964.
38. Vardavas CI, Mathioudakis AG, Nikitara K, et al. Prognostic factors for mortality, intensive care unit and hospital admission due to SARS-CoV-2: a systematic review and meta-analysis of cohort studies in Europe. *Eur Respir Rev*. 2022;**31**(166):220098.

39. Lemarie B, Boussaid G, Gault E, et al. Predictors of hospitalization and superinfection in viral respiratory tract infections between influenza and paramyxoviruses: the SUPERFLUOUS study. *J Infect Dis*. 2022;**226**(6):1027–35.
40. Metlay JP, Waterer GW, Long AC, et al. Diagnosis and treatment of adults with community-acquired pneumonia. *Am J Respir Crit Care Med*. 2019;**200**(7):E45–67.
41. Armstrong CAFP. Community-acquired pneumonia: updated recommendations from the ATS and IDSA. *Am Fam Physician*. 2020;**102**(2):121–4.
42. National Institute for Health and Care Excellence. *Pneumonia in adults: diagnosis and management (CG191)* [internet]. London: National Institute for Health and Care Excellence; 2014. Accessed May 7, 2025. Available at: https://www.nice.org.uk/guidance/cg191.
43. Verheij T, Hopstaken R, Prins J, et al. *NHG-Standaard Acuut hoesten* [internet]. Utrecht: Nederlands Huisartsen Genootschap; 2020. Accessed May 6, 2025. Available at: https://richtlijnen.nhg.org.
44. McNally M, Curtain J, O'Brien KK, Dimitrov BD, Fahey T. Validity of British Thoracic Society guidance (the CRB-65 rule) for predicting the severity of pneumonia in general practice: systematic review and meta-analysis. *Br J Gen Pract*. 2010;**60**(579):423–33.
45. van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the achilles heel of predictive analytics. *BMC Med*. 2019;**17**(1):230.
46. van Calster B, Nieboer D, Vergouwe Y, et al. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol*. 2016;**74**:167–76.

SUPPLEMENTARY MATERIALS FOR CHAPTER 2

Table S1: search strategy

Search date: 27 June 2022

Pubmed	Embase	The Cochrane Library
(((((“RTI”[Title/Abstract] OR “respiratory tract infection*”[Title/Abstract] OR “respiratory infection*”[Title/Abstract] OR “pneumonia*”[Title/Abstract] OR “bronchitis”[Title/Abstract] OR “lung infection*”[Title/Abstract] OR “chest infection*”[Title/Abstract]) OR (“Respiratory Tract Infections”[Mesh])) AND (((Validat* OR Predict*[Title] OR Rule*) OR (Predict* AND (Outcome* OR Risk* OR Model*)) OR ((History OR Variable* OR Criteria OR Scor* OR Characteristic* OR Finding* OR Factor*) AND (Predict* OR Model* OR Decision* OR Identif* OR Prognos*)) OR (Decision* AND (Model* OR Clinical* OR Logistic Models)) OR (Prognostic AND (History OR Variable* OR Criteria OR Scor* OR Characteristic*	(('rti':ti,ab,kw OR 'respiratory tract infection*':ti,ab,kw OR 'respiratory infection*':ti,ab,kw OR 'pneumonia*':ti,ab,kw OR 'bronchitis':ti,ab,kw OR 'lung infection*':ti,ab,kw OR 'chest infection*':ti,ab,kw) OR ('respiratory tract infection'/exp OR 'respiratory tract inflammation'/exp)) AND ((validat* OR predict*:ti OR rule* OR (predict* AND (outcome* OR risk* OR model*)) OR ((history OR variable* OR criteria OR scor* OR characteristic* OR finding* OR factor*) AND (predict* OR model* OR decision* OR identif* OR prognos*)) OR (decision* AND (model* OR clinical* OR logistic) AND models) OR (prognostic AND (history OR variable* OR criteria OR scor* OR characteristic* OR finding* OR OR 'receiver operating	(RTI OR respiratory tract infection OR respiratory infection OR pneumonia OR bronchitis OR lung infection OR chest infection) AND (hospitalisation OR hospitalization OR hospital OR complication OR complicated OR morbidity OR death OR mortality OR survival OR survive) AND (Primary care OR general practice OR general practitioner OR family practice OR GP) in Title Abstract Keyword
OR Finding* OR Factor* OR Model*))) OR (“Stratification” OR “ROC Curve”[Mesh] OR “Discrimination” OR “Discriminate” OR “c-statistic” OR “c statistic” OR “Area under the curve” OR “AUC” OR “Calibration” OR “Indices” OR “Algorithm” OR “Multivariable”))) AND ((“hospital*”[Title/Abstract] OR “complicat*”[Title/Abstract] OR “morbidity”[Title/Abstract] OR “death”[Title/Abstract] OR “mortality”[Title/Abstract] OR “surviv*”[Title/Abstract]) OR (“Mortality”[Mesh] OR “Hospitalization”[Mesh] OR “Death”[Mesh]))) AND ((“Primary care”[Title/Abstract] OR “general practi*”[Title/Abstract] OR “family practice”[Title/Abstract] OR “GP”[Title/Abstract]) OR (“Primary Health Care”[Mesh] OR “Family Practice”[Mesh] OR “General Practice”[Mesh]))) NOT ((“child*”[Title/Abstract] OR	characteristic'/exp OR 'discrimination' OR 'discriminate' OR 'c-statistic' OR 'c statistic' OR 'area under the curve' OR 'auc' OR 'calibration' OR 'indices' OR 'algorithm' OR 'multivariable')) AND (('hospital*':ti,ab,kw OR 'complicat*':ti,ab,kw OR 'morbidity':ti,ab,kw OR 'death':ti,ab,kw OR 'mortality':ti,ab,kw OR 'surviv*':ti,ab,kw) OR ('hospitalization'/exp OR 'hospital admission'/exp OR 'death'/exp OR 'mortality'/exp)) AND (('primary care':ti,ab,kw OR 'general practi*':ti,ab,kw OR 'family practice':ti,ab,kw OR 'gp':ti,ab,kw) OR ('primary medical care'/exp OR 'primary health care'/exp OR 'general practice'/exp)) NOT (('child*':ti,ab,kw OR 'pediatric*':ti,ab,kw	

Table S1: Continued

Search date: 27 June 2022

Pubmed	Embase	The Cochrane Library
"pediatric*"[Title/Abstract] OR "paediatric*"[Title/Abstract] OR "infan*"[Title/Abstract] OR "adolescen*"[Title/Abstract] OR "young*"[Title/Abstract] OR "newborn*"[Title/Abstract]) OR ("Infant"[Mesh] OR "Child"[Mesh] OR "Adolescent"[Mesh] OR "Pediatrics"[Mesh] OR "Young Adult"[Mesh]))	OR 'paediatric*':ti,ab,kw OR 'infan*':ti,ab,kw OR 'adolescen*':ti,ab,kw OR 'young*':ti,ab,kw OR 'newborn*':ti,ab,kw) OR ('child'/exp OR 'adolescent'/exp OR 'pediatrics'/exp))	
Filters: none Language restrictions: none Publication date restrictions: none	Filters: none Language restrictions: none Publication date restrictions: none	Filters: none Language restrictions: none Publication date restrictions: none
Records identified: 1592	Records identified: 1392	Records identified: 207

Figure S2: Overall assessment of risk of bias and applicability of included studies

QUIPS assessment of prognostic factor studies

Study	RoB					Applicability					Overall	
	Study participation	Study attrition	PF measurement	Outcome measurement	Statistical analysis and reporting	Participants	Outcome	Timing of PF measurement	Timing of outcome measurement	Setting	RoB	Applicability
Houston, 1997	±	+	-	+	-	+	+	+	+	±	-	±
Seppä, 2001	-	±	-	+	-	+	+	+	+	+	-	+
Hak, 2005	+	+	±	+	-	+	±	+	+	+	-	±
Myles, 2009	-	+	+	±	-	+	+	+	+	+	-	+
van de Nadort, 2009	+	+	+	±	-	±	+	+	+	+	-	±
Winchester, 2009	+	+	-	+	-	±	±	+	+	+	-	±
Millett, 2015	+	+	±	+	-	+	+	+	+	+	-	+
Moore, 2019	-	+	-	+	±	±	±	+	+	+	-	±
Hamilton, 2021	+	-	-	+	±	±	+	+	+	+	-	±
Martínez-Redondo, 2021	-	+	-	-	-	±	+	-	±	+	-	-

PROBAST assessment of prediction model studies

Study	RoB				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	RoB	Applicability
Bont, 2007	+	+	-	-	+	+	+	-	+
Bont, 2008	-	+	?	-	+	+	+	-	+
Ochoa-Gondar, 2011	+	?	?	-	?	+	+	-	?
Francis, 2012	-	+	-	-	+	+	+	-	+
Ochoa-Gondar, 2013	+	+	+	-	?	+	+	-	?
Bruyndonckx, 2018	+	+	+	-	+	+	?	-	?
Moore, 2019	+	+	+	-	+	+	?	-	?

Figure S3: Risk of bias assessment and applicability per individual included study

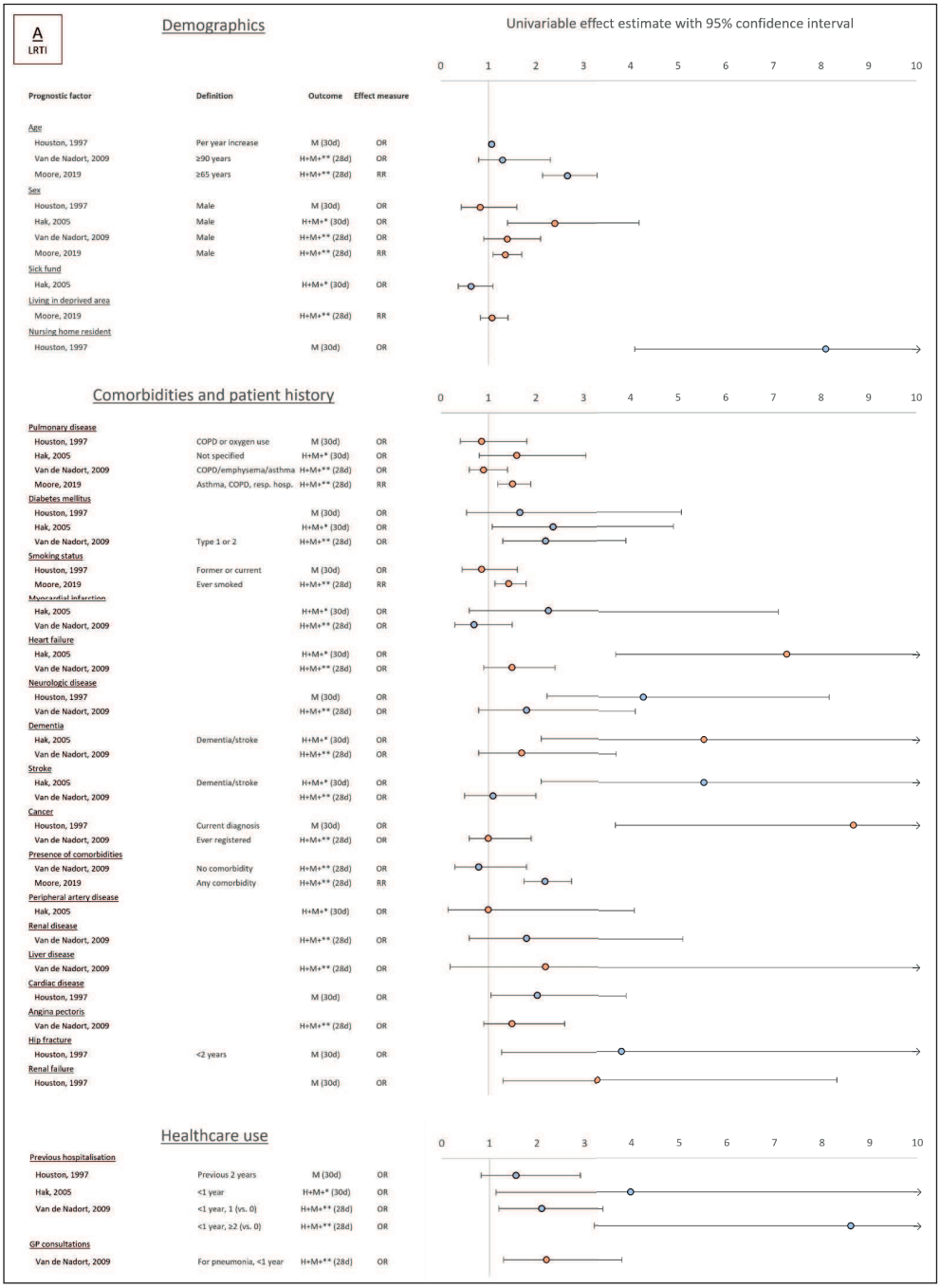


Figure S4: figure of all PF effect estimates based on univariable analysis

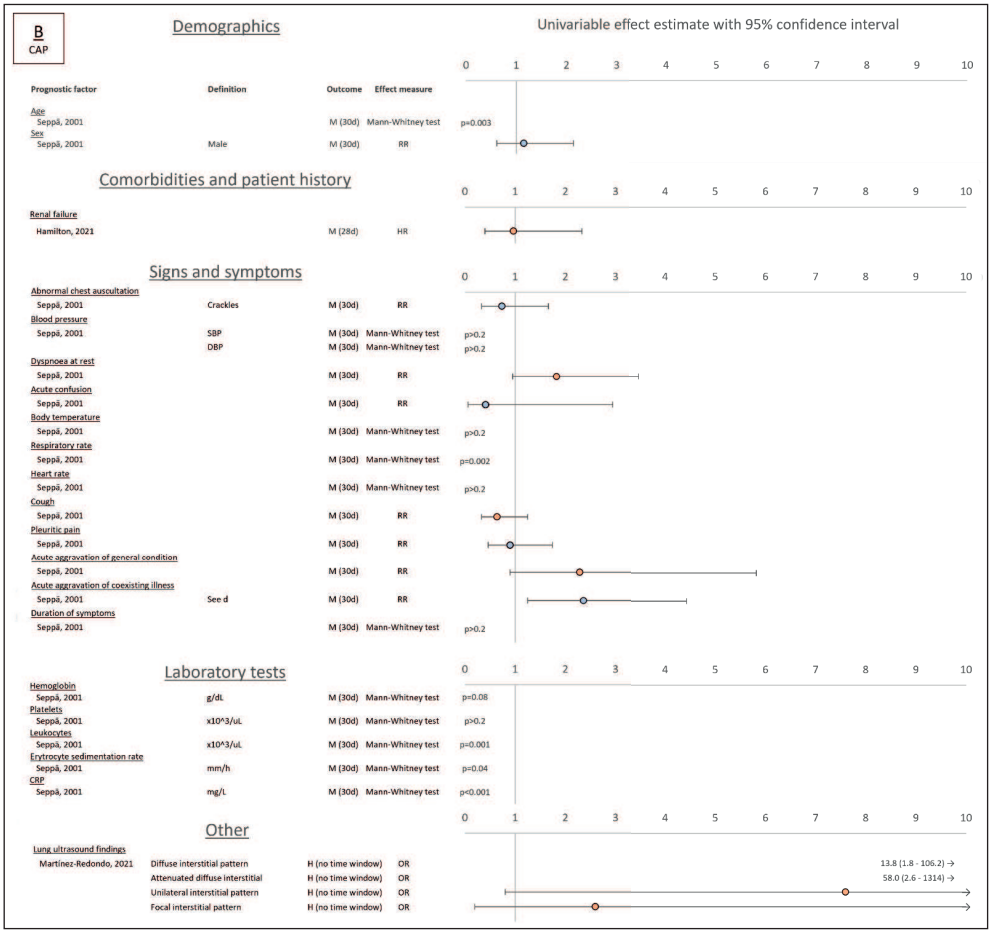


Figure S4: Continued

A. Univariable analyses of prognostic factors from studies on LRTI patients. B. Univariable analysis of prognostic factors from studies on pneumonia patients.

* Composite outcome also includes dysregulation of diabetes, stroke, heart failure, MI.

** Composite outcome also includes late onset pneumonia.

a: rigor, hemoptysis, pleuritic pain.

b: chills, cough, congestion, fever, dyspnea, sputum production.

c: poor eating, confusion, lethargy.

d: impairment of glucose balance in diabetic patients, deterioration of congestive heart failure.

Abbreviations: LRTI, lower respiratory tract infection; H, hospitalisation; M, mortality; OR, odds ratio; RR, risk ratio; HR, hazard ratio; COPD, chronic obstructive pulmonary disease; GP, general practitioner; ICS, inhalation corticosteroids; ACE, angiotensin converting enzyme; SBP, systolic blood pressure; DBP, diastolic blood pressure; min, minute; g, gram; dL, deciliter; uL, microliter; mm, millimeter; h, hour; L, liter.

Table S5. Absolute risks according to absence and presence of individual (categorical) prognostic factors

Study	Houston, 1997	Seppä, 2001	Hak, 2005	Winchester, 2009		Van de Nadort, 2009	Myles, 2009	Millett, 2015	Moore, 2019	Hamilton, 2021	Martínez-Redondo, 2021
Absolute risks provided or calculable?	No	Yes	Yes	Yes		Yes	Yes	Yes	Yes	No	Yes
Outcome definition	Mortality <30 days	Mortality <30 days	Composite including mortality <30 days	Hosp. <3 months	Mortality <3 months	Hosp. or mortality <30 days	Mortality <30 days	Hosp. <28 days	Composite including hosp. or mortality <30 days	Mortality <28 days	Hosp.
Baseline risk of outcome	10.7 %	4.0 %	14.3%	0.8 %	1.4 %	12.7 %	24.6 %	76.5 %	1.1 %	8.9 %	23.0 %
Prognostic factor	**Absolute risk according to prognostic factor (present/absent)**										
Demographics											
Male sex		4.3/3.7 %	20.2/9.5 %	0.8/0.7 %	1.5/1.3 %	14.2/11.8 %		79.5/73.8 %	1.3/0.9 %		
Sick fund			12.3/18.1 %								
Living in deprived area									1.2/1.1 %		
Medical history											
Heart failure			46.7/10.7 %			16.4/10.8 %		79.6/75.5 %			
Dementia/ stroke			45.0/12.9 %			17.0/12.4 % 12.0/12.8 %		66.8/78.2 %			
Myocardial infarct.			26.7/13.9 %			11.5/12.8 %		83.2/75.4 %			
Periph. arterial dis.			14.3/14.3 %					82.7/75.5 %			
Diabetes mellitus			26.2/13.1 %			23.1/10.8 %		82.6/75.2 %			
Prior hospitalisation			38.5/13.6 %			36.0/10.1 %					
Smoking				0.6/0.8 %	0.9/1.7 %				13.1/0.9 %		

Table S5. Continued

Study	Houston, 1997	Seppä, 2001	Hak, 2005	Winchester, 2009		Van de Nadort, 2009	Myles, 2009	Millett, 2015	Moore, 2019	Hamilton, 2021	Martínez-Redondo, 2021
Alcohol				0.7/0.8 %	1.1/2.3 %						
Asthma/COPD			19.7/13.3 %	0.8/0.8 %	0.7/1.5 %	12.3/12.9 %		83.4/71.8 %	1.5/1.0 %		
All. rhinitis				0.7/0.8 %	0.5/1.5 %						
Malignancies						11.9/12.8 %		80.7/75.6 %			
Angina pectoris						18.2/11.4 %		81.5/75.2 %			
Neurological dis.						18.0/12.3 %		73.1/76.8 %			
Renal disease						19.0/12.5 %		88.6/73.8 %			
Liver disease						25.0/12.6 %		85.5/76.4 %			
Connective tissue dis.								81.9/75.9 %			
Peptic ulcer								81.1/76.0 %			
Hemiplegia								76.4/76.5 %			
Leukaemia/lymphoma								85.0/76.2			
Cerebrovascular dis.								74.5/77.2 %			
Terminal illness								67.1/76.9 %			
Medication use											
Antidepressant/ benzodiazepines			18.1/12.2 %			15.4/11.6 %					
Corticosteroids			13.8/14.4 %								
Oral prednisone			19.2/13.6 %			37.5/11.7 %		83.0/75.4 %			
Immunosuppressants								85.3/76.3 %			
ICS								84.7/74.6 %			
Bronchodilator			19.0/13.6 %						1.4/1.0 %		

Table S5. Continued

Study	Houston, 1997	Seppä, 2001	Hak, 2005	Winchester, 2009		Van de Nadort, 2009	Myles, 2009	Millett, 2015	Moore, 2019	Hamilton, 2021	Martínez-Redondo, 2021
Recent antibiotics			4.5/14.8 %	2.8/0.7 %	3.4/1.4 %	24.2/11.8 %		74.8/77.2 %			
Vaccination											
Influenza				0.9/0.4 %	2.4/0.8 %			76.2/69.7 %	1.5/0.9 %		
Pneumococcal				1.0/0.7 %	2.6/1.1 %			81.3/33.6 %	1.8/1.0 %		
Statin use							6.8/25.5 %				
ACEi use							18.6/25.1 %				
PPI use							26.1/24.4 %				
H2 antagonist							31.7/24.3 %				
Presenting signs and symptoms											
Cough		3.6/5.6 %									
Dyspnoea		5.3/2.9 %							1.3/0.8 %		
Pleuritic pain		3.8/4.2 %									
Confusion		1.8/4.4 %							1.3/1.1 %		
Aggravation of											
General condition		3.7/4.7 %									
Coexisting illness		7.6/3.2 %									
Abn. auscultation		3.7/5.0%									
Fever									1.1/1.1 %		
Chills									1.4/1.0 %		
Chest pain									1.4/1.0 %		
No coryza									1.4/0.9 %		
Muscle aches									1.1/1.2 %		
Diarrhoea									1.1/1.1 %		
Purulent sputum									1.0/1.3 %		

Table S5. Continued

Study	Houston, 1997	Seppä, 2001	Hak, 2005	Winchester, 2009	Van de Nadort, 2009	Myles, 2009	Millett, 2015	Moore, 2019	Hamilton, 2021	Martínez-Redondo, 2021
Bloody sputum								1.6/1.1 %		
Severity assessment >5/10								1.7/0.7 %		
Respiratory rate >24/min								2.1/1.0 %		
Temperature >37.8								2.4/0/1 %		
Pulse >100/min								1.6/1.1 %		
O_2 saturation <95 %								3.5/0.9 %		
SBP<90 or DBP<60 mmHg								1.8/1.1 %		
Crackles								15.7/0.8 %		
Bronchial breathing								1.7/1.1 %		
Wheezing								1.5/1.0 %		
LRTI diagnosis										
Pneumonia			22.6/6.1 %		27.3/8.0 %					
Exacerbation COPD					9.9/14.5 %					
Acute bronchitis					5.8/16.5 %					
Frailty factors										
Recent carer							79.6/76.3 %			
Visual impairment							78.8/75.3 %			
Self-care							79.7/76.4 %			
Anxious/depressed							76.0/76.5 %			
Bedsore/ulcer							59.7/77.0 %			

Table S5. Continued

Study	Houston, 1997	Seppä, 2001	Hak, 2005	Winchester, 2009	Van de Nadort, 2009	Myles, 2009	Millett, 2015	Moore, 2019	Hamilton, 2021	Martínez-Redondo, 2021
Mobility issues							79.0/76.3 %			
Tired							74.4/76.6 %			
Low weight/nutrition							75.1/76.7 %			
Incontinence							71.7/77.0			
History of falling							76.4/76.5 %			
Lung ultrasound										
Abnormal										28.2/3.33 %

Abbreviations: NR, not reported; COPD, chronic obstructive pulmonary disease; ICS, inhaled corticosteroids; ACEi, angiotensin converting enzyme inhibitor; PPI, protonpump inhibitor; H2, histamine-2; SBP, systolic blood pressure; DBP, diastolic blood pressure; LRTI, lower respiratory tract infection.

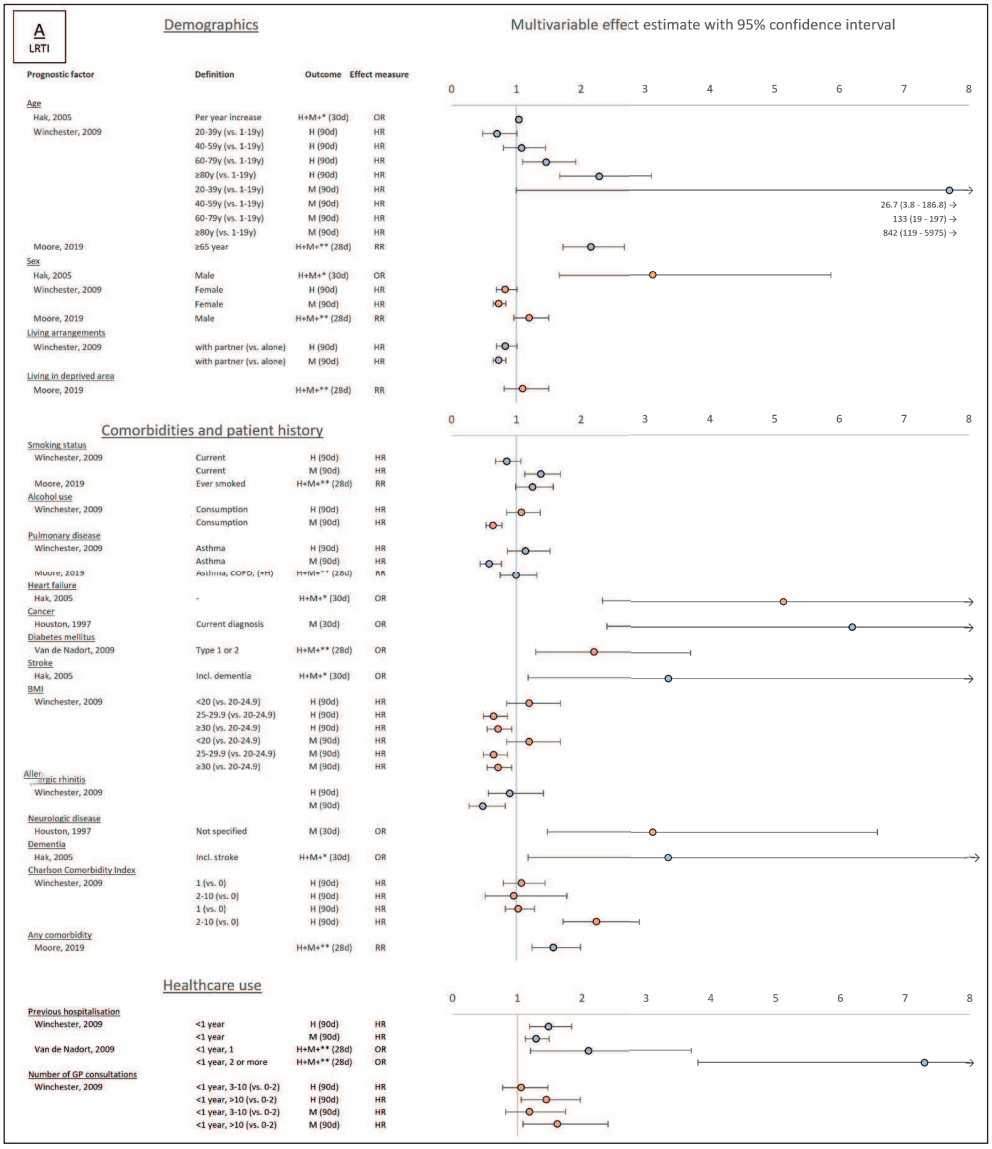


Figure S6: figure of all PF effect estimates based on multivariable effect estimates

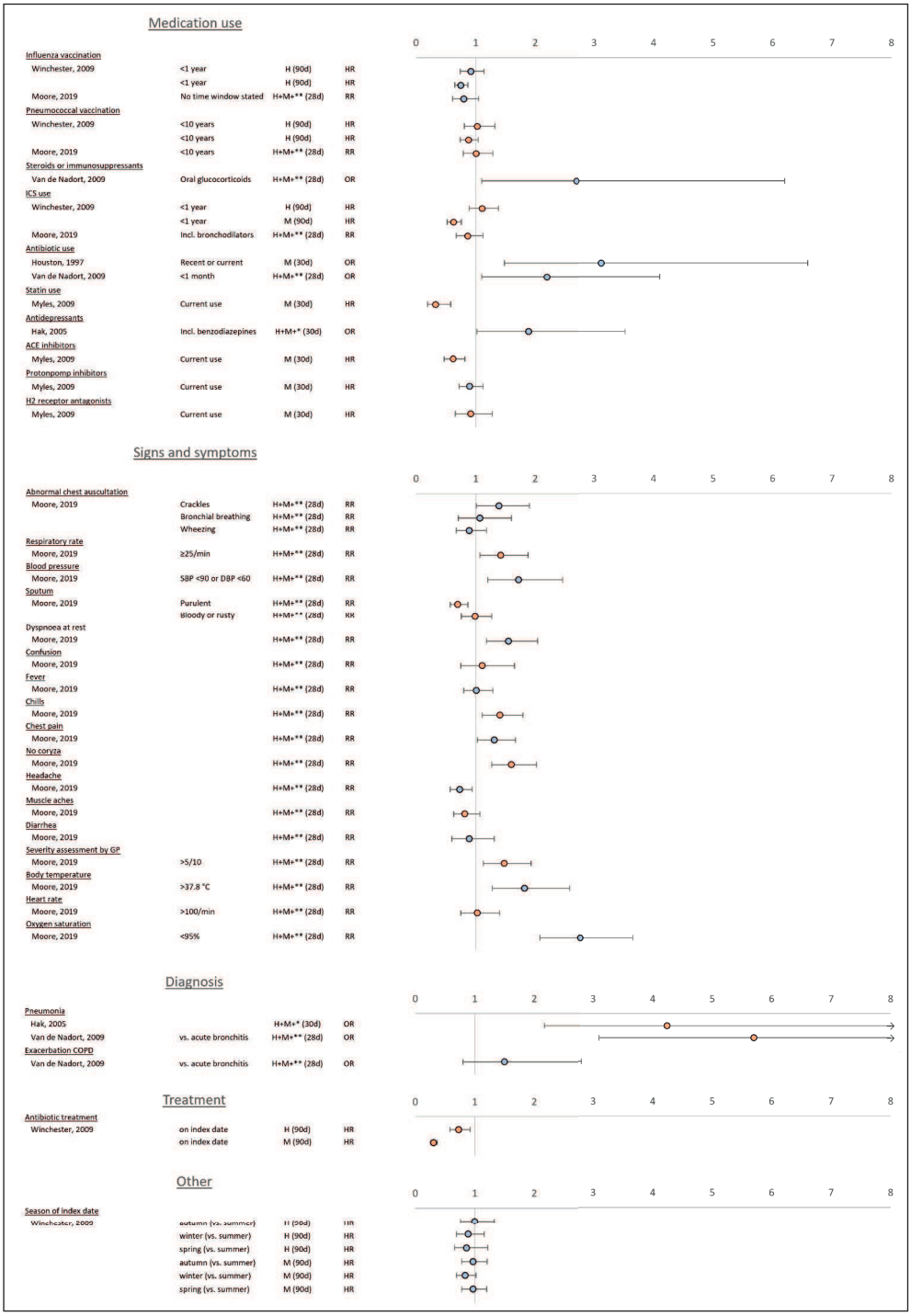


Figure S6: Continued

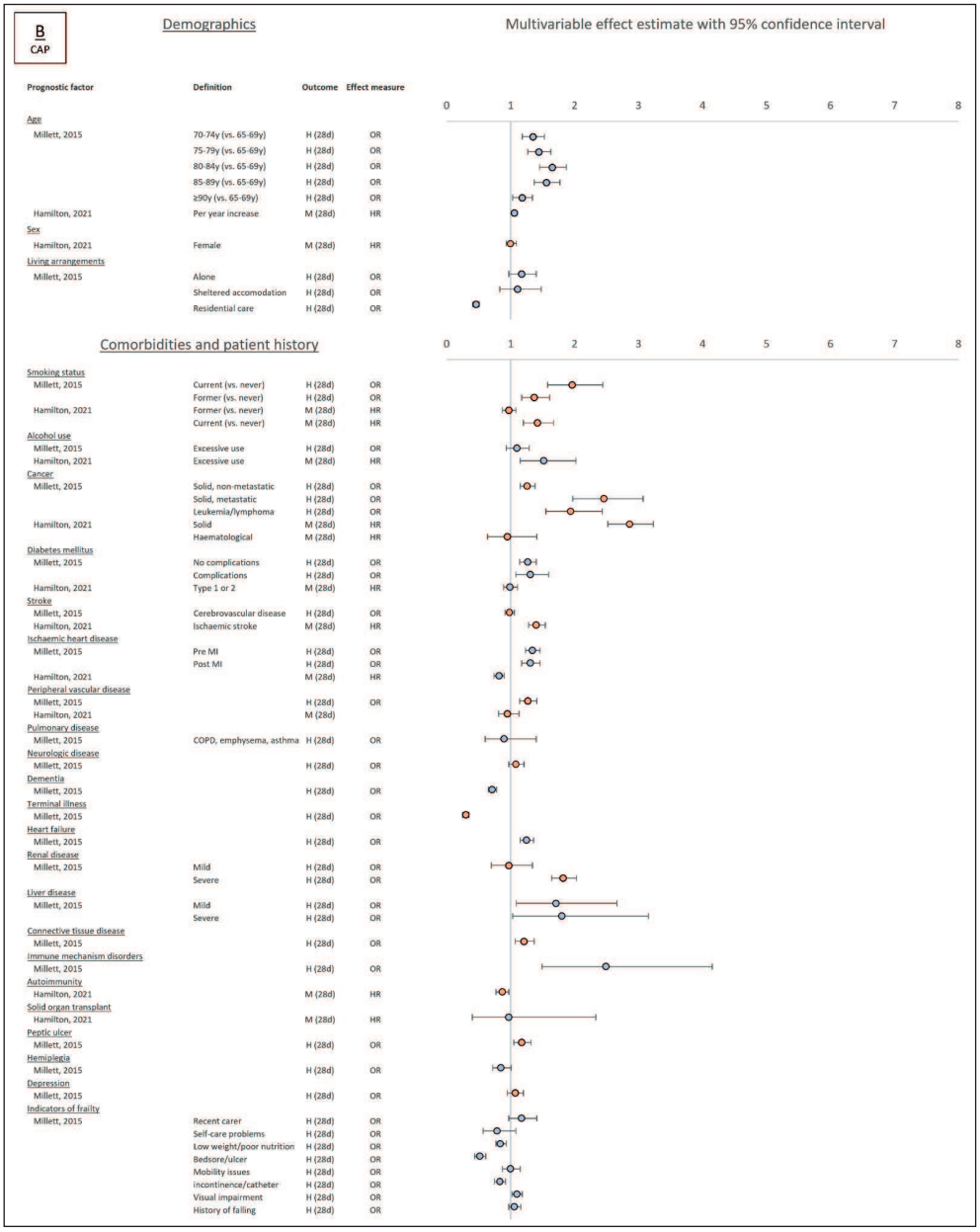


Figure S6: Continued

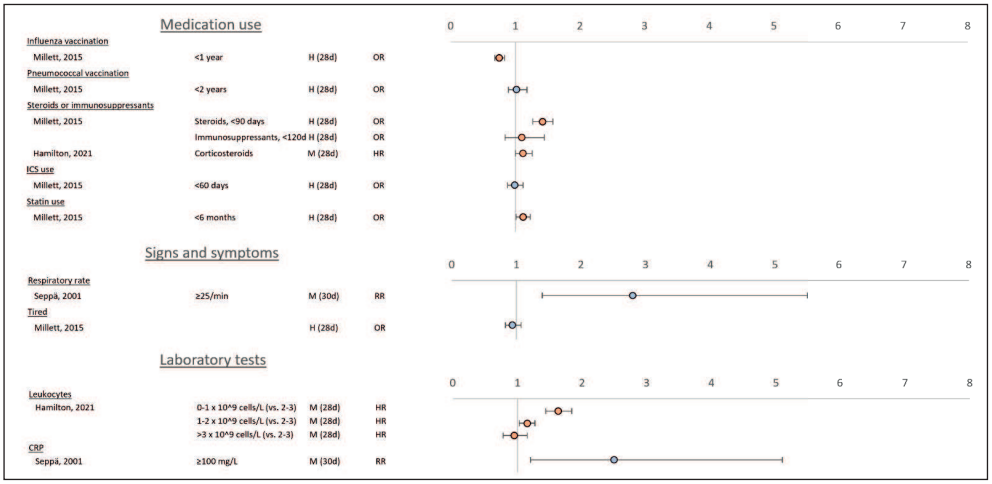


Figure S6: Continued

A. Multivariable analyses of prognostic factors from studies on LRTI patients. B. Multivariable analysis of prognostic factors from studies on pneumonia patients.

* Composite outcome also includes dysregulation of diabetes, stroke, heart failure, MI.

** Composite outcome also includes late onset pneumonia.

Abbreviations: LRTI, lower respiratory tract infection; H, hospitalisation; M, mortality; OR, odds ratio; RR, risk ratio; HR, hazard ratio; COPD, chronic obstructive pulmonary disease; BMI, body mass index; MI, myocardial infarction; GP, general practitioner; ICS, inhalation corticosteroids; ACE, angiotensin converting enzyme; SBP, systolic blood pressure; DBP, diastolic blood pressure; min, minute; L, liter; mg, milligram.

Table S7: Rating of the quality of evidence on promising prognostic factors based on the GRADE framework

		GRADE criteria						
Prognostic factor	**Study design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Publication bias**	**Upgrading factors**	**Quality of evidence**
Age	Cohort (high quality)	-2	0	0	0	0	0	Low
Sex	Cohort (high quality)	-2	0	-1	-1	0	0	Very low
Smoking status	Cohort (high quality)	-2	-1	-1	-1	0	0	Very low
Diabetes	Cohort (high quality)	-2	0	-1	0	0	0	Very low
Stroke	Cohort (high quality)	-2	-1	-1	-1	0	0	Very low
Cancer	Cohort (high quality)	-2	0	-1	0	0	0	Very low
Heart failure	Cohort (high quality)	-2	0	-1	0	0	0	Very low
Previous hospitalisation	Cohort (high quality)	-2	0	0	0	0	0	Low
Systemic corticosteroids	Cohort (high quality)	-2	0	-1	-1	0	0	Very low
Influenza vaccination	Cohort (high quality)	-2	0	0	-1	0	0	Very low
Recent antibiotic use	Cohort (high quality)	-2	0	0	-1	0	0	Very low
Respiratory rate	Cohort (high quality)	-2	0	-1	0	0	0	Very low
Diagnosis of pneumonia	Cohort (high quality)	-2	0	0	0	0	0	Low

	Certainty of evidence		
High	Moderate	Low	Very low
4	3	2	1

Application of the GRADE framework to rate the quality of evidence was based on prognostic research-specific guidance (Foroutan et al., 2020).

Abbreviations: GRADE, grading of recommendations, assessment, development, and evaluations

Table S8: Overview of prediction models included in the synthesis

Prediction model	Predictors included in model	Predictor weight
Bont, 2007 (new)	- Diagnosis (either acute bronchitis, COPD exacerbation, or pneumonia)	0, 2, or 4
	- Age ≥80 years	2
	- Congestive heart failure	1
	- Diabetes	2
	- Oral glucocorticoid use	3
	- Hospitalisations in previous year (0, 1, or ≥2)	0, 2, or 3
	- Antibiotic use in previous month	2
CRB(-65)	- Confusion	1
	- Respiratory rate ≥30/minute	1
	- Blood pressure (SBP ≤90mmHg or DBP ≤60mmHg)	1
	(- Age ≥65 years)	1
CURB(-65)	- Predictors of CRB(-65)	0-4
	- Blood urea nitrogen >7 mmol/L	1
PSI (stage 1)	- Age >50 years	Any versus none
	- Altered mental status	
	- Pulse >125/minute	
	- Respiratory rate >30/minute	
	- SBP <90mmHg	
	- Temperature <35°C or ≥40°C	
	- Neoplastic disease	
	- Congestive heart failure	
	- Cerebrovascular disease	
	- Renal disease	
	- Liver disease	
CORB-75 (new)	- Confusion	1
	- Peripheral oxygen saturation ≤90%	1
	- Respiratory rate ≥30/minute	1
	- Blood pressure (SBP ≤90mmHg or DBP ≤60mmHg)	1
	- Age ≥75 years	1
RISSC85 (new)	- Risk of poor outcome, grouped by country (A: Spain, B: Belgium, the Netherlands, Poland, UK, C: Germany)	A = 0, B/C = 1
	- Interference in daily activities (some versus severe)	1
	- Number of years stopped smoking (>45 years)	1
	- Severe sputum	1
	- Presence of crackles	1
	- Diastolic blood pressure (<85mmHg)	1

Table S8: Continued

Prediction model	Predictors included in model	Predictor weight
Moore, 2019 (new)	- Oxygen saturation <95%	1
	- Age >65 years	1
	- Blood pressure (SBP <90mmHg or DBP <60mmHg)	1
	- Temperature >37.8°C	1
	- Any comorbidity (cardiovascular, cerebrovascular or lung comorbidities)	1
	- No coryza	1
	- Severity assessment >5/10 by GP	1
	- Chest pain	1

Abbreviations: COPD, chronic obstructive pulmonary disease; SBP, systolic blood pressure; DBP, diastolic blood pressure; GP, general practitioner.

Table S9: Rating of the quality of evidence on prediction models based on the GRADE framework

		GRADE criteria						
Prediction model	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Upgrading factors	Quality of evidence
Bont, 2007	Cohort (high quality)	-2	0	0	-1	0	0	Very low
CRB	Cohort (high quality)	-2	0	-1	-1	0	0	Very low
CRB-65	Cohort (high quality)	-2	0	0	0	0	0	Low
CURB	Cohort (high quality)	-2	0	-1	-1	0	0	Very low
CURB-65	Cohort (high quality)	-2	0	-1	-1	0	0	Very low
PSI	Cohort (high quality)	-2	0	-1	-1	0	0	Very low
CORB-75	Cohort (high quality)	-2	0	-1	-1	0	0	Very low
RISCC85	Cohort (high quality)	-2	0	-1	-1	0	0	Very low
Moore, 2019	Cohort (high quality)	-2	0	-1	-1	0	0	Very low

	Certainty of evidence		
High	Moderate	Low	Very low
4	3	2	1

Application of the GRADE framework to rate the quality of evidence was based on prognostic research-specific guidance (Brozek et al., 2021).
Abbreviations: GRADE, grading of recommendations, assessment, development, and evaluations.

3

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
Tamara N. Platteel
Geert-Jan Geersing
Monika Hollander
Bert L.G.P. Dalmolen
Paul Little
Frans H. Rutten
Maarten van Smeden
Roderick P. Venekamp

Diagnostic and Prognostic Research. 2023;7(1):23.

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 ≥40 years with a GP-diagnosis of LRTI between 2016-2019 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.

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.

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 a poor prognosis of 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 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 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 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 previous year, and antibiotic use in 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 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 for 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 relevant individual disease trajectories. Details on the various data sources and their coverage is 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 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]

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

Characteristics			Type of data					Outcomes			Intended use
Data source	**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.

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) (Figure 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), 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.

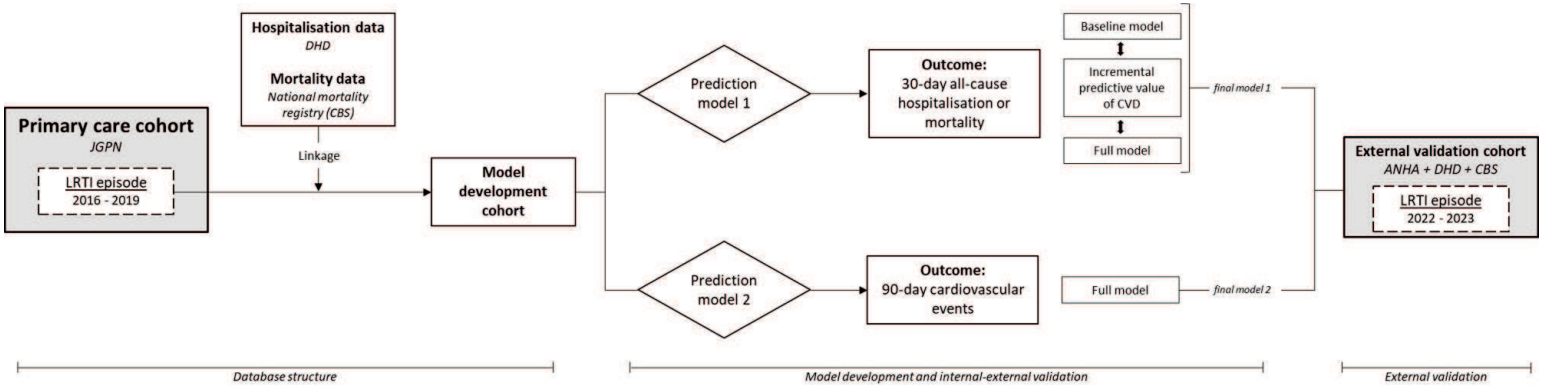


Figure 1. Graphical representation of database structure and anticipated prediction model development process.

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.

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 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 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 imputation with chained equations.[34]

Statistical analysis

Candidate predictor selection will be based on existing literature, clinical expertise, data availability, and 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 difference in c-statistics calculated by bootstrapping, change in 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 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 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 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 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 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 guideline 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 risk of poor prognosis in primary care LRTI patients, which may ultimately allow for early detection and prevention of deterioration.

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, 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,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, et al. Antibiotic prescription strategies and adverse outcome for uncomplicated lower respiratory tract infections: prospective cough complication cohort (3C) study. *BMJ*. 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, 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, 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, et al. 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, et al. 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, 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.
11. Smeeth L, Thomas SL, Hall AJ, et al. 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, Calvieri 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 Card*. 2021;**18**(9):666–82.
14. Macias AE, McElhaney JE, Chaves SS, et al. The disease burden of influenza beyond respiratory illness. *Vaccine*. 2021;**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, 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, 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, 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 CRB-65 severity assessment tool. *Arch Intern Med*. 2008;**168**(13):1465–8.
20. Ochoa-Gondar O, Vila-Corcoles A, Rodriguez-Blanco T, 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, 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, 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, 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 Family Doctors (WONCA). *ICPC-2: International Classification of Primary Care*. Oxford: Oxford University Press; 1998.
25. World Health Organization (WHO) Collaborating Centre for Drug Statistics Methodology. *ATC Classification Index with DDDs*. Oslo: WHO Collaborating Centre for Drug Statistics Methodology; 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)*. Available from: https://www.vumc.nl/research/overzicht/academisch-netwerk-huisartsgeneeskunde-amsterdam-umc-anha.html.
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: pmsampsize: calculates the minimum sample size required for developing a multivariable prediction model. Version 1.1.2. 2022. Available from: https://CRAN.R-project.org/package=pmsampsize.
30. R Core Team. *R: A Language and Environment for Statistical Computing* [Internet]. R Foundation for Statistical Computing, Vienna, Austria; 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**(2):414.
33. Linden M van der, Westert G, Bakker D de, Schellevis F. *Tweede nationale studie naar ziekten en verrichtingen in de huisartspraktijk: klachten en aandoeningen in de bevolking en in de huisartspraktijk*. Utrecht: NIVEL, 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;**32**(8):1461–1477.
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, et al. Validation of the CORB75 (confusion, oxygen saturation, respiratory rate, blood pressure, and age ≥75 years) as a simpler pneumonia severity rule. *Infection*. 2013;**42**(2):371–8.
38. Moore M, Stuart B, Lown M, 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, 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, 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**(1):187–96.
42. Davidson JA, Banerjee A, Smeeth L, 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**(3):195–210.
44. Wells BJ, Nowacki AS, Chagin K, Kattan MW. Strategies for handling missing data in electronic health record derived data. *EGEMS*. 2013;**1**(3):1035.
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):1167–1176.
46. Metlay JP, Waterer GW, Long AC, et al. Diagnosis and treatment of adults with community-acquired pneumonia. *Am J Respir Crit Care Med*. 2019;**200**(7):E45–67.

SUPPLEMENTARY MATERIALS FOR CHAPTER 3

Additional file 2 (TRIPOD checklist) is not printed in this thesis but is available online.

Additional file 1. Overview of variables considered as candidate predictors

Parameter	Underlying registrations, if applicable	Codes, if applicable		Data source
		ICPC	ATC	
Demographics				
Age (years)				JGPN
Sex (male, female)				JGPN
Frailty index	Algorithm based on ICPC codes (Drubbel et al., 2013)			JGPN
Use of medical care				
Number of GP consultations ≤12 months				JGPN
Hospitalisation ≤12 months				DHD
Vaccination status				
Pneumococcal conjugate vaccine	Immunisation/preventive medicine of airways	R44	J07BB02 (<5 years)	JGPN
Influenza vaccine	Immunisation/preventive medicine of airways	R44	J07AL01 (<1 year)	JGPN
Comorbidities: CVD				
Aortic aneurysm	Aortic aneurysm	K99.01		JGPN
Atrial fibrillation	Atrial fibrillation/flutter	K78		JGPN
Cardiomyopathy	Cardiomyopathy	K84.03		JGPN
Cerebrovascular disease	CVA TIA	K90 K89		JGPN
Coronary artery disease	Angina pectoris Other chronic/ischemic heart disease	K74 K76		JGPN
Heart failure	Decompensatio cordis	K77		JGPN
Myocardial infarction	Acute myocardial infarction	K75		JGPN
Other arrhythmias	Paroxysmal tachycardia Sick sinus syndrome Wolff-Parkinson-White syndrome Atrioventricular block Long QT-syndrome Presence of pacemaker/ICD	K79 K80.03 K84.01 K84.02 K84.07 A89.01		JGPN
Pregnancy related CVD	Deep vein thrombosis in pregnancy Postpartum thrombosis	W77.03 W99.03		JGPN
Thrombo-embolic disease	Pulmonary embolism Thrombophlebitis/ phlebothrombosis	K93 K94		JGPN
Valve disease	Acute rheumatic fever Non-rheumatic valve disease	K71 K83		JGPN

Additional file 1. Continued

Parameter	Underlying registrations, if applicable	Codes, if applicable		Data source
		ICPC	ATC	
Other comorbidities				
Smoking status	Algorithm for text mining (Boer et al., 2022)			JGPN
Family history of CVD	Family history of CVD	A29.01		JGPN
Hypertension	High blood pressure	K85		JGPN
	Primary hypertension with organ damage	K86		
	Secondary hypertension/organ damage	K87		
Atherosclerosis	Atherosclerosis	K91		JGPN
Intermittent claudication	Intermittent claudication	K92.01		JGPN
Rheumatoid arthritis	Rheumatoid arthritis	L88.01		JGPN
History of pneumonia	Pneumonia	R81		JGPN
COPD	Emphysema/COPD	R95		JGPN
	Chronic bronchitis/bronchiectasis	R91		
Asthma	Asthma	R96		JGPN
Allergic rhinitis	Allergic rhinitis	R97		JGPN
Adipositas	Adipositas	T82		JGPN
	Obesity	T83		
Diabetes mellitus	Diabetes mellitus type 1 and 2	T90		JGPN
Disorders of fat metabolism	Disorders of fat metabolism	T93		JGPN
Renal dysfunction	Renal dysfunction/insufficiency	U99.01		JGPN
Pregnancy related CVD risk factors	Pre-eclampsia	W81		JGPN
	Gestational diabetes	W84.02		
Premature menopause	Premature menopause	T99.07		JGPN
Coagulation disorders	Purpura/coagulation disorder/abnormal thrombocytes	B83		JGPN
PCOS	PCOS	T99.06		JGPN
Dementia	Cognitive disorders	P20		JGPN
	Dementia	P70		
Thyroid disease	Hyperthyroidism/thyrotoxicosis	T85		JGPN
	Hypothyroidism/myxoedema	T86		
IBD	Ulcerative colitis/Crohn's disease	D94		JGPN
HIV infection	HIV-infection	B90		JGPN
Viral hepatitis	Viral hepatitis	D72		JGPN
Cirrhosis	Cirrhosis/other liver disease	D97		JGPN
Skin ulcer	Ulcer cruris/decubitus/chronic ulcer	S97		JGPN
Gastrointestinal ulcer	Ulcer duodeni	D85		JGPN
	Ulcer pepticum	D86		
Paralysis	Paralysis	N18		JGPN
Hip fracture	Fratcure of collum femoris	L75.01		JGPN

Additional file 1. Continued

Parameter	Underlying registrations, if applicable	Codes, if applicable		Data source
		ICPC	ATC	
Oncology				
Hematologic cancer	Hodgkin lymphoma	B72.01		JGPN
	Non-Hodgkin lymphoma	B72.02		
	Leukaemia	B73		
	Hematologic/lymphatic malignancy NOS	B74		
	Benign hematologic/lymphatic Neoplasm	B75		
Digestive tract cancer	Stomach malignancy	D74		JGPN
	Colon/rectum malignancy	D75		
	Pancreatic malignancy	D76		
	Digestive tract malignancies NOS	D77		
Pulmonary cancer	Malignancy of bronchus/lung	R84		JGPN
	Other airway malignancy	R85		
Skin cancer	Malignancy skin/subcutis	S77		JGPN
Urological cancer	Malignancy of kidney	U75		JGPN
	Malignancy of bladder	U76		
	Other malignancy of urinary tract	U77		
	Malignancy prostate	Y77		
	Other urologic malignancy	Y78		
Gynaecological cancer	Malignancy of cervix uteri	X75		JGPN
	Malignancy of breast (female)	X76		
	Other gynaecological malignancies	X77		
Other malignancies	Malignancy of unknown primary origin	A79		JGPN
	Malignancy of eye/adnexa	F74.01		
	Malignancy of ear	H75.01		
	Cardiovascular malignancy	K72.01		
	Musculoskeletal malignancy	L71.01		
	Neurologic malignancy	N74		
	Malignancy of thyroid	T71		
Recent laboratory tests and measurements (timeframe in months prior to LRTI-episode)				
Body mass index (kg/m^2)	<12 months			JGPN
eGFR (ml/min/1,73 m^2)	<12 months			JGPN
Haemoglobin (mmol/l)	<3 months			JGPN
Leukocytes (x10^9/L)	<6 months			JGPN
BNP or NT-proBNP (pmol/l)	<6 months			JGPN
HbA1$_c$ (mmol/mole)	<6 months			JGPN
TSH (mU/l)	<3 months			JGPN
D-dimer (mg/l)	<3 months			JGPN
LDL (mmol/l)	<6 months			JGPN
Measurements linked to LRTI-related episode				
Oxygen saturation (%)				JGPN
Respiratory rate (/min)				JGPN

Additional file 1. Continued

Parameter	Underlying registrations, if applicable	Codes, if applicable		Data source
		ICPC	ATC	
Heart rate (/min)				JGPN
Systolic blood pressure (mmHg)				JGPN
Diastolic blood pressure (mmHg)				JGPN
Body temperature (°C)				JGPN
C-reactive protein (mg/ml)				JGPN
Chronic use of medication[a]				
Recent antibiotic use	Within 1 month prior to LRTI episode		J01	JGPN
Platelet aggregation inhibitors	platelet aggregation inhibitors		B01AC	JGPN
Other anticoagulants	Vitamin K antagonists Dalteparin Enoxaparin Nadroparin Tinzaparin Dabigatran Direct factor Xa inhibitors Fondaparinux		B01AA B01AB04 B01AB05 B01AB06 B01AB10 B01AE07 B01AF B01AX05	JGPN
Diuretics	Diuretics		C03	JGPN
Beta blockers	Beta blockers		C07A	JGPN
Cardioselective beta blockers	Cardioselective beta blockers		C07AB	JGPN
Calcium channel blockers	Calcium channel blockers		C08	JGPN
ACE-inhibitors	ACE-inhibitors, plain ACE-inhibitors, combinations		C09A C09B	JGPN
ARBs	Angiotensin II receptor blockers, plain Angiotensin II receptor blockers, combinations		C09C C09D	JGPN
Nitrates	Isosorbidemononitrate Isosorbidedinitrate Nitrogycerin		C01DA14 C01DA08 C01DA02	JGPN
Lipid-lowering medication	Lipid-lowering agents		C10A	JGPN
Digoxin	Digoxin		C01AA05	JGPN
Flecainide	Flecainide		C01BC04	JGPN
Immunosuppressive drugs	Systemic glucocorticoids Predniso(lo)n ≥7.5mg/day Immunosuppressants		H02AB L04A	JGPN
Benzodiazepines	Benzodiazepines		N05CD + N05CF	JGPN
Antidepressants	Antidepressants		N06A	JGPN

Additional file 1. Continued

Parameter	Underlying registrations, if applicable	Codes, if applicable		Data source
		ICPC	ATC	
Insulin	Insulin and analogs		A10A (- A10AF)	JGPN
Inhalation medication	Long acting beta antagonists (LABA):			JGPN
	Formoterol		R03AC13	
	Indacaterol		R03AC18	
	Olodaterol		R03AC19	
	Salmeterol		R03AC12	
	Long azcting muscarine antagonists (LAMA):			
	Aclidinium		R03BB05	
	Glycopyrronium		R03BB06	
	Tiotropium		R03BB04	
	Umeclidinium		R03BB07	
	LABA+LAMA:			
	Aclidinium/formoterol		R03AL05	
	Glycopyrronium/formoterol		R03AL07	
	Indacaterol/glycopyrronium		R03AL04	
	Tiotropium/olodaterol		R03AL06	
	Umeclidinium/vilanterol		R03AL03	
	Inhalation corticosteroids (ICS):			
	Beclomethasone		R03BA01	
	Budesonide		R03BA02	
	Fluticasone		R03BA05	
	ICS+LABA:			
	Formoterol/beclomethasone		R03AK08	
	Formoterol/budesonide Salmeterol/		R03AK07	
	fluticasone propionate		R03AK06	
	Vilanterol/fluticasone furoate		R03AK10	
	ICS+LABA+LAMA			
	Beclomethasone/formoterol/		R03AL09	
	glycopyrronium		R03AL08	
	Fluticasone/umeclidinium/vilanterol		R03AL11	
	Formoterol/glycopyrronium/budesonide			

Abbreviations: CVD, cardiovascular disease; ICPC, International Classification of Primary Care; ATC, anatomical therapeutic chemical; JGPN, Julius General Practitioners' Network; GP, general practitioner; DHD, Dutch Hospital Data; COVID-19, coronavirus disease 2019; CIMS, COVID-vaccination Information and Monitoring System; COPD, chronic obstructive pulmonary disease; PCOS, polycystic ovarium syndrome; NOS, not otherwise specified; IBD, inflammatory bowel disease; HIV, human immunodeficiency virus; LRTI, lower respiratory tract infection; kg, kilogram; m, meter; eGFR, estimated glomerular filtration rate; ml, millilitre; min, minute; mmol, millimole; l, litre; BNP, B-type natriuretic peptide; NT-proBNP, N-terminal (NT)-pro hormone BNP; pmol, picomole; $HbA1_c$, haemoglobin A1c; mU, milliunits; mg, milligram; LDL, low density lipoproteins; mmHg, millimeters of mercury; °C, degrees celcius.

[a] Medication was selected only when commonly used in a primary care setting and is based upon guidelines issued by the Dutch College of General Practitioners and the Dutch pharmacotherapeutic compass.

4

Incomplete and possibly selective recording of signs, symptoms and measurements in free text fields of primary care electronic health records of adults with lower respiratory tract infections

Merijn H. Rijk
Tamara N. Platteel
Marissa M.M. Mulder
Geert-Jan Geersing
Frans H. Rutten
Maarten van Smeden
Roderick P. Venekamp
Artuur M. Leeuwenberg

Journal of Clinical Epidemiology. 2024;**166**:111240.

Abstract

Objective To assess the completeness of recording of relevant signs, symptoms, and measurements in Dutch free text fields of primary care electronic health records (EHR) of adults with lower respiratory tract infections (LRTI).

Study design and setting Retrospective cohort study embedded in a prediction modelling project using routine health care data of the Julius General Practitioners' Network of adult patients with LRTIs. Free text fields of 1,000 primary care consultations of LRTI episodes between 2016-2019 were manually annotated to retrieve data on the recording of sixteen relevant signs, symptoms, and measurements.

Results For 12/16 (75%) of the relevant signs, symptoms, and measurements, more than 50% of the values was not recorded. The patterns of recorded values indicated selective recording of positive or abnormal values. Recording rates varied across consultation type (physical consultation vs. home visit), diagnosis (acute bronchitis vs. pneumonia), antibiotic prescription issued (yes vs. no), and between practices.

Conclusion In EHR of primary care LRTI patients, recording of signs, symptoms, and measurements in free text fields is incomplete and possibly selective. When using free text data in EHR-based research, careful consideration of its recording patterns and appropriate missing data handling techniques is therefore required.

INTRODUCTION

Electronic health records (EHR) hold information on large numbers of individual patients and contain structured data on health indicators such as medical history, coded disease episodes, presenting signs or symptoms, and treatments. Although EHR data are not primarily collected for research purposes, its potential for conducting research is increasingly recognised and the availability of longitudinal data on a large number of patients enables researchers to study relatively rare events.[1,2] Currently, the majority of EHR-based research is focussed on structured (often limited) data, since manual extraction of information from unstructured free text fields is time- and resource-consuming.[2,3] However, advances in natural language processing methods increasingly allow for automated retrieval of unstructured information from free text fields of the EHR, in which more detailed information on signs, symptoms, and measurements is often documented.[2–4] This increasing availability of unstructured data provides relevant opportunities for EHR-based research, for example on the prognosis of acute lower respiratory tract infections (LRTI) in primary care. Identifying LRTI patients at risk of hospitalisation or mortality is challenging and existing prediction models that can assist general practitioners (GP) in risk stratification – of which CRB-65 (confusion, respiratory rate, blood pressure and age ≥65 years) is the most promising – have not been properly validated.[5] Predictors of the CRB-65, as well as other potential relevant signs, symptoms, and measurements, are mainly recorded in free text fields.[3] However, prior to including free text variables in primary care EHR-based research, it is instrumental to gain insight in the type and quality of information captured in free text primary care EHR data and to assess the completeness of this information. Yet, this has been rarely done previously.

In this study, we therefore aimed to assess the completeness of recording of signs, symptoms, and measurements in free text fields of EHRs of adult patients presenting to primary care with an LRTI. This provides insight on the recording in routine GP practice of such variables and informs on the quality and potential challenges in the use of free text data in primary care EHR-based research.

STUDY DESIGN AND SETTING

Design and participants

For this study, we used primary care routine health care data from the Julius General Practitioners' Network (JGPN).[6] The JGPN contains data on patient demographics, encoded diseases and prescriptions (using the International Classification of Primary Care (ICPC) and Anatomical Therapeutic Chemical (ATC) codes), coded measurements of vital parameters, laboratory results, and free text fields of consultations, and covers approximately 450,000 patients from both urban and rural general practices in the region of Utrecht, the

Netherlands. This study is embedded in a project aimed at developing EHR-based prediction models for adverse outcomes in adult patients presenting with LRTI to primary care, which has been reported on elsewhere.[7] In short, the cohort comprises of patients aged 40 years and older consulting their GP with an LRTI episode (defined as ICPC registration of either acute bronchitis (R78) or pneumonia (R81)) between 1 January 2016 and 31 December 2019. An episode starts at the day of the first LRTI-related consultation (i.e. index consultation), and consecutive episodes within individual patients are separated by a period of at least 28 days without LRTI-related consultations. For this study, a non-stratified random sample of 1,000 episodes from our JGPN cohort was drawn. In case of multiple episodes per individual patient, only the first episode was included and we only included episodes of which the index consultation was a face-to-face appointment, since these are likely to contain information on physical examination and vital and laboratory measurements.

Data collection

For all 1,000 episodes, information on patient demographics (age, sex), comorbidities (pulmonary diseases, diabetes, heart failure, and history of malignancy or cerebrovascular accident), smoking status and medication use (immunosuppressants, inhalation medication) was collected at the day of index consultation. Smoking status was automatically extracted using a free text algorithm that has been developed within JGPN.[8] Using free text fields and structured input on measurements, we extracted data on signs, symptoms and measurements recorded at index consultations which may be of value for predicting poor outcome in primary care LRTI patients.[5] These include patient-reported symptoms (cough, shortness of breath, sputum, chest pain, chills, fever, and confusion), GP-reported signs (ill appearance, confusion, crackles on lung auscultation), and vital and laboratory measurements (blood pressure, body temperature, respiratory rate, heart rate, oxygen saturation, and C-reactive protein (CRP)). Recording of these variables was manually retrieved from free text fields of index consultations by one author (AM). During this process, a second author (MR) manually cross-checked the first 100 annotated episodes (10% of the total sample) to assure correct coding of free text derived variables in the remaining sample, with a high inter-rater reliability (Cohen's Kappa 0.98 (95% confidence interval (CI) 0.97 to 0.99), indicating almost perfect agreement). When recorded, the absence or presence of signs and symptoms (e.g. cough yes/no) and the values of measurements were documented. Antibiotic prescription at index consultation was extracted using ATC codes of LRTI-related prescriptions including amoxicillin, amoxicillin/clavulanate, azithromycin, clarithromycin, doxycycline, and erythromycin.[9]

Statistical analysis

We performed an exploratory analysis of the extracted data on signs, symptoms, and measurements, in which the proportion of recording, the distribution of positive and negative values of recorded signs and symptoms, and the distribution of recorded mea-

surements were descriptively summarised. We explored differences in the recording of signs, symptoms, and measurements across the strata of consultation type, diagnosis, antibiotic treatment at index consultation, and age, and we assessed heterogeneity in recording patterns between general practices.

Proportions of registered signs, symptoms, and measurements were compared using a two-sample Z-test, and mean number of recordings were compared using the Wilcoxon rank-sum test. Correlation between continuous variables was assessed using Pearson's correlation coefficient. All analyses were performed in R version 4.2.2.[10]

RESULTS

Characteristics of cohort

Characteristics of the cohort are summarised in Table 1. The mean age of the patients was 63.0 (standard deviation (SD) 13.8) years, and 56.0% were female. Most patients (56.3%) were diagnosed with pneumonia and 31.9% were treated with antibiotics at index consultation.

Table 1. Baseline characteristics of cohort

Characteristic	Total = 1000
Mean age (SD)	63 (13.8)
Female sex	560 (56.0%)
Year	
2016	343 (34.3%)
2017	241 (24.1%)
2018	246 (24.6%)
2019	170 (17.0%)
Consultation type	
Physical consultation	850 (85.0%)
Home visit	150 (15.0%)
Diagnosis	
Acute bronchitis	437 (43.7%)
Pneumonia	563 (56.3%)
Smoking status	
Current	266 (26.6%)
Former	209 (20.9%)
Never	202 (20.2%)
No information	323 (32.3%)
Comorbidities	
COPD	102 (10.2%)
Asthma	121 (12.1%)
Malignancy †	78 (7.8%)
Diabetes	52 (5.2%)
Heart failure	37 (3.7%)
CVA	38 (3.8%)

Table 1. Continued

Characteristic	Total = 1000
Medication use	
Inhalation medication [§]	45 (4.5%)
Immunosuppressants or systemic corticosteroids	9 (0.9%)
Antibiotic prescription (at index consultation) [‡]	319 (31.9%)

[†] Excluding malignancies of the skin.

[§] Includes long-acting bronchodilators, corticosteroid inhalers, and combinations.

[‡] Includes only antibiotic prescriptions that are appropriate for treating lower respiratory tract infections based on primary care guidelines, i.e. amoxicillin, amoxicillin/clavulanate, azithromycin, clarithromycin, doxycycline, and erythromycin.

Abbreviations: SD, standard deviation; COPD, chronic obstructive pulmonary disease; CVA, cerebrovascular accident.

Recording of signs, symptoms, and measurements

On average, 3.6 signs and symptoms (median: 4) were recorded per patient (Figure 1). Lung auscultation (86.1%) and cough (76.6%) were recorded most frequently whereas information on chills (4.4%) and confusion (3.5%) was recorded least frequently (Table 2). When recorded, patient reported symptoms were mostly present – most notably cough (98.4%), chills (95.5%), sputum (91.1%) and chest pain (78.3%) – indicating selective recording.

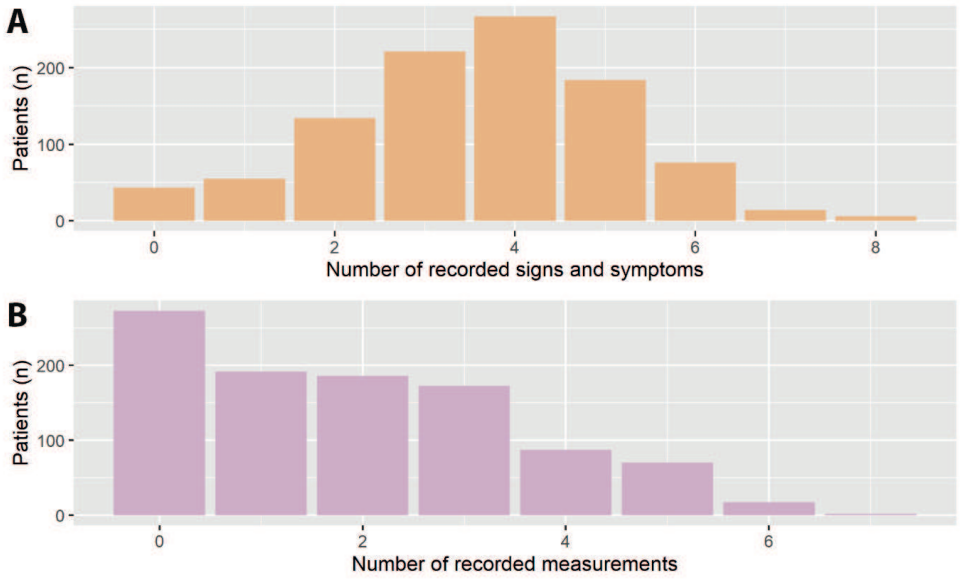


Figure 1. Number of (**A**) signs and symptoms and (**B**) measurements recorded per individual patient.

An average number of 1.9 measurements (median: 2) was recorded per patient (Figure 1). For all measurements more than 50% of the values were not recorded (Table 2). Body temperature (45.4%), oxygen saturation (45.1%), and heart rate (40.0%) were recorded most frequently whereas respiratory rate (5.7%) was infrequently recorded. The distributions of recorded measurement values – especially respiratory rate, oxygen saturation, and CRP – reveal a substantial number of abnormal values, potentially indicating selective recording in those more severely affected (Figure 2).

Table 2. Recording and distribution of signs, symptoms, and measurements in free text fields of electronic medical records

	Signs and symptoms			
	Recording (%)			**Positive if recorded (%)**
	Pneumonia	**Acute bronchitis**	**Overall**	
Patient reported				
Cough	71.9	82.6	76.6	98.4
Fever	57.0	50.3	54.1	57.7
Shortness of breath	54.4	51.7	53.2	70.9
Sputum	26.1	30.7	28.1	91.1
Chest pain	22.9	13.5	18.8	78.7
Chills	5.7	2.7	4.4	95.5
Patient or GP reported				
Confusion	5.3	1.1	3.5	17.1
GP reported				
Crackles (auscultation)	83.7	89.2	86.1	26.6
Ill appearance	39.1	24.7	32.8	39.3
	Measurements			
	Recording (%)			**Median value (IQR)**
	Pneumonia	**Acute bronchitis**	**Overall**	
Body temperature	52.0	36.8	45.4	37.2 (36.8 – 38.0) ^{0}C
Oxygen saturation	51.0	37.5	45.1	97 (95 – 98)%
Heart rate	46.9	31.1	40.0	88 (76 – 100) beats/min
CRP	20.1	16.5	18.5	39 (10 – 94) ml/mL
Systolic BP	21.7	12.1	17.5	132 (120 – 150) mmHg
Diastolic BP	21.7	12.1	17.5	80 (70 – 84.5) mmHg
Respiratory rate	6.4	4.8	5.7	18 (16 – 20) breaths/min

Abbreviations: GP, general practitioner; CRP, C-reactive protein; BP, blood pressure.

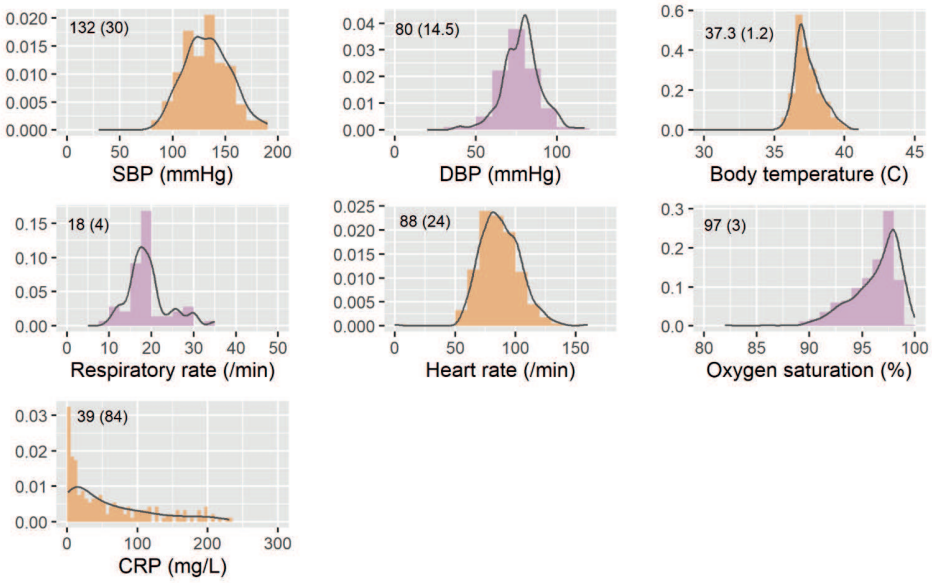


Figure 2. Distribution of recorded measurements with density curve and median (interquartile range)

Abbreviations: SBP, systolic blood pressure; DBP, diastolic blood pressure; C, Celsius; min, minute; CRP, c-reactive protein.

Recording patterns across subgroups

The mean number of recorded signs and symptoms was significantly higher in patients with ICPC registration of pneumonia compared to acute bronchitis (3.7 vs. 3.5, p=0.006) and in those receiving immediate antibiotic treatment compared to those who did not (3.9 vs. 3.4, p<0.001) (Table 3). On average, the number of recorded measurements was significantly higher at home visits compared to physical consultations (2.7 vs. 1.8, p<0.001), in patients with pneumonia compared to acute bronchitis (2.2 vs. 1.5, p<0.001), and in those receiving immediate antibiotic treatment compared to those who did not (2.2 vs. 1.8, p<0.001). There was a significant difference in the recording of 9/15 variables between consultation type (home visit vs. physical consultation), 12/15 variables between diagnosis (pneumonia vs. acute bronchitis), and 8/15 variables between antibiotic prescription issued at index consultation (yes vs. no) (Table S1). Shortness of breath was the only variable that did not significantly differ across all three strata. Age was weakly negatively associated with the number of registered signs and symptoms included in our study (Pearson's correlation coefficient (r) -0.07, 95% CI -0.13 to -0.01) but showed no significant association with the number of registered measurements (r 0.06, 95% CI 0.00 to 0.12) (Figure 3). The number of recorded signs and symptoms was positively associated with the number of recorded measurements (r 0.30, 95% CI 0.24 to 0.35).

Table 3. Differences in mean number of recordings between LRTI episodes stratified according to consultation type, diagnosis, and immediate antibiotic prescription.

	Consultation type		
	Mean number of recordings		
	Home visit	**Physical consultation**	**p value †**
Signs and symptoms	3.7	3.5	0.106
Measurements	2.7	1.8	<0.001
	Diagnosis		
	Mean number of recordings		
	Pneumonia	**Acute bronchitis**	**p value †**
Signs and symptoms	3.7	3.5	0.006
Measurements	2.2	1.5	<0.001
	Treatment		
	Mean number of recordings		
	Immediate AB	**No immediate AB**	**p value †**
Signs and symptoms	3.9	3.4	<0.001
Measurements	2.2	1.8	<0.001

† Based on Wilcoxon rank sum test.

Abbreviations: LRTI, lower respiratory tract infection; AB, antibiotics.

Summarising the recording patterns of different primary care practices, a mean number of 3.6 (SD 0.7, range 1.8 to 5.3) signs and symptoms and 1.9 (SD 0.8, range 0.0 to 3.8) measurements were recorded per patient per practice. On a primary care practice level, the mean number of recorded signs and symptoms per patient was positively associated with the mean number of recorded measurements per patient on a primary care practice level (r 0.44, 95% CI 0.25 to 0.60) (Figure 4).

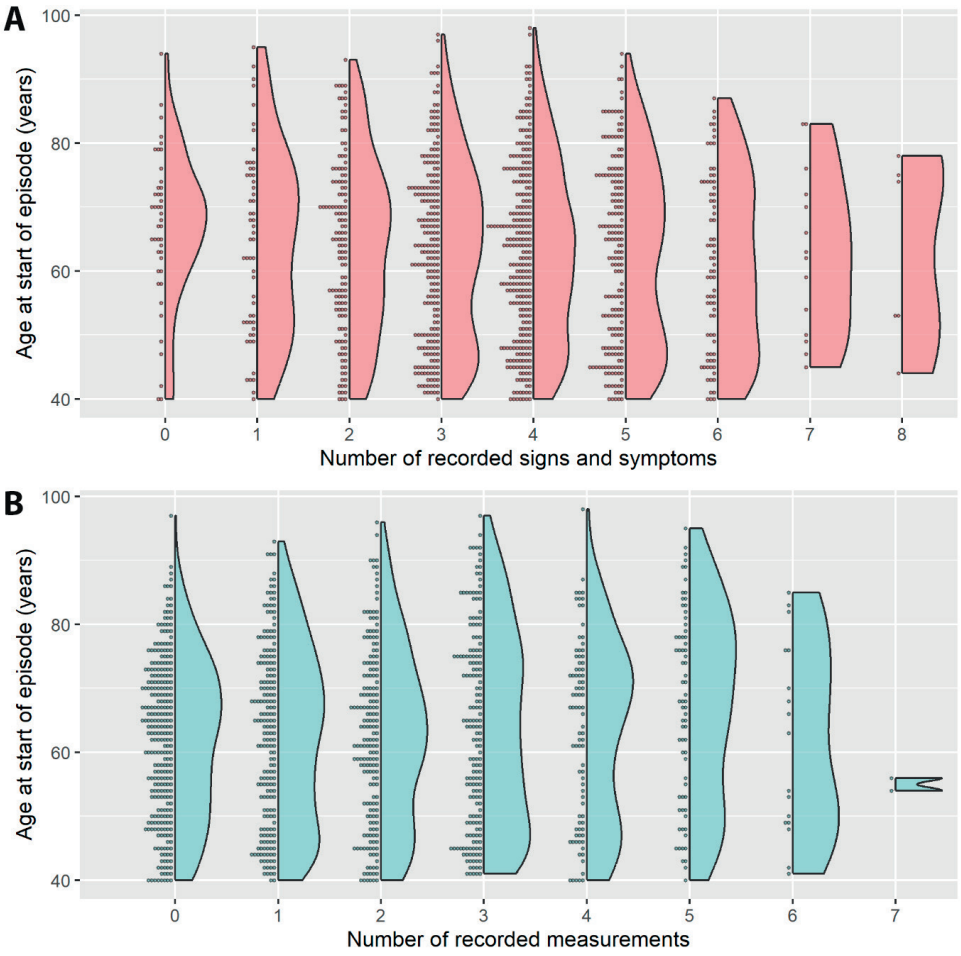


Figure 3. Association between age and (**A**) number of signs and symptoms, and (**B**) number of measurements recorded

Association between age and number of recorded signs and symptoms (**A**) and between age and number of recorded measurements (**B**) visualised in half-violin half-dot plot. Dots represent individual data points (i.e. episodes) and violin plots visualise the distribution of age per number of recordings.

Pearson's correlation coefficient (r) for association between age and the number of recordings:
A: r -0.07, 95% confidence interval -0.13 to -0.01
B: r 0.06, 95% confidence interval 0.00 to 0.12

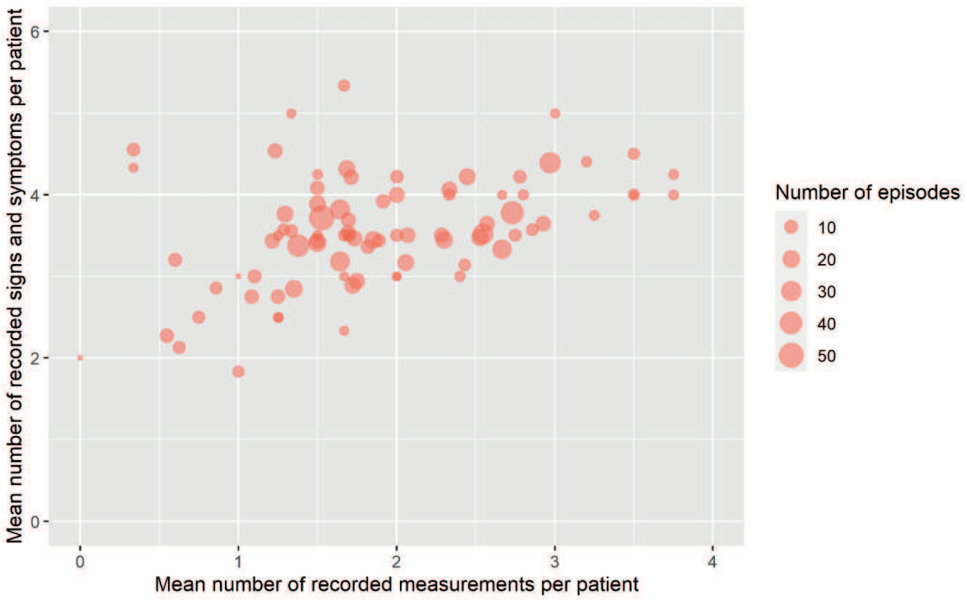


Figure 4. Association between mean number of signs and symptoms and mean number of measurements recorded, on a general practice level

Each data point represents a unique general practice in the Julius General Practitioners' Network. The size of the data points is determined by the number of episodes contributed to the study cohort.

Pearson's correlation coefficient (r) 0.44, 95% confidence interval 0.25 to 0.60

DISCUSSION

Main findings

In this study, we aimed to assess the completeness of recording of relevant signs, symptoms, and measurements in free text EHR data of primary care LRTI patients. Our most notable finding is the substantial amount of non-recorded values. Apart from incomplete recording, our study reveals several issues indicating selective recording of these variables, potentially posing a challenge when using these data in EHR-based research. First, the distribution of recorded values suggests that signs, symptoms, and measurements are more likely to be recorded in case of positive or abnormal findings. Second, we found that their recording is associated with observed factors such as consultation type, diagnosis, and antibiotic treatment, raising uncertainty on potential associations with non-observed factors. Lastly, we observed heterogeneity in recording behaviour between primary care practices.

Comparison with existing literature

Primary care guidelines currently aid GPs in identifying LRTI patients with increased risk of adverse outcomes by highlighting factors associated with such outcomes, including

several signs, symptoms, and measurements. E.g., the Dutch primary care guideline includes ill appearance, confusion, lung auscultation, respiratory rate, heart rate and blood pressure in its severity assessment.[9] UK guidance on pneumonia patients issued by the National Institute of Health and Care Excellence (NICE) suggests to stratify risk of poor prognosis in primary care LRTI patients based on CRB-65.[11] Nevertheless, the findings in our study indicate that the factors addressed in these guidelines are not routinely recorded in LRTI patients, which is largely in line with findings of previous studies. A prospective evaluation of the potential use of CRB-65 in primary care LRTI patients conducted in 13 European countries reported that respiratory rate (22.7%) and blood pressure (31.9%) were infrequently measured as part of routine practice, whereas confusion (99.8%) was recorded in almost all patients as it was part of a standard case report form.[12] Another retrospective UK EHR-based study among primary care patients with community-acquired pneumonia found that complete recording of CRB-65 at index consultation was 0.4%; confusion (0.2%), respiratory rate (3.6%), and blood pressure (17.6%) were all infrequently recorded.[13] Of note, in a primary care-based prospective cohort study of patients with LRTI with near-complete recording, the positivity rates of signs and symptoms were substantially lower than in our study (e.g. fever and chest pain reported in 38.1% and 37.0% versus 57.7% and 78.7%, respectively) and the distributions of measurements revealed less abnormal values.[14] Hence, selective recording of signs, symptoms, and measurements in our study is likely. This might be explained by the fact that GPs both use their severity assessment to selectively assess these factors in a subset of patients and mainly report positive or abnormal values. Extrapolating these positivity rates to our data does, however, not imply that all non-recorded values in our study are negative or normal.

When considering free text variables for use in observational EHR-based prognostic research, non-recorded values could be considered 'missing'. Proper handling of missing data and careful consideration of its mechanisms is therefore essential.[15] Our study identifies two issues that might pose a challenge when using free text EHR data for this purpose: (i) the substantial amount of non-recorded values and (ii) the selective recording of variables. Although the maximum proportion of missing data on variables that is generally accepted in prognostic studies has been subject of debate, some of the observed proportions of missingness in our study are likely to pose a challenge when applying techniques such as multiple imputation (MI).[16–18] Selective recording could potentially be even more problematic, since this might violate assumptions regarding the type of missingness that should be met before applying MI (Box 1).19 In our study, we identified several variables (i.e. consultation type, diagnosis, and immediate antibiotic treatment) that are associated with the proportion of missingness – indicating MAR – and it is likely that even more observed variables are associated with the level of missingness. For example, our observation that signs, symptoms, and measurements are presumably recorded more often in case of positive or abnormal findings might indicate MNAR, which violates one of the assumptions of MI.[19] Alternative methods for handling incomplete

recording of variables could be by considering missing values as 'negative' or 'normal' (i.e. zero imputation) or to include a missing indicator.[20,21] However, when such variables are eventually included in a prediction model their assessment becomes routine care, and GPs will increasingly assess (and record) negative or normal values inherently altering the mechanism of missingness. This may lead to changing predictor effects, which is also referred to as 'the curse of knowing'.[22]

Box 1. Different types of missing data
Missing data is commonly classified into three distinct types[19]:

- Missing completely at random (MCAR; i.e. no systematic differences between observed and missing values)
- Missing at random (MAR; i.e. systematic differences between observed and missing values which can be explained by other variables in the observed data)
- Missing not at random (MNAR; i.e. systematic differences between observed and missing values remain after taking other variables in the observed data into account).

Strengths and limitations

Strengths of our study include the large sample that was derived from a cohort representative of Dutch adult primary care LRTI patients, the selection of relevant signs, symptoms, and measurements that was based on a recent systematic review, and the manual annotation of the variable values from free text EHR data by two authors with high inter-rater reliability.

Some important limitations should, however, be recognized. First, our study should foremost be regarded as an exploratory analysis of the completeness of recording in free text primary care EHR data, and it was not our aim to draw strong conclusions regarding the value of free text EHR data for future research. Nevertheless, we believe our results to be relevant as they highlight some important issues to consider when using free text EHR data in research. Second, it is important to emphasise that the absence of recorded signs, symptoms, and measurements does not necessarily imply that these were not assessed by the GP. It is likely that GPs do not routinely register all information that has been acquired during a consultation while our findings are solely based on recorded data. Therefore, no firm conclusions should be drawn regarding the patient characteristics that GPs routinely assess in primary care LRTI patients. Third, our findings only reflect routine primary health care for LRTI patients during regular working hours, since we did not include out-of-hours consultations. However, even in acutely ill patients presenting to out-of-hours primary care facilities vital parameters such as respiratory rate are often not measured according to a previous report.[23] Lastly, our data do not include information on

the course of the LRTI episodes and whether adverse outcomes such as hospitalisation or mortality occurred. As these are objective indicators of disease severity, it would have been valuable to assess the association between the recording of signs, symptoms, and measurements and adverse outcomes.

Implications for future research

In spite of the descriptive nature of this study, our findings are relevant to other future EHR-based research aiming to include free text variables. Before applying natural language processing to free text EHR data, it is important to gain insight into the completeness of recording of free text variables and identify potential challenges that come with its use. Our findings imply that, prior to including free text variables, recording patterns should be carefully assessed and appropriate missing data handling techniques should be considered based on these patterns. To place our findings in the context of future (prognostic) research and highlight the challenges of using free text EHR data for validating existing prediction models, we have used the CRB-65 model as an illustrative example in Box 2. Nevertheless, the potential value of information captured in free text fields justifies further exploration.[24] Therefore, to properly assess the added value of free text variables in prognostic research, future studies should compare the performance of a prediction model developed with only structured data to a model that also includes free text predictors using appropriate methods[25], and with proper handling of missing data depending on recording patterns and how the model will be applied in practice.

Box 2. Recording of CRB-65 predictors: an illustrative example

To place our findings in the context of future (prognostic) research and highlight the challenges of using free text EHR data for validating existing prediction models, we summarised the recording of CRB-65 predictors (a model developed to predict mortality in pneumonia patients) on a patient level. In our sample, there was no missing data on age, but confusion (3.5%), respiratory rate (5.7%), and blood pressure (17.5%) were all incompletely recorded. Overall, 778 (77.8%) LRTI patients only had information on age available whereas four (0.4%) had complete recording of CRB-65 items. Due to the substantial amount of non-recorded values (i.e. equivalent to missing data since this information is not available elsewhere), external validation of CRB-65 using EHR-data is challenging and would require tailored missing data handling techniques based on the observed recording patterns and underlying assumptions. For example, one could argue to consider non-recorded values of confusion as 'negative' or 'normal'.[21] Whether this assumption also holds for non-recorded values of respiratory rate and blood pressure might be less certain.

		Age (complete data)			
Respiratory rate	Recorded	4 (0.4%)	20 (0.2%)	Recorded	***Blood pressure***
	Missing	15 (1.5%)	136 (13.6%)	Recorded	
	Recorded	2 (0.2%)	31 (3.1%)	Missing	
	Missing	14 (1.4%)	778 (77.8%)	Missing	
		Recorded	Missing		
		Confusion			

Figure. Recording of CRB-65 predictors on a patient level based on manual annotation of 1,000 primary care LRTI consultations.

REFERENCES

1. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using electronic health records for population health research: a review of methods and applications. *Annu Rev Public Health*. 2016;**37**:61–81.
2. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012;**13**(6):395–405.
3. Fu S, Wang L, Moon S, et al. Recommended practices and ethical considerations for natural language processing-assisted observational research: a scoping review. *Clin Transl Sci*. 2023;**16**(3):398–411.
4. Seinen TM, Fridgeirsson EA, Ioannou S, et al. Use of unstructured text in prognostic clinical prediction models: a systematic review. *J Am Med Inform Assoc*. 2022;**29**(7):1292–1302.
5. Rijk MH, Platteel TN, van den Berg TMC, et al. Prognostic factors and prediction models for hospitalisation and all-cause mortality in adults presenting to primary care with a lower respiratory tract infection: a systematic review. *BMJ Open*. 2024;**14**(3):e075475.
6. Smeets H, Kortekaas M, Rutten F, 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.
7. Rijk MH, Platteel TN, Geersing GJ, et al. 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. *Diagn Progn Res*. 2023;**7**(1):23.
8. de Boer AR, de Groot MCH, Groenhof TKJ, et al. Data mining to retrieve smoking status from electronic health records in general practice. *Eur Heart J Digit Health*. 2022;**3**(3):437–44.
9. Verheij T, Hopstaken R, Prins J, et al. *NHG-Standaard Acuut hoesten* [internet]. Utrecht: Nederlands Huisartsen Genootschap; 2020. Accessed May 6, 2025. Available at: https://richtlijnen.nhg.org.
10. R Core Team. *R: A Language and Environment for Statistical Computing* [Internet]. R Foundation for Statistical Computing, Vienna, Austria; 2022. Available from: https://www.R-project.org/.
11. National Institute for Health and Care Excellence. *Pneumonia in adults: diagnosis and management (CG191)* [internet]. London: National Institute for Health and Care Excellence; 2014. Accessed May 7, 2025. Available at: https://www.nice.org.uk/guidance/cg191.
12. Francis NA, Cals JW, Butler CC, 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.
13. Launders N, Ryan D, Winchester C, et al. Management of Community-Acquired Pneumonia: an Observational Study in UK Primary Care. *Pragmat Obs Res*. 2019;**10**:53–65.
14. Moore M, Stuart B, Lown M, et al. Predictors of adverse outcomes in uncomplicated lower respiratory tract infections. *Ann Fam Med*. 2019;**17**(3):231–8.
15. Nijman SWJ, Leeuwenberg AM, Beekers I, et al. Missing data is poorly handled and reported in prediction model studies using machine learning: a literature review. *J Clin Epidemiol*. 2022;**142**:218–29.
16. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. *BMC Med Res Methodol*. 2017;**17**(1):162.
17. Lee KJ, Carlin JB. Recovery of information from multiple imputation: a simulation study. *Emerg Themes Epidemiol*. 2012;**9**(1):3.
18. Madley-Dowd P, Hughes R, Tilling K, Heron J. The proportion of missing data should not be used to guide decisions on multiple imputation. *J Clin Epidemiol*. 2019;**110**:63–73.

19. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;**338**:b2383.
20. Sperrin M, Martin GP. Multiple imputation with missing indicators as proxies for unmeasured variables: simulation study. *BMC Med Res Methodol*. 2020;**20**(1):185.
21. Wells BJ, Nowacki AS, Chagin K, Kattan MW. Strategies for handling missing data in electronic health record derived data. *EGEMS*. 2013;**1**(3):1035.
22. Groenwold RHH. Informative missingness in electronic health record systems: the curse of knowing. *Diagn Progn Res*. 2020;**4**:8.
23. Loots FJ, Dekker I, Wang RC, et al. The accuracy and feasibility of respiratory rate measurements in acutely ill adult patients by GPs: a mixed-methods study. *BJGP Open*. 2022;**6**(4):BJGPO.2022.0029.
24. Seinen TM, Kors JA, van Mulligen EM, Fridgeirsson E, Rijnbeek PR. The added value of text from Dutch general practitioner notes in predictive modeling. *J Am Med Inform Assoc*. 2023;**30**(12):1973–1984.
25. Cook NR. Quantifying the added value of new biomarkers: how and how not. *Diagn Progn Res*. 2018;**2**:14.

SUPPLEMENTARY MATERIALS FOR CHAPTER 4

Table S1. Differences in recording of signs, symptoms, and measurements between LRTI episodes stratified according to type of index consultation, diagnosis, and immediate antibiotic prescription.

	Signs and symptoms								
	Recording (%)			Recording (%)			Recording (%)		
	Consultation	Home visit	p value †	Acute bronchitis	Pneumonia	p value †	No immediate AB	Immediate AB	p value †
Patient reported									
Cough	76.6	76.7	0.983	82.6	71.9	**<0.001**	73.9	82.4	**0.003**
Fever	55.2	48.0	0.104	50.3	57.0	**0.036**	51.4	59.9	**0.012**
Shortness of breath	52.1	59.3	0.103	51.7	54.4	0.407	51.8	56.1	0.206
Chest pain	17.9	24.0	0.077	13.5	22.9	**<0.001**	16.9	22.9	**0.024**
Patient or GP reported									
Confusion	2.1	11.3	**<0.001**	1.1	5.3	**<0.001**	2.9	4.7	0.157
GP reported									
Crackles (auscultation)	86.7	82.7	0.187	89.2	83.7	**0.011**	83.3	92.2	**<0.001**
Ill appearance	31.6	39.3	0.065	24.7	39.1	**<0.001**	31.1	36.4	0.100
	Measurements								
	Recording (%)			Recording (%)			Recording (%)		
	Consultation	Home visit	p value †	Acute bronchitis	Pneumonia	p value †	No immediate AB	Immediate AB	p value †
Body temperature	42.6	61.3	**<0.001**	36.8	52.0	**<0.001**	41.4	53.9	**<0.001**
Oxygen saturation	42.9	57.3	**0.001**	37.5	51.0	**<0.001**	40.8	54.2	**<0.001**
Heart rate	37.3	55.3	**<0.001**	31.1	46.9	**<0.001**	36.6	47.3	**0.001**
CRP	21.1	4.0	**<0.001**	16.5	20.1	0.146	18.5	18.5	0.998
Systolic BP	13.9	38.0	**<0.001**	12.1	21.7	**<0.001**	17.0	18.5	0.571
Diastolic BP	13.9	39.1	**<0.001**	12.1	21.7	**<0.001**	17.0	18.5	0.571
Respiratory rate	4.5	12.7	**<0.001**	4.8	6.4	0.282	5.7	5.6	0.957

† Based on two sample test Z-test for equality of proportions. P values in bold are statistically significant (at alpha 0.05).

Abbreviations: LRTI, lower respiratory tract infection; GP, general practitioner; CRP, C-reactive protein; BP, blood pressure; AB, antibiotics.

5

Predicting 30-day hospitalisation or mortality in patients presenting to general practice with a lower respiratory tract infection: a prediction model development and external validation study

Merijn H. Rijk
Tamara N. Platteel
Geert-Jan Geersing
Artuur M. Leeuwenberg
Monika Hollander
Paul Little
Frans H. Rutten
Roderick P. Venekamp
Maarten van Smeden

Submitted for publication

Abstract

Objectives To develop and validate a model that predicts individual risk of 30-day all-cause hospitalisation or mortality in patients presenting to their general practitioner (GP) with a lower respiratory tract infection (LRTI), and to evaluate the incremental predictive value of cardiometabolic disease (CMD) for these outcomes.

Design Prediction model development, internal and external validation study, using large electronic health records (EHR) based general practice cohorts linked to hospitalisation and mortality data.

Setting General practice.

Participants Two EHR-based cohorts of patients aged ≥40 years with a GP diagnosis of LRTI. The pre-pandemic development cohort comprised 11,065 patients from the region of Utrecht, the Netherlands, diagnosed with an LRTI between 2016 and 2019. The post-pandemic external validation cohort comprised 6,284 patients from the region of Amsterdam, the Netherlands, diagnosed with an LRTI between 2022 and 2023.

Main outcome measures All-cause hospitalisation or mortality within 30 days from LRTI diagnosis.

Results The outcome occurred in 863 (7.8%) patients of the development cohort and 717 (11.4%) of the external validation. The final prediction model included: age, sex, number of CMD, medical history (pneumonia, malignancy, lung disease (COPD or asthma), dementia, and hospitalisation <1 year), influenza vaccination <1 year, antibiotic prescription <1 month, current medication use (immunosuppressants, inhalation medication, and antidepressants), and diagnosis of pneumonia. External validation yielded a c-statistic of 0.71 (0.69 to 0.73), a calibration intercept of 0.28 (0.20 to 0.36), and a calibration slope of 0.95 (0.85 to 1.05). Predicted risks ranged from 1.5% to 51.3% (median 6.9%). The incremental predictive value of CMD beyond age and sex was low (Dc-statistic +0.02).

Conclusions The developed prediction model holds promise to aid GPs by identifying LRTI patients at highest risk of 30-day hospitalisation or mortality. Before adoption in everyday general practice, future research should assess the impact of model application on individual patient outcomes.

INTRODUCTION

Community-acquired lower respiratory tract infections (LRTI) are a common reason for general practitioner (GP) consultation.[1,2] Although uncomplicated LRTIs are usually self-limiting with limited benefit of antibiotic treatment,[3,4] adverse outcomes such as hospitalisation or mortality occur more often in complicated LRTIs, notably in patients with community-acquired pneumonia (CAP).[5] However, distinguishing patients with CAP from uncomplicated LRTI is challenging in general practice,[6,7] leaving GPs with uncertainties on how to identify LRTI patients with poor prognosis.

Several prognostic factors for hospitalisation and mortality in LRTI patients presenting to general practice have been established,[5] and concurrent cardiometabolic disease (CMD) – e.g. cardiovascular disease and diabetes – has particularly been linked to poor prognosis.[5,8,9] In contrast to individual prognostic factors, which do not directly translate into absolute risks and are thus difficult to interpret in a clinical setting, prediction models that estimate individual risk of adverse outcomes can assist GPs in identifying LRTI patients with poor prognosis, and may inform subsequent management decisions to mitigate this risk. However, a prediction model that can be applied for this task in general practice is currently lacking.[5]

In this study, we therefore aimed to develop, internally, and externally validate a prediction model that estimates individual risk of 30-day all-cause hospitalisation or mortality in adult LRTI patients presenting to general practice. We used large cohorts based on electronic health records (EHR) data and applied an incremental value approach, which allowed us to determine the incremental predictive value of CMD and select the most optimal model for external validation.

METHODS

While reporting this study, we adhered to the prediction model risk of bias assessment tool (PROBAST+AI) criteria and transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD+AI) statement.[10,11]

Study design, setting, and participants

The design of this model development and validation study, including sample size calculations, has been reported in detail elsewhere.[12] In short, two cohorts were constructed from two large general practice EHR databases, covering data on patient demographics, consultation, coded disease episodes and medical history (using the International Classification of Primary Care (ICPC)[13]), prescriptions (using Anatomical Therapeutic Chemical (ATC) codes[14]), measurements (e.g. blood pressure or respiratory rate, when reported),

and laboratory results. The development cohort was derived from the Julius General Practitioners' Network (JGPN)[15] in the region of Utrecht, the Netherlands and comprised patients aged ≥40 years presenting to their GP with an LRTI (defined as an ICPC diagnosis code of either acute bronchitis (R78) or pneumonia (R81)) between 1 January 2016 and 31 December 2019. The validation cohort was derived from the Academic Network of General Practitioners Amsterdam (ANHA)[16] – which has a similar data structure and coverage as JGPN – and comprised similarly selected LRTI patients presenting to primary care in a post coronavirus disease 2019 (COVID-19) pandemic period, i.e. between 1 January 2022 and 1 December 2023, with the remaining 30 days in 2023 only used for follow-up. For both cohorts, only the individual's first LRTI episode during follow-up was included. Subsequently, the general practice databases were linked to data on hospitalisation (Dutch Hospital Data (DHD)) and mortality (Statistics Netherlands (CBS)) on a patient level, resulting in a comprehensive database covering the full trajectory of the included LRTI episodes.

Outcome

Prediction models were developed estimating absolute individual risk of all-cause hospitalisation or mortality within 30 days from GP diagnosis of LRTI (yes/no).

Candidate predictors

Selection of candidate predictors was based on a recent systematic review and clinical expertise.[5] The following candidate predictors were included in the model development process: patient demographics (age and sex), CMD (i.e., a history of diabetes, heart failure, ischaemic heart disease, stroke, transient ischaemic attack (TIA), pulmonary embolism, deep venous thrombosis, atrial fibrillation, and peripheral artery disease), medical history other than CMD (history of pneumonia, non-dermatological malignancies, chronic obstructive pulmonary disease (COPD), asthma, dementia, and hospitalisation <1 year), influenza vaccination <1 year, antibiotic prescription <1 month, current medication use (immunosuppressants, inhalation medication, and antidepressants), and type of LRTI diagnosis code (i.e., either acute bronchitis or pneumonia). In addition, some candidate predictors were automatically extracted through EHR-based algorithms, including the number of GP-consultations in the previous year, smoking status (retrieved by text mining of free text consultation data[17]), and a frailty index for those ≥60 years based on structured EHR data.[18] All candidate predictors were measured on the day of GP diagnosis of LRTI (i.e. moment of prediction). An overview of all candidate predictors and their definition is presented in the Supplementary Table 1 (Table S1).

Missing data

Missing data on patient demographics, frailty index (calculated based on registered EHR-codes) and number of GP-consultations in the previous year were not anticipated. The absence of registered medical history, comorbidities, and prescriptions was considered

as the absence of the condition or medication use. Candidate predictors retrieved with EHR-based algorithms are ideally incorporated in a model that runs on EHR input data, which would produce a pattern of missingness similar to our data. Therefore, a missing indicator was considered appropriate in the case of missing values for smoking status.

Statistical analysis

Characteristics of the development and validation cohorts were descriptively summarised. An incremental value approach was adopted during model development, which allowed us to evaluate the added predictive value of adding extra sets of candidate predictors to the model. To account for possible non-linearity of continuous candidate predictors, restricted cubic splines (RCS) were considered using four knots on the percentiles 0.05, 0.35, 0.65, and 0.95. The prediction models were developed using multivariable elastic net penalised logistic regression, which accounts for both predictor selection and shrinkage of their coefficients. Hyperparameter tuning was performed using 10-fold cross-validation to determine the optimal regularisation penalty.

The model development process comprised four steps (Figure 1). First, in step 1, a basic model – including age, sex, and an interaction term – was fitted, and the optimal modelling strategy for age was examined (i.e., linear or using RCS). Second, the incremental predictive value of CMD was explored by adding CMDs to the modelling process, modelled as both separate candidate predictors and as a cumulative number of CMDs (0, 1, or ≥2) to find the optimal modelling strategy; step 2. Third, in step 3, a clinical model was fitted adding all other candidate predictors that are readily available for GPs in everyday clinical practice. Lastly, a full EHR-based model – adding number of GP-consultations in the previous year, smoking status, and frailty index to the clinical model – was fitted, while examining the optimal modelling strategy for the added continuous candidate predictors (i.e. linear or using RCS); step 4. During each step the fitted models were internally validated by 1,000 bootstrapping samples, and model performance was assessed based on measures of discrimination (c-statistic), calibration (intercept and slope), and the range of predicted risks with accompanying 95% confidence intervals where appropriate. The final model, to be used in subsequent validation procedures of this study, was chosen out of the four developed models based on model performance at internal validation and complexity (e.g. a more complex model with only a slight increase in performance might not be preferable).

To assess heterogeneity of predictor effects by time, temporal internal-external cross-validation (IECV) of the final model was performed by repeating the complete model development procedure in the development cohort while leaving one year out at a time and subsequently validating the model in the year that was left out. Heterogeneity in model discrimination and calibration was assessed visually through forest plots and by fitting random effects models to calculate pooled performance estimates.

Lastly, the final model was externally validated in the post-pandemic ANHA cohort. Measures of discrimination and calibration, and the range of predicted risks were evaluated.

All analyses were performed in R version 4.4.3.[19]

Patient and Public Involvement

A patient representative was consulted throughout all phases of this study, providing input to the initial research idea, design of the study, and interpretation of the results to assure these aligned with the patient's perspective.

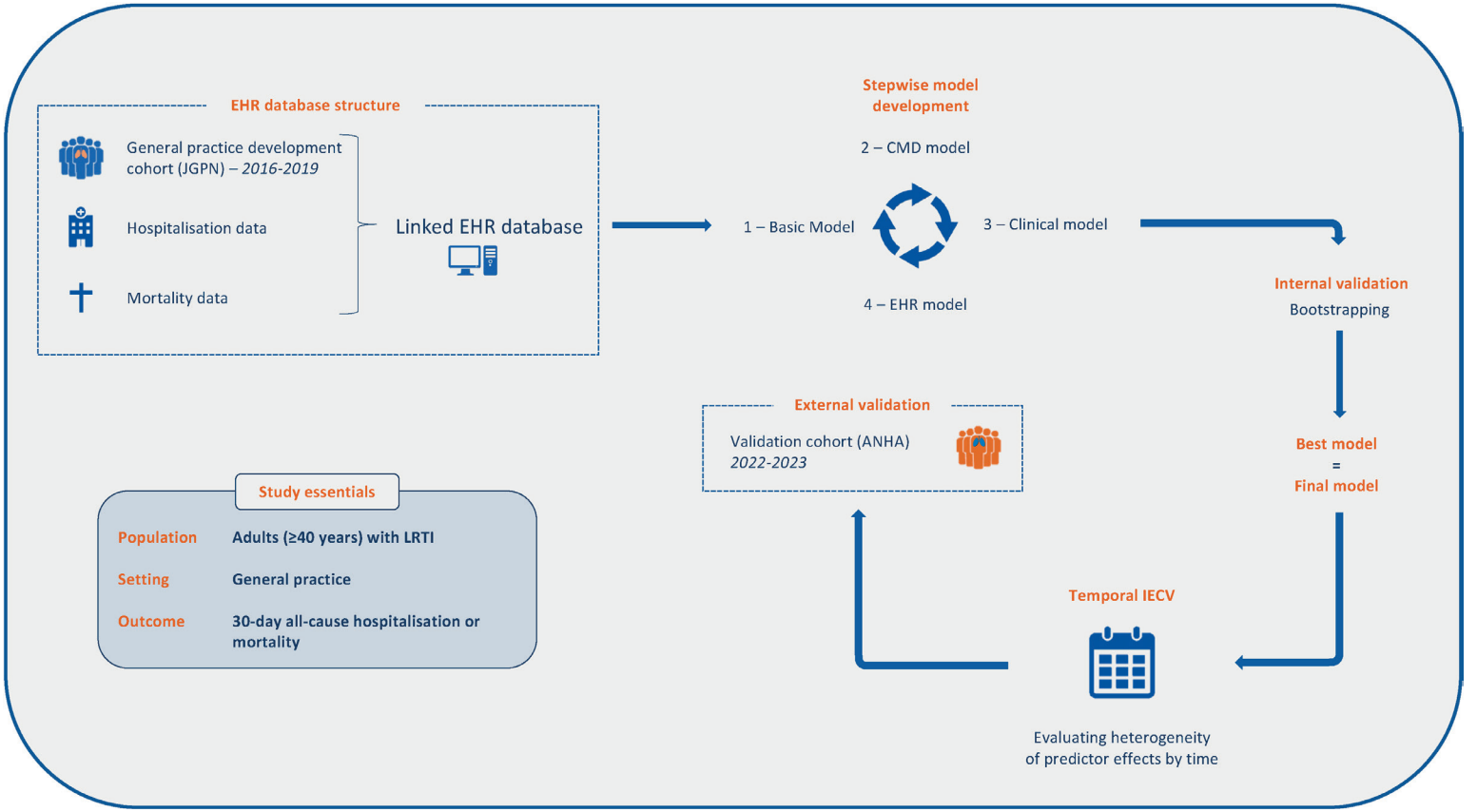


Figure 1. Visual representation of cohort construction and prediction model development and validation process.

LRTI, lower respiratory tract infections; JGPN, Julius General Practitioners' Network; EHR, electronic health records; CMD, cardiometabolic disease; IECV, internal-external cross-validation; ANHA, Academic Network of General Practitioners Amsterdam.

Table 1. Characteristics of model development and validation cohort

	Development cohort		Validation cohort	
Characteristic	**Total**	**With outcome**	**Total**	**With outcome**
n (%)	11065	863 (7.8)	6284	717 (11.4)
Median age (IQR)	64 (53 – 74)	70 (61 – 78)	66 (55 – 76)	72 (62 – 81)
Female sex, n (%)	6206 (56.1)	451 (52.3)	3575 (56.9)	395 (55.1)
LRTI diagnosis, n (%)				
Acute bronchitis	547 (41.1)	164 (19.0)	1858 (29.6)	79 (11.0)
Pneumonia	6518 (58.9)	699 (81.0)	4426 (70.4)	638 (89.0)
Year, n (%)				
2016	2720 (24.6)	196 (22.7)	-	-
2017	2555 (23.1)	186 (21.6)	-	-
2018	3065 (27.7)	237 (27.5)	-	-
2019	2725 (24.6)	244 (28.3)	-	-
2022	-	-	3089 (49.2)	367 (51.2)
2023	-	-	3195 (50.8)	350 (48.8)
Median frailty index (IQR)	0.18 (0.0 – 0.32)	0.30 (0.0 – 0.40)	0.18 (0.0 – 0.30)	0.26 (0.14 – 0.36)
Smoking status, n (%)				
Current	3053 (27.6)	271 (31.4)	1762 (28.0)	211 (29.4)
Former	2409 (21.8)	204 (23.6)	1501 (23.9)	195 (27.2)
Never	2238 (20.2)	144 (16.7)	1334 (21.2)	112 (15.6)
No information	3365 (30.4)	244 (28.3)	1687 (26.8)	199 (27.8)
Comorbidities, n (%)				
Diabetes	578 (5.2)	67 (7.8)	246 (3.9)	37 (5.2)
Malignancy	1073 (9.7)	154 (17.8)	756 (12.0)	133 (18.5)
Dementia	526 (4.8)	68 (7.9)	498 (7.9)	84 (11.7)
COPD or asthma	2279 (20.6)	239 (27.7)	1414 (22.5)	188 (26.2)
Number of CMD, n (%)				
0	7859 (71.0)	498 (57.7)	4199 (66.8)	348 (48.5)
1	2161 (19.5)	211 (24.4)	1366 (21.7)	232 (32.4)
≥2	1045 (9.4)	154 (17.8)	719 (11.4)	137 (19.1)
Medical history, n (%)				
Pneumonia	3462 (31.3)	340 (39.4)	2079 (33.1)	282 (39.3)
Hospitalisation <1 year	2387 (21.6)	388 (45.0)	1377 (21.9)	291 (40.6)
Median number of GP-consultations <1 year (IQR)	8 (3 – 15)	12 (6 – 21)	5 (1 – 12)	8 (2 – 17)
Medication use, n (%)				
Immunosuppressant	118 (1.1)	16 (1.9)	125 (2.0)	18 (2.5)
Inhalation medication	806 (7.3)	102 (11.8)	358 (5.7)	39 (5.4)
Antidepressant	649 (5.9)	68 (7.9)	363 (5.8)	52 (7.3)
Antibiotics <1 month	804 (7.3)	116 (13.4)	482 (7.7)	88 (12.3)
Influenza vaccination <1 year, n (%)	3327 (30.1)	328 (38.0)	1880 (29.9)	252 (35.1)

IQR, interquartile range; LRTI, lower respiratory tract infection; COPD, chronic obstructive pulmonary disease; CMD, cardiometabolic disease; GP, general practitioner.

Table 2. Modelling strategies and model performance upon internal validation (1,000 bootstrapping samples) of the developed prediction models.

Model	Predictors with varying modelling strategies				Candidate predictors	#RC	Model performance (95% CI)			Range of predicted risks (median)
	Age	CMD	Frailty index	Number of GP-consultations <1 year			c-statistic	Calibration intercept	Calibration slope	
Basic 1	L	-	-	-	Age, sex, and interaction term	4	0.61 (0.59 – 0.63)	0.00 (-0.07 – 0.07)	1.04 (0.86 – 1.23)	3.8% – 20.8% (7.4%)
Basic 2	RCS	-	-	-		8	0.62 (0.60 – 0.64)	0.00 (-0.07 – 0.07)	1.04 (0.86 – 1.23)	3.5% - 24.3% (7.7%)
CMD 1	L	Cumulative[1]	-	-	Basic model + diabetes, heart failure, ischaemic heart disease, stroke, TIA, PE, DVT, atrial fibrillation, and peripheral artery disease	6	0.63 (0.61 – 0.65)	0.00 (-0.07 – 0.07)	1.04 (0.88 – 1.20)	3.9% - 24.7% (7.0%)
CMD 2	L	Separate	-	-		11	0.63 (0.61 – 0.65)	0.00 (-0.07 – 0.07)	1.04 (0.88 – 1.19)	3.9% - 37.1% (7.0%)
Clinical	L	Cumulative[1]	-	-	CMD model + LRTI diagnosis code, hospitalisation <1 year, history of pneumonia, malignancy, COPD/ asthma, or dementia, influenza vaccination <1 year, use of immunosuppressants, inhalation medication, or antidepressants, and antibiotic use <1 month	17	0.73 (0.71 – 0.74)	0.00 (-0.07 – 0.07)	1.03 (0.94 – 1.12)	1.5% - 53.4% (5.9%)
EHR 1	L	Cumulative[1]	L	L	Clinical model + smoking status, frailty index, and number of GP consultations <1 year	22	0.74 (0.72 – 0.75)	0.00 (-0.07 – 0.07)	1.03 (0.94– 1.11)	1.4% - 51.6% (5.7%)
EHR 2	L	Cumulative[1]	RCS	RCS		26	0.74 (0.72 – 0.75)	0.00 (-0.07 – 0.07)	1.03 (0.94 – 1.11)	1.4% - 52.4% (5.7%)

[1] Modelling of CMD as cumulative predictor: 0, 1, or ≥2 CMD based on medical history

#RC, number of regression coefficients; CI, confidence interval; CMD, cardiometabolic disease; GP, general practitioner; L, linear; RCS, restricted cubic splines; TIA, transient ischaemic attack; PE, pulmonary embolism; DVT, deep venous thrombosis; LRTI, lower respiratory tract infection; EHR, electronic health record.

RESULTS

Cohort characteristics

The development cohort comprised 11,702 LRTI patients of whom 11,065 (94.6%) could be linked to hospitalisation and mortality registries. Among these, 863 (7.8%) were hospitalised or died within 30 days after diagnosis (Table 1). All-cause hospitalisation and mortality occurred in 859 (7.8%) and 21 (0.2%) of the participants, respectively, and the median time to the composite outcome was 4 days (IQR 0 – 14.5) (Figure S4). The median age was 64 years (IQR 53 – 74), and 6,206 (56.1%) were female. Pneumonia was registered as LRTI diagnosis in 6,518 (58.9%) participants.

The validation cohort comprised 6,513 LRTI patients, of whom 6,284 (96.5%) could be linked to hospitalisation and mortality registries. Among these patients, 717 (11.4%) outcomes were registered: 679 (10.8%) participants were hospitalised and 73 (1.2%) died within 30 days from LRTI diagnosis, and the median time to the composite outcome was 3 days (IQR 0 to 12) (Figure S4). The median age was 66 years (IQR 55 – 76), 3,575 (56.9%) were female, and pneumonia was diagnosed in 4,426 (70.4%) participants.

Missing data

There was no missing data on patient demographics or number of GP consultations in the previous year. Based on text mining of free-text consultation data, there was no information on smoking status of 3,365 (30.4%) participants from the development cohort and 1,687 (26.8%) participants from the validation cohort.

Stepwise model development and internal validation

Apparent performance of the developed models is presented in the Supplementary Table 2 (Table S2). Model performance was stable upon internal validation by bootstrapping (Table 2). The basic model in step 1 showed similar performance for both modelling strategies of age: linear versus RCS resulted in a c-statistic of 0.61 (0.59 to 0.63) versus 0.62 (0.59 to 0.63), respectively, with a calibration intercept of 0.00 (-0.07 to 0.07) and a calibration slope of 1.04 (0.86 to 1.23) for both modelling strategies. The basic model with the lowest number of regression coefficients – with age modelled linear – was therefore used in subsequent steps of the development process. The two strategies for modelling CMD (as separate candidate predictors or a cumulative number) in step 2 also resulted in similar model performance: c-statistics of 0.63 (0.61 – 0.65) and calibration intercepts of 0.00 (-0.07 to 0.07) were similar for both modelling strategies, with a calibration slope of 1.04 (0.88 to 1.20) and 1.04 (0.88 to 1.19) for cumulative and separate CMD, respectively. Although predicted risks increased with the cumulative number of CMDs (Figure 2), the incremental predictive value of CMD was low: Dc-statistic +0.02 without improvement of model calibration (Figure S5). The CMD model including the cumulative number of CMDs was carried forward in the development process based on the (lower) number of

regression coefficients and model simplicity. In step 3, model performance increased after adding the candidate predictors of the clinical model, with a c-statistic 0.73 (0.71 to 0.74), a calibration intercept 0.00 (-0.07 to 0.07), and a calibration slope 1.03 (0.94 to 1.12). The two EHR models in step 4 – with additional continuous candidate predictors modelled as either linear or with RCS – showed similar performance compared to the clinical model, with c-statistics 0.74 (0.72 to 0.75), calibration intercepts 0.00 (-0.07 to 0.07), and calibration slopes 1.03 (0.94 to 1.11) for both models. Considering overall model performance, the range of predicted risks and model complexity, the clinical model of step 3 was considered the most optimal and therefore final model to continue the model validation process. The final model consisted of 14 parameters: age, an interaction term between age and sex, number of concurrent CMDs, history of pneumonia, malignancy, lung disease (defined as COPD or asthma), dementia, hospitalisation <1 year, influenza vaccination <1 year, antibiotic prescription <1 month, use of immunosuppressants, use of inhalation medication, use of antidepressants, and GP diagnosis of pneumonia.

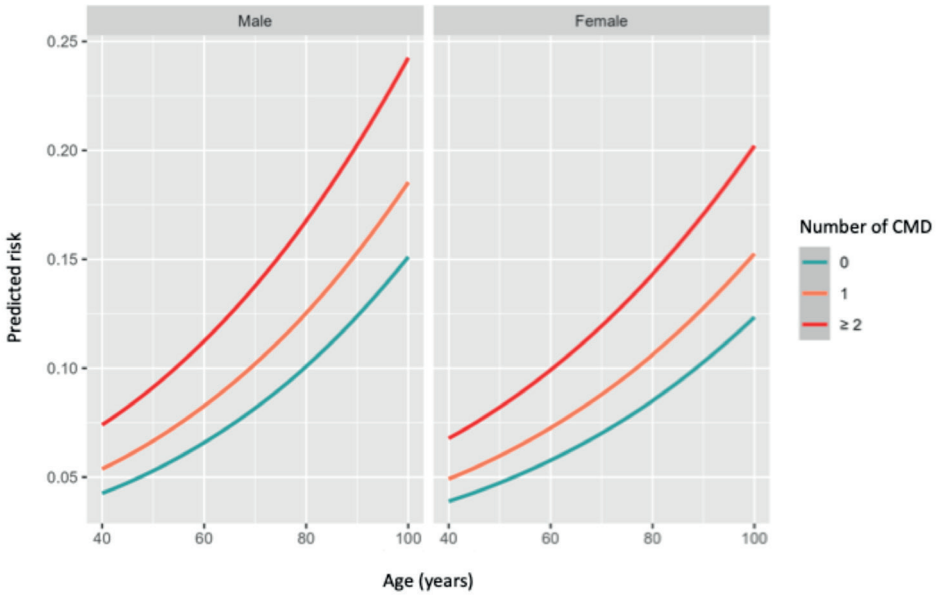


Figure 2. Risks predicted by CMD 1 model across age, sex, and number of CMD

Individual predicted risks by the CMD 1 model, based on all possible combinations of predictor vales for age, sex, and number of CMD. CMD, cardiometabolic disease.

In the development cohort, risks predicted by the final model ranged from 1.5% to 53.4% (median: 5.9%) (Figure 3A). Linear predictors and apparent model calibration curves of all developed models are presented in the Supplementary materials (Table S3 and Figure S6).

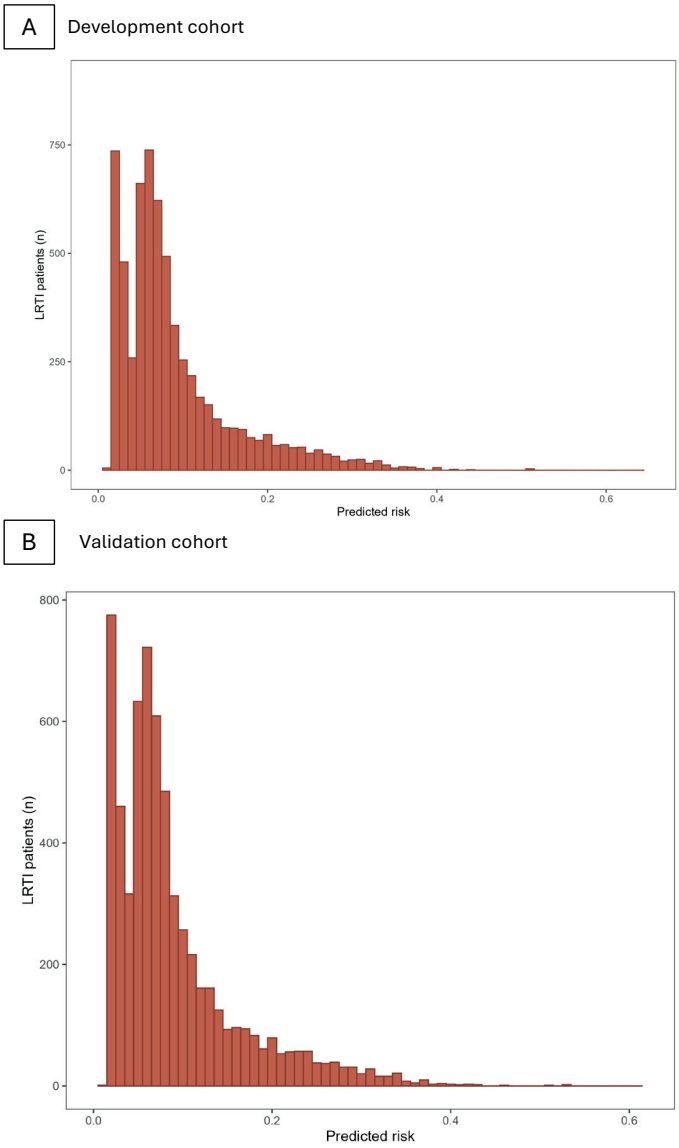


Figure 3. Individual risks predicted by the final model in the development (**A**) and validation (**B**) cohort

Temporal IECV

Upon temporal IECV of the final model, discriminative performance was stable over the model development study period, with a random effects pooled c-statistic of 0.72 (0.70 to 0.74) (Figure 4A). Model calibration was also fairly stable over the model development years, with random effects pooled estimates of the calibration intercept at 0.00 (-0.06 to 0.07) (Figure 4B) and slope at 0.98 (0.84 to 1.13) (Figure 4C).

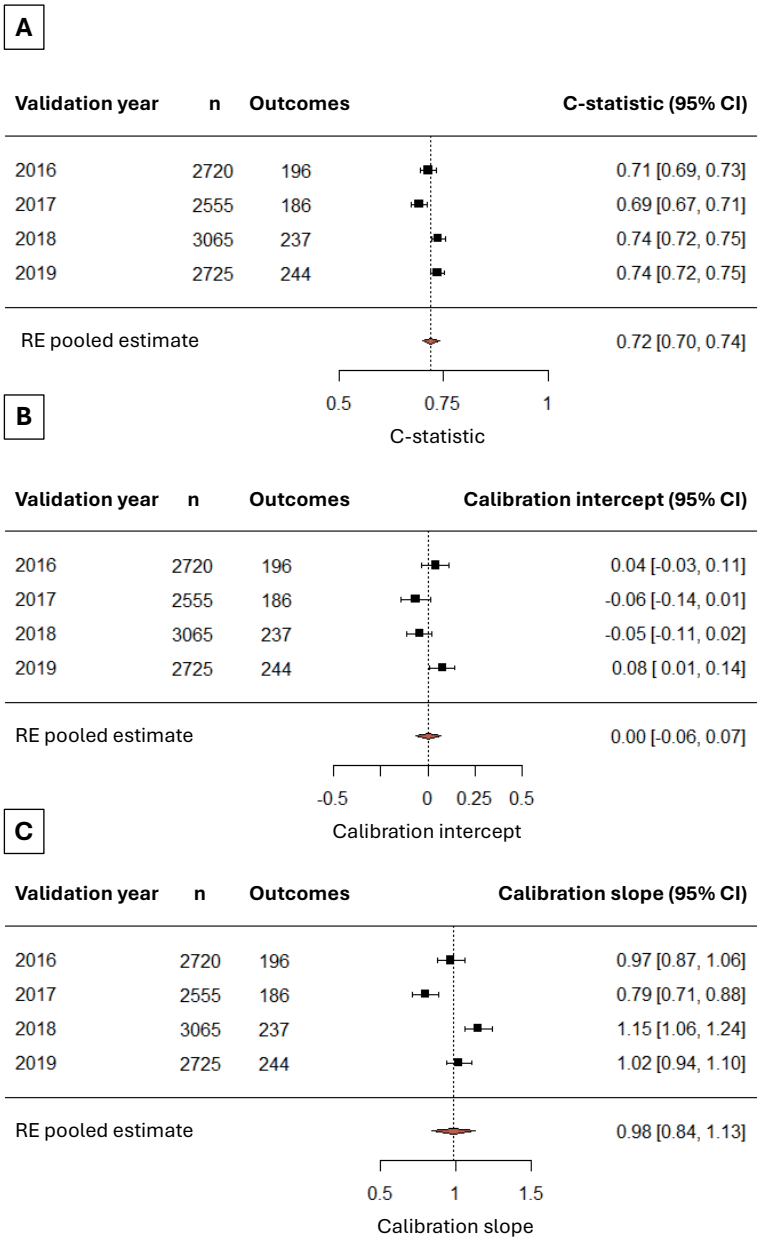


Figure 4. Results of temporal IECV of the final model

Forests plots of random effects models fitted by developing the final model in three years from the 2016 – 2019 period and subsequently validating the model in the year that was left out, for every year separately. Random effects pooled estimates of c-statistic (**A**), calibration intercept (**B**), and calibration slope (**C**).

CI, confidence interval; RE, random effects.

External validation

Upon external validation of the final model in the post-pandemic validation cohort, the c-statistic was 0.71 (0.69 to 0.73) with a calibration intercept of 0.28 (0.20 to 0.36) and slope 0.95 (0.85 to 1.05). Visual inspection of the flexible calibration curve indicated some underprediction in the lower risk strata up to predicted risks of approximately 0.20 (Figure 5). Predicted risks in the validation cohort ranged from 1.5% to 51.3% (median 6.9%) (Figure 3B).

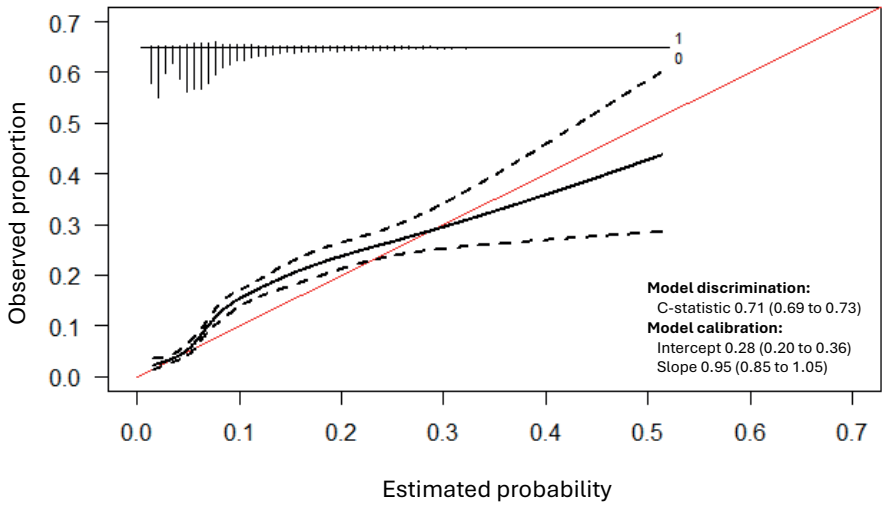


Figure 5. Calibration curve of external validation of the final model

Flexible calibration curve of external validation of the final model with summary model discrimination and calibration statistics: 0.71 (0.69 to 0.73), calibration intercept 0.28 (0.20 to 0.36) and slope 0.95 (0.85 to 1.05).

DISCUSSION

In this study, we developed, internally, and externally validated an EHR-based prediction model estimating individual risk of all-cause hospitalisation or mortality within 30 days in adult LRTI patients presenting to general practice. Albeit included as one of the predictors, the incremental predictive value of concurrent CMD beyond age and sex was low. Indeed, predictive performance was higher when including additional clinical variables beyond CMD. External validation of our final model in a recent post-pandemic cohort showed promising predictive performance, with predicted risks ranging up to 51.3%. The final model includes 14 predictors readily available from EHR: age, sex, number of concurrent CMDs, history of pneumonia, malignancy, lung disease (COPD or asthma), dementia, hospitalisation <1 year, influenza vaccination <1 year, antibiotic prescription <1 month,

use of immunosuppressants, use of inhalation medication, use of antidepressants, and GP diagnosis of pneumonia.

Strengths of this study include the state-of-the-art prediction modelling methods that were applied – using a step-wise and incremental value approach for model development and a rigorous validation process of the final model – and the use of two large EHR-based cohorts of LRTI patients presenting to general practice linked to hospitalisation and mortality data on a patient level, with external validation in a post-pandemic cohort. Our results are therefore applicable to the current clinical context of LRTI, with SARS-CoV-2 as persistent cause of LRTI among other pathogens.

Nevertheless, some limitations deserve mentioning. First, EHR data are not primarily collected for research purposes, and codes may be inaccurately or incompletely registered. Its use to select LRTI patients and define candidate predictors therefore potentially introduced misclassification.[20,21] As a result, predictor effects will likely differ when used with prospectively collected patient data (e.g. as a 'scoring rule'), and the developed model should thus be integrated in the EHR environment, with similar patterns of misclassification.[22] Second, although plausibly important as well[23], presenting signs and symptoms were not included as predictors in this study, since these are captured in EHR clinical notes and would require advanced natural language processing (NLP) techniques for data extraction. Third, we did not explicitly include asthma and COPD exacerbations – part of LRTI definitions in related studies[24–26] – in the study population, as there is no separate ICPC-code that distinguished asthma and COPD exacerbations from asthma and COPD follow-up consultations. Model performance should therefore be carefully evaluated in patients presenting to general practice with asthma and COPD exacerbations before the model can be applied to this population. Lastly, the outcome definition was broad, including all-cause hospitalisation, and outcomes in our study may have therefore included non-acute hospitalisations as well. However, this outcome definition allowed us to also include clinically relevant deterioration of non-respiratory conditions (e.g. heart failure, functional decline in patients with malignancies) as adverse outcome following LRTI.

Previously developed models to predict adverse outcomes in LRTI patients presenting to general practice include CRB-65 (confusion, respiratory rate, blood pressure, and age ≥65 years) and a model developed by Bont et al.[24,27] The current UK general practice guideline on CAP suggests using CRB-65 in addition to clinical judgement to determine disease severity.[28] It has, however, been developed to predict 30-day all-cause mortality in hospitalised patients with CAP, without proper validation or impact assessment for use in a general practice setting.[5] Predictors of the model developed by Bont et al. were included in the prediction model that was developed and validated in this study.

The model developed in this study showed promising predictive performance and can identify LRTI patients presenting to general practice with a clinically relevant risk of adverse outcomes. It could therefore aid GPs in the clinical decision-making process for patients with LRTI, in addition to clinical judgement, by tailoring close monitoring strategies or early intervention. However, before adoption in general practice, the impact of applying our model on individual patient outcomes, e.g. hospitalisations or antibiotic prescriptions, should be evaluated in an impact study.[29,30] Of note, this study focused on prediction, not on causal inference. The predictors included in our multivariable prediction model should therefore not be explained as individual 'risk factors' as this will lead to unsubstantiated claims of causal effects.[31,32] With recent advancements in the field of NLP, future research may also focus on the incremental predictive value of adding presenting signs and symptoms retrieved from EHR clinical notes to the current model.[33,34]

REFERENCES

1. Millett ERC, Quint JK, Smeeth L, Daniel RM, Thomas SL. Incidence of community-acquired lower respiratory tract infections and pneumonia among older adults in the United Kingdom: a population-based study. *PLoS One*. 2013;**8**(9):e75131.
2. Winchester CC, Macfarlane T V., Thomas M, Price D. Antibiotic prescribing and outcomes of lower respiratory tract infection in UK primary care. *Chest*. 2009;**135**(5):1163–72.
3. Little P, Stuart B, Smith S, et al. Antibiotic prescription strategies and adverse outcome for uncomplicated lower respiratory tract infections: prospective cough complication cohort (3C) study. *BMJ*. 2017;**357**:j2148.
4. Moore M, Stuart B, Coenen S, 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.
5. Rijk MH, Platteel TN, van den Berg TMC, et al. Prognostic factors and prediction models for hospitalisation and all-cause mortality in adults presenting to primary care with a lower respiratory tract infection: a systematic review. *BMJ Open*. 2024;**14**(3):e075475.
6. Teepe J, Broekhuizen BDL, Loens K, 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, 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. van Royen FS, Joosten LPT, van Smeden M, 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.
9. Islam N, Shabnam S, Khan N, Gillies C, Zaccardi F, Banerjee A, et al. Combinations of multiple long term conditions and risk of hospital admission or death during winter 2021-22 in England: population based cohort study. *BMJ Med*. 2024;**3**(1):e001016.
10. Moons KGM, Damen JAA, Kaul T, et al. PROBAST+AI: An updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ*. 2025;**388**:e082505.
11. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;**385**:e078378.
12. Rijk MH, Platteel TN, Geersing GJ, et al. 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. *Diagn Progn Res*. 2023;**7**(1):23.
13. World Organization of Family Doctors (WONCA). *ICPC-2: International Classification of Primary Care*. Oxford: Oxford University Press; 1998.
14. World Health Organization (WHO) Collaborating Centre for Drug Statistics Methodology. *ATC Classification Index with DDDs*. Oslo: WHO Collaborating Centre for Drug Statistics Methodology; 2022.
15. Smeets H, Kortekaas M, Rutten F, 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.
16. Amsterdam UMC. *Academisch Netwerk Huisartsgeneeskunde Amsterdam (ANHA)*. Available from: https://www.vumc.nl/research/overzicht/academisch-netwerk-huisartsgeneeskunde-amsterdam-umc-anha.html.

17. de Boer AR, de Groot MCH, Groenhof TKJ, et al. Data mining to retrieve smoking status from electronic health records in general practice. *Eur Heart J Digit Health*. 2022;**3**(3):437–44.
18. Drubbel I, De Wit NJ, Bleijenberg N, et al. Prediction of adverse health outcomes in older people using a frailty index based on routine primary care data. *J Gerontol A Biol Sci Med Sci*. 2013;**68**(3):301–8.
19. R Core Team. *R: A Language and Environment for Statistical Computing* [Internet]. R Foundation for Statistical Computing, Vienna, Austria; 2022. Available from: https://www.R-project.org/.
20. Bots SH, Groenwold RHH, Dekkers OM. Using electronic health record data for clinical research: a quick guide. *Eur J Endocrinol*. 2022;**186**(4):E1–6.
21. Gianfrancesco MA, Goldstein ND. A narrative review on the validity of electronic health record-based research in epidemiology. *BMC Med Res Methodol*. 2021;**21**(1):234.
22. Luijken K, Groenwold RHH, Van Calster B, Steyerberg EW, van Smeden M. Impact of predictor measurement heterogeneity across settings on the performance of prediction models: a measurement error perspective. *Stat Med*. 2019;**38**(18):3444–59.
23. Moore M, Stuart B, Lown M, et al. Predictors of adverse outcomes in uncomplicated lower respiratory tract infections. *Ann Fam Med*. 2019;**17**(3):231–8.
24. Bont J, Hak E, Hoes AW, et al. A prediction rule for elderly primary-care patients with lower respiratory tract infections. *Eur Resp J*. 2007;**29**(5):969–75.
25. 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.
26. van de Nadort C, Smeets HM, Bont J, et al. Prognosis of primary care patients aged 80 years and older with lower respiratory tract infection. *Br J Gen Pract*. 2009;**59**(561):261–5.
27. Lim WS, Van Der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax*. 2003;**58**(5):377–82.
28. National Institute for Health and Care Excellence. *Pneumonia: diagnosis and management (NG250)* [internet]. London: National Institute for Health and Care Excellence; 2025. Accessed September 6, 2025. Available at: https://www.nice.org.uk/guidance/ng250.
29. van Royen FS, Moons KGM, Geersing GJ, van Smeden M. Developing, validating, updating and judging the impact of prognostic models for respiratory diseases. *Eur Respir J*. 2022;**60**(3):2200250.
30. Van Calster B, Van Smeden M, Van Amsterdam W, et al. The enemies of reliable and useful clinical prediction models: a review of statistical and scientific challenges. *Annu Rev Stat Applic*. 2025;**13**.
31. Vickers AJ, Assel M, Dunn RL, et al. Guidelines for reporting observational research in urology: the importance of clear reference to causality. *Eur Urol*. 2023;**84**(2):147–51.
32. Carriero A, de Hond A, Cappers B, et al. Explainable AI in healthcare: to explain, to predict, or to describe? *Diagn Progn Res*. 2025;**9**:29.
33. Seinen TM, Fridgeirsson EA, Ioannou S, et al. Use of unstructured text in prognostic clinical prediction models: a systematic review. *J Am Med Inform Assoc*. 2022;**29**(7):1292–1302.
34. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *J Am Med Inform Assoc*. 2019;**26**(4):364–79.

SUPPLEMENTARY MATERIALS FOR CHAPTER 5

File S1 and S2 (R codes) are not printed in this thesis but will be available online.

Table S1. Details on candidate predictors and definitions

Candidate predictor	Definition or underlying registration(s)	Codes, if applicable		Data source		Values	Model(s)			
		ICPC	ATC	Development	Validation		Basic	CMD	Clinical	EHR
Demographics										
Age (years)	Age in years at day of LRTI diagnosis			JGPN	ANHA	40 – infinity (integer)	X	X	X	X
Sex (male, female)	Biological sex			JGPN	ANHA	Male, female	X	X	X	X
Cardiometabolic comorbidities										
Diabetes	Diabetes mellitus type 1 and 2	T90		JGPN	ANHA	Yes, no		X	X	X
Atrial fibrillation	Atrial fibrillation/flutter	K78		JGPN	ANHA	Yes, no		X	X	X
Cerebrovascular disease	CVA TIA	K90 K89		JGPN	ANHA	Yes, no		X	X	X
Coronary artery disease	Acute myocardial infarction Angina pectoris Other chronic/ischemic heart disease	K75 K74 K76		JGPN	ANHA	Yes, no		X	X	X
Heart failure	Congestive heart failure	K77		JGPN	ANHA	Yes, no		X	X	X
Venous thromboembolism	Pulmonary embolism Deep venous thrombosis	K93 K94.01		JGPN	ANHA	Yes, no		X	X	X
Peripheral artery disease	Intermittent claudication	K92.01		JGPN	ANHA	Yes, no		X	X	X

Table S1. Continued

Candidate predictor	Definition or underlying registration(s)	Codes, if applicable		Data source		Values	Model(s)			
		ICPC	ATC	Development	Validation		Basic	CMD	Clinical	EHR
Health care use										
Number of GP consultations ≤12 months	Number of GP consultations in the year prior to LRTI diagnosis			JGPN	ANHA	0 – infinity (integer)				X
Hospitalisation ≤12 months	Hospitalisation in the year prior to LRTI diagnosis			DHD	DHD	Yes, no			X	X
Vaccination status										
Influenza vaccination in previous year	Immunisation/preventive medicine of airways	R44	J07AL01 (<1 year)	JGPN	ANHA	Yes, no			X	X
Other comorbidities										
History of pneumonia	Pneumonia	R81		JGPN	ANHA	Yes, no			X	X
COPD or asthma	Emphysema/COPD Chronic bronchitis/bronchiectasis Asthma	R95 R91 R96		JGPN	ANHA	Yes, no			X	X
Dementia	Cognitive disorders Dementia	P20 P70		JGPN	ANHA	Yes, no			X	X
Non-dermatological malignancy	Hodgkin lymphoma	B72.01		JGPN	ANHA	Yes, no			X	X
	Non-Hodgkin lymphoma	B72.02								
	Leukaemia	B73								
	Hematologic/lymphatic malignancy NOS	B74								
	Benign hematologic/lymphatic Neoplasm	B75								
	Stomach malignancy	D74								

Table S1. Continued

Candidate predictor	Definition or underlying registration(s)	Codes, if applicable		Data source		Values	Model(s)			
		ICPC	ATC	Development	Validation		Basic	CMD	Clinical	EHR
	Colon/rectum malignancy	D75								
	Pancreatic malignancy	D76								
	Digestive tract malignancies NOS	D77								
	Malignancy of bronchus/lung	R84								
	Other airway malignancy	R85								
	Malignancy of kidney	U75								
	Malignancy of bladder	U76								
	Other malignancy of urinary tract	U77								
	Malignancy prostate	Y77								
	Other urologic malignancy	Y78								
	Malignancy of cervix uteri	X75								
	Malignancy of breast (female)	X76								
	Other gynaecological Malignancies	X77								
	Malignancy of unknown primary origin	A79								
	Malignancy of eye/adnexa	F74.01								
	Malignancy of ear	H75.01								
	Cardiovascular malignancy	K72.01								
	Musculoskeletal malignancy	L71.01								
	Neurologic malignancy	N74								
	Malignancy of thyroid	T71								

Table S1. Continued

Candidate predictor	Definition or underlying registration(s)	Codes, if applicable		Data source		Values	Model(s)			
		ICPC	ATC	Development	Validation		Basic	CMD	Clinical	EHR
Medication use[1]										
Antibiotic prescription within one month prior to LRTI diagnosis	Prescription within one month prior to LRTI diagnosis		J01	JGPN	ANHA	Yes, no			X	X
Immunosuppressants	Systemic glucocorticoids Predniso(lo)n ≥7.5mg/day		H02AB	JGPN	ANHA	Yes, no			X	X
	Immunosuppressants		L04A							
Antidepressants	Antidepressants		N06A	JGPN	ANHA	Yes, no			X	X
Inhalation medication	Long acting beta antagonists (LABA):			JGPN	ANHA	Yes, no			X	X
	Formoterol		R03AC13							
	Indacaterol		R03AC18							
	Olodaterol		R03AC19							
	Salmeterol		R03AC12							
	Long acting muscarine antagonists (LAMA):									
	Aclidinium		R03BB05							
	Glycopyrronium		R03BB06							
	Tiotropium		R03BB04							
	Umeclidinium		R03BB07							
	LABA+LAMA:									
	Aclidinium/formoterol		R03AL05							
	Glycopyrronium/formoterol		R03AL07							
	Indacaterol/glycopyrronium		R03AL04							
	Tiotropium/olodaterol		R03AL06							
	Umeclidinium/vilanterol		R03AL03							

Table S1. Continued

Candidate predictor	Definition or underlying registration(s)	Codes, if applicable		Data source		Values	Model(s)			
		ICPC	ATC	Development	Validation		Basic	CMD	Clinical	EHR
	Inhalation corticosteroids (ICS):									
	Beclomethasone		R03BA01							
	Budesonide		R03BA02							
	Fluticasone		R03BA05							
	ICS+LABA:									
	Formoterol/beclomethasone		R03AK08							
	Formoterol/budesonide		R03AK07							
	Salmeterol/fluticasone propionate		R03AK06							
	Vilanterol/fluticasone furoate		R03AK10							
	ICS+LABA+LAMA									
	Beclomethasone/formoterol/glycopyrronium		R03AL09							
	Fluticasone/umeclidinium/vilanterol		R03AL08							
	Formoterol/glycopyrronium/budesonide		R03AL11							
Algorithm-based candidate predictors										
Frailty index	Algorithm based on ICPC codes (Drubbel et al., J Gerontol A Biol Sci Med Sci, 2013)			JGPN	ANHA	1.0 – 1.00 (two digits)				X
Smoking status	Algorithm for text mining (de Boer et al., Eur Heart J Digit Health, 2022)			JGPN	ANHA	Current, former, never, no information				X

Abbreviations: ICPC, International Classification of Primary Care; ATC, anatomical therapeutic chemical; CMD, cardiometabolic disease; LRTI, lower respiratory tract infection; JGPN, Julius General Practitioners' Network; ANHA, Academic Network of General Practitioners Amsterdam; GP, general practitioner; DHD, Dutch Hospital Data; COPD, chronic obstructive pulmonary disease; NOS, not otherwise specified; mg, milligram; LABA, long-acting beta antagonist; LAMA, long-acting muscarine antagonist; ICS, inhalation corticosteroids.

[1] Medication was selected only when commonly used in a primary care setting and is based upon guidelines issued by the Dutch College of General Practitioners and the Dutch pharmacotherapeutic compass.

Table S2. Modelling strategy and apparent model performance in development cohort of developed models

Model	Predictors with varying modelling strategies				Candidate predictors	#RC	Model performance (95% CI)		
	Age	CMD	Frailty index	Number of GP-consultations <1 year			c-statistic	Calibration intercept	Calibration slope
Basic 1	L	-	-	-	Age, sex, and interaction term	4	0.61 (0.59 – 0.63)	0.00 (-0.07 – 0.07)	1.04 (0.86 – 1.23)
Basic 2	RCS	-	-	-		8	0.62 (0.60 – 0.64)	0.00 (-0.07 – 0.07)	1.04 (0.86 – 1.23)
CMD 1	L	Cumulative[a]	-	-	Basic model + diabetes, heart failure, ischaemic heart disease, CVA, TIA, PE, DVT, atrial fibrillation, and peripheral artery disease	6	0.63 (0.61 – 0.65)	0.00 (-0.07 – 0.07)	1.04 (0.88 – 1.20)
CMD 2	L	Separate	-	-		11	0.63 (0.61 – 0.65)	0.00 (-0.07 – 0.07)	1.05 (0.89 – 1.20)
Clinical	L	Cumulative[a]	-	-	CMD model + LRTI diagnosis code, hospitalisation <1 year, history of pneumonia, malignancy, COPD/ asthma, or dementia, influenza vaccination <1 year, use of immunosuppressants, inhalation medication, or antidepressants, and antibiotic use <1 month	17	0.73 (0.71 – 0.74)	0.00 (-0.07 – 0.07)	1.02 (0.93 – 1.11)
EHR 1	L	Cumulative[a]	L	L	Clinical model + smoking status, frailty index, and number of GP consultations <1 year	22	0.74 (0.72 – 0.75)	0.00 (-0.07 – 0.07)	1.04 (0.95 – 1.12)
EHR 2	L	Cumulative[a]	RCS	RCS		26	0.74 (0.72 – 0.75)	0.00 (-0.07 – 0.07)	1.03 (0.94 – 1.11)

[a] Modelling of CMD as cumulative predictor: 0, 1, or ≥2 CMD based on medical history

CI, confidence interval; CMD, cardiometabolic disease; GP, general practitioner; RC, regression coefficients; L, linear; RCS, restricted cubic splines; CVA, cerebrovascular accident; TIA, transient ischaemic attack; PE, pulmonary embolism; DVT, deep venous thrombosis; COPD, chronic obstructive pulmonary disease; LRTI, lower respiratory tract infection; EHR, electronic health records.

Table S3. Linear predictors and regression coefficients of all developed models

Predictor	Regression coefficients						
	Basic 1	**Basic 2**	**CMD 1**	**CMD 2**	**Clinical**	**EHR 1**	**EHR 2**
Intercept	-4.298	-4.710	-4.043	-4.044	-4.618	-4.047	-4.060
Age (linear)	0.029	-	0.023	0.023	0.015	0.005	0.005
Female sex	0	0	0	0	0	0	0
*Age (linear)*female sex*	-0.003	-	-0.002	-0.003	-0.001	-0.002	-0.002
RCS(age, c(43, 58, 70, 86))	-	0.035	-	-	-	-	-
RCS(age, c(43, 58, 70, 86))'	-	0.001	-	-	-	-	-
RCS(age, c(43, 58, 70, 86))"	-	0	-	-	-	-	-
*RCS(age, c(43, 58, 70, 86))*female sex*	-	-0.0002	-	-	-	-	-
*RCS(age, c(43, 58, 70, 86))' *female sex*	-	-0.0004	-	-	-	-	-
*RCS(age, c(43, 58, 70, 86))" *female sex*	-	-0.067	-	-	-	-	-
No. of CMD = 1	-	-	0.245	-	0.115	0.016	0.021
No of. CMD ≥ 2	-	-	0.583	-	0.344	0.143	0.147
Diabetes	-	-	-	0.215	-	-	-
Heart failure	-	-	-	0.399	-	-	-
Ischaemic heart disease	-	-	-	0.118	-	-	-
CVA or TIA	-	-	-	0.464	-	-	-
PE or DVT	-	-	-	0.392	-	-	-
Atrial fibrillation	-	-	-	0.035	-	-	-
Peripheral artery disease	-	-	-	0.114	-	-	-
LRTI diagnosis = pneumonia	-	-	-	-	1.015	1.009	1.023
Hospitalisation <1 year	-	-	-	-	0.922	0.818	0.822
History of pneumonia	-	-	-	-	-0.137	-0.136	-0.149
Malignancy	-	-	-	-	0.393	0.338	0.432

Table S3. Continued

Predictor	Regression coefficients						
	Basic 1	**Basic 2**	**CMD 1**	**CMD 2**	**Clinical**	**EHR 1**	**EHR 2**
COPD or asthma	-	-	-	-	0.112	0.018	0.021
Dementia	-	-	-	-	0.222	0.139	0.145
Influenza vaccination <1 year	-	-	-	-	0.076	0.045	0.049
Immunosuppressant use	-	-	-	-	0.182	0.110	0.128
Inhalation medication use	-	-	-	-	0.082	0.040	0.041
Antibiotics <1 month	-	-	-	-	0.335	0.267	0.272
Antidepressant use	-	-	-	-	0.254	0.158	0.167
Smokingstatus = former	-	-	-	-	-	-0.074	-0.092
Smoking status = never	-	-	-	-	-	-0.186	-0.203
Smoking status = no information	-	-	-	-	-	-0.107	-0.124
Frailty index (linear)	-	-	-	-	-	1.231	-
No. of GP-consultations <1 year (linear)	-	-	-	-	-	0.008	-
RCS(frailty index, c(0, 0.001, 0.28, 0.48))	-	-	-	-	-	-	1.224
RCS(frailty index, c(0, 0.001, 0.28, 0.48))'	-	-	-	-	-	-	0
RCS(frailty index, c(0, 0.001, 0.28, 0.48))"	-	-	-	-	-	-	0
RCS(No. of GP-consultations <1 year, c(0, 5, 12, 33))	-	-	-	-	-	-	0.008
RCS(No. of GP-consultations <1 year, c(0, 5, 12, 33))'	-	-	-	-	-	-	0
RCS(No. of GP-consultations <1 year, c(0, 5, 12, 33))"	-	-	-	-	-	-	0

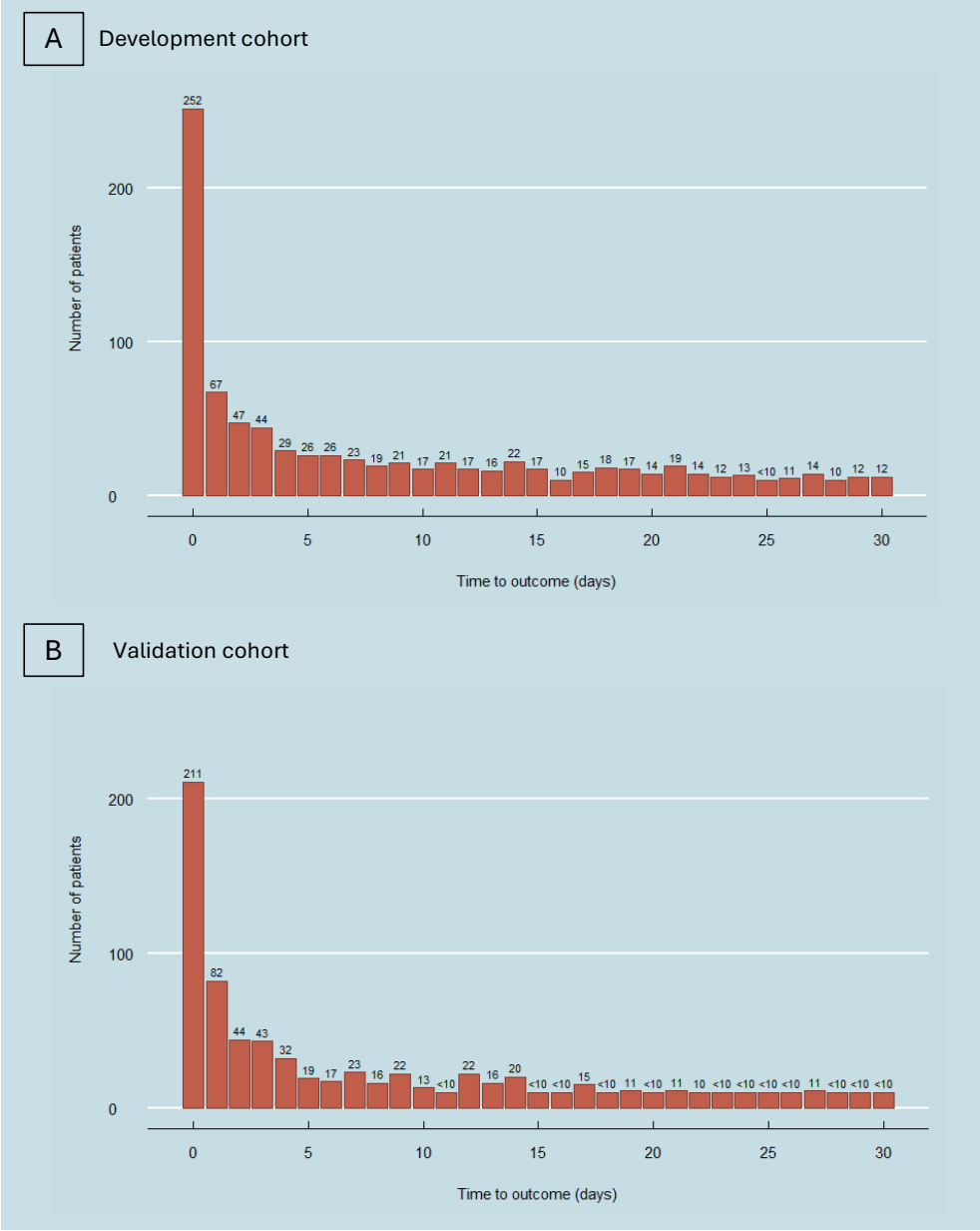


Figure S4. Days between LRTI diagnosis and hospitalisation or mortality in the model development cohort (A) and validation (B) cohort

Due to output restrictions of CBS, which do not allow to present data based on less than 10 individual observations, values below 10 have been indicated as '<10' in this figure with a column size corresponding to 10.

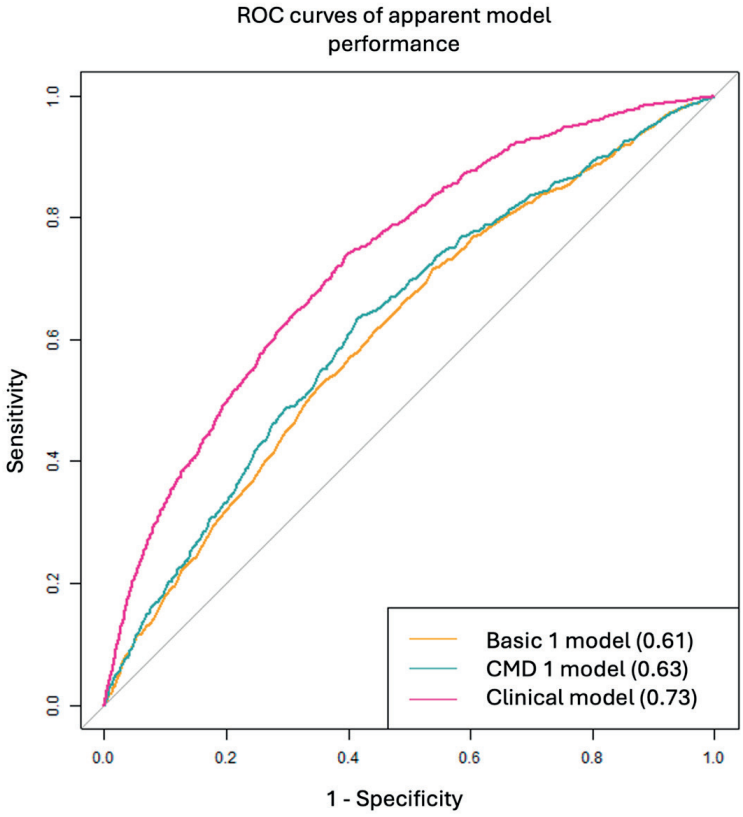


Figure S5. Visual representation of apparent model discrimination (c-statistic) performance of basic 1, CMD 1, and clinical models

ROC, receiver operator characteristic; CMD, cardiometabolic disease.

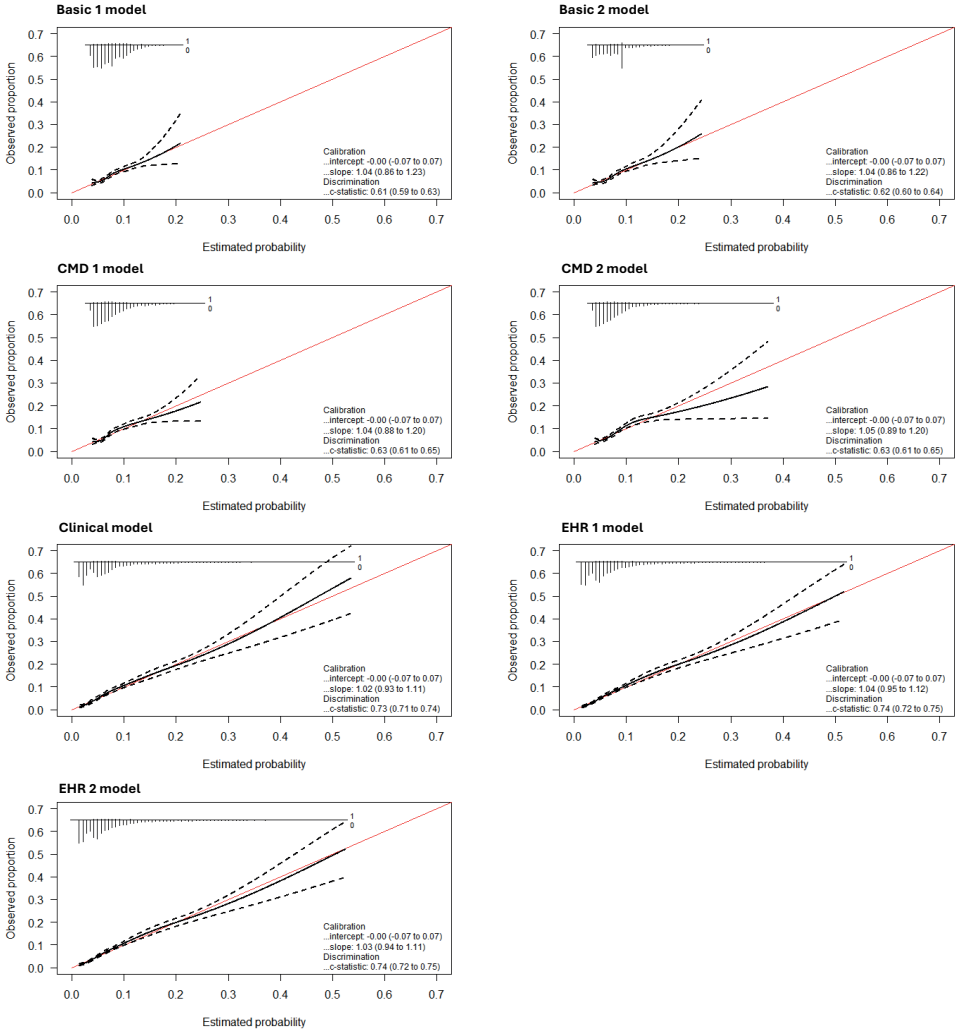


Figure S6. Flexible calibration curves of apparent model performance of all developed models

Apparent performance measures and flexible calibration curves in the development cohort of all developed models.

CMD, cardiometabolic disease; EHR, electronic health records.

6

Comparison of local large language models for extraction of signs and symptoms data from electronic health records

Isa Spiero*
Merijn H. Rijk*
Matthew A. Scheeres
Frans H. Rutten
Geert-Jan Geersing
Tamara N. Platteel
Karel G.M. Moons
Lotty Hooft
Johanna A.A. Damen
Roderick P. Venekamp
Artuur M. Leeuwenberg

* Authors contributed equally
Accepted for publication in PLoS One

Abstract

Electronic health records (EHRs) provide a large source of data that can be used for research purposes. Extraction of information from unstructured clinical notes in EHRs can be automated by large language models (LLMs). Although LLMs are promising for this task, challenges remain in reliable application of LLMs to EHR, including the lack of development and validation for languages other than English. Here, we identified Dutch LLMs and compared their performance in a case study. We selected the MedRoBERTa.nl and RobBERT models based on local applicability, Dutch language compatibility, and model architecture. We evaluated their performance in a case study on the extraction of signs and symptoms from comprehensive Dutch primary care EHRs of patients with a lower respiratory tract infection. Using manually annotated clinical notes, models were trained as direct and prompt-based classifiers with varying amounts of training samples. Performance was expressed by precision, recall, and F1-score. The MedROBERTa.nl and RobBERT models showed good performance as direct classifiers, with a macro-averaged F1-score of 0.74 (range 0.56–0.87) and 0.69 (range 0.46–0.86) using 1600 training samples, respectively. The prompt-based classifiers performed worse with F1-scores of 0.08 (range 0.02–0.30) and 0.08 (range 0.02–0.22), respectively. In general, performance of the models was negatively affected by class imbalance and missingness of signs and symptoms. A minimum of 800 annotated training samples were required to obtain sufficient performance. The selected LLMs showed good performance as direct classifiers in extracting signs and symptoms from Dutch primary care EHRs. However, prompt-based models require performance improvement by further prompt engineering, and caution is warranted with imbalanced or partially missing EHR data.MedROBERTa.nl and RobBERT models, used as direct classifiers, can be considered for clinical research to extract information from clinical notes from Dutch primary care EHRs, potentially reducing manual annotation time and accelerating real-world research and evidence generation.

INTRODUCTION

Electronic health records (EHRs) are increasingly used in clinical research as they form a valuable source of information on patients and health care use.[1] Generally, EHRs contain structured data, such as disease codes and demographics, and unstructured data, such as clinical notes.[2] While structured data can easily be retrieved, a large amount of valuable information is described by clinicians in unstructured data as narrative, free text in clinical notes. Manual screening of these clinical notes to retrieve information for research purposes is laborious and costly.

To facilitate extraction of unstructured data from clinical notes, natural language processing (NLP) techniques have been developed to automate information extraction for a wide range of EHR extraction tasks.[1,3] Current growth in NLP techniques for extraction mainly involves the fast-paced development of large language models (LLMs), which are, in contrast to more simple NLP techniques, able to process texts in a more context-aware manner.[4] These NLP techniques have already shown promising performance in extracting data from clinical notes in a wide range of studies and extraction targets such as specific clinical conditions[5], chronic diseases[6], symptoms[7], or social determinants of health[8].

Given the potential of LLMs to automatically extract unstructured EHR data, several areas of clinical research – such as prognostic prediction model studies – could benefit. Prognostic prediction modelling could be improved by applying LLMs to automatically extract large numbers of eligible patients, (candidate) predictors, and outcomes from EHRs. Using LLMs, study sample sizes can thus be increased and predictors reported or hidden in unstructured data can be included. This can subsequently improve the performance and the generalizability of the prediction model.[9]

Still several challenges remain for reliable extraction of data from clinical notes in EHRs by LLMs. For example, EHR data are not developed for research purposes and therefore the information in clinical notes tends to contain abbreviations, grammatical errors, ambiguous language and incomplete or selective information, which could negatively impact the LLM's extraction performance.[10,11] In addition, training and validating LLMs for extraction often requires manually annotated clinical notes, which are costly to obtain and therefore only available in small amounts.[10] Furthermore, the best performing LLMs are only available online and therefore not locally applicable in secured environments where EHR data are stored, which is crucial for the confidential data present in EHRs. And finally, LLMs are mostly developed for the English language[12], but the majority of patients worldwide receive care in other languages, creating a need for locally trained LLMs. Although development is growing for other languages[13], relatively little is known about the performance of LLMs for most languages including the application of LLMs to EHR notes in Dutch language.

We address these issues by exploring the performance of locally applicable LLMs for extraction of data from Dutch EHRs. We gain insight on the potential of LLMs for the extraction of EHR data to be used in, for example, prediction model research. First, we performed a literature search for locally applicable LLMs available for Dutch clinical texts. Subsequently, we applied the selected models to extract signs and symptoms from EHR data of adult patients presenting to primary care with a lower respiratory tract infection (LRTI), which are the target predictors for a prediction modelling case study described elsewhere.[14] We additionally estimated the sample size of the training data needed for sufficient LLM performance.

MATERIALS & METHODS

Case study: predicting 30-day hospitalisation or mortality in primary care LRTI patients

We evaluated the performance of Dutch LLMs in a case study aiming to extract nine signs and symptoms from EHR data of adult LRTI patients presenting to their general practitioner (GP). These signs and symptoms, summarised in Table 1, are candidate predictors for a prognostic modelling study aiming to predict individual risk of 30-day hospitalisation or mortality using EHR data of LRTI patients as described elsewhere.[14] However, data on signs and symptoms are recorded in unstructured clinical notes and would require manual extraction for incorporation into such a prediction model. As this would be infeasible on large scales, LLMs could be implemented to automatically extract signs and symptoms from EHR clinical notes from large numbers of patients.

Data

We used data from a sample of 2,000 patients aged ≥40 years presenting to their GP with an LRTI between 2016 and 2019 in the region of Utrecht, The Netherlands.[15] The data were derived from the comprehensive Julius General Practitioners' Network[16], and contains both structured (e.g. demographics, medical history, medication use) and unstructured (e.g. descriptions of signs, symptoms, physical examination) data of the first GP consultation within an LRTI episode. Unstructured data was covered in 'subjective' (patient reported), 'objective' (GP reported), 'evaluation', and 'plan' text fields. We used the subjective and objective texts as our target for extraction, hereafter called 'clinical notes'. Most clinical notes had between 5 and 100 tokens (words) with an average number of 63 tokens and a few large notes containing up to 240 tokens.

Target variables

The target variables for extraction consisted of both patient and GP reported LRTI signs and symptoms (Table 1). The clinical notes were manually annotated for these signs and symptoms in a previous study by *Rijk et al*.[15] Each sign or symptom was annotated for each

patient as (a) 'present', (b) 'absent', or (c) 'not reported'. These annotations functioned as the reference standard for the LLM. The LRTI signs and symptoms show a strong variation in class imbalance/prevalence (i.e., some occur more rarely across patients than others) as well as missingness (i.e., some signs and symptoms are often not reported) (Table 1).

Table 1. Recording of signs and symptoms in clinical notes from EHR of 2,000 primary care LRTI patients, that are targets for extraction by the LLMs, from *Rijk et al*. (2024).

	Overall	Recorded as 'present" [overall], %	Recorded as 'absent' [overall], %	Not recorded, %
Patient reported [*Dutch*]				
Cough [*Hoesten]*	76.8	98.2 [*75.4*]	1.8 [*1.4*]	23.2
Fever [*Koorts]*	58.3	43.1 [*25.1*]	56.9 [*33.2*]	61.7
Shortness of Breath [*Kortademigheid*]	51.4	76.0 [*39.1*]	24.0 [*12.3*]	58.6
Sputum [*Sputum*]	26.2	89.5 [*23.4*]	10.5 [*2.8*]	73.8
Chest pain [*Pijn op de borst*]	14.7	72.4 [*10.6*]	27.6 [*4.1*]	85.3
Chills [*Koude rillingen*]	4.9	92.9 [*4.6*]	7.1 [*0.3*]	95.1
Patient or GP reported [*Dutch*]				
Confusion [*Verwardheid*]	3.1	31.1 [*1.0*]	68.9 [*2.1*]	96.9
GP reported [*Dutch*]				
Crackles upon auscultation [*Crepitaties*]	84.3	27.5 [*23.2*]	32.5 [*61.1*]	15.7
Ill appearance [*Zieke indruk]*	30.3	35.3 [*10.7*]	64.7 [*19.6*]	69.7

GP, general practitioner; LRTI, lower respiratory tract infections; EHR, electronic health records; LLM, large language model.

Possible approaches of LLMs

Generally, there are three approaches to implement LLMs for extraction of EHR data:

- Training from scratch: This approach involves completely building the LLM from the ground up, including designing the architecture of the LLM (in terms of layers, attention heads, embedding sizes, etc.). This allows for complete freedom in designing and altering the model which enables to tailor the model to the specific task at hand, at the cost of (generally) very high computational power needed as well as vast amounts of training data. In our setting, this computation was not feasible.

- Using a pretrained model: This is a simple approach in which a model is used that had previously been trained on either a general dataset or a domain-specific dataset. The pre-trained model is used as-is and is directly applied to perform the desired task. Although this approach requires minimal effort and is the most

straightforward, the model might not perform optimally as it is not optimized for the specific data and task at hand.

- Fine-tuning a pretrained model: This approach involves using a pretrained model which is then fine-tuned for the specific task at hand by further training the model with the available data. This allows the model to adapt to the desired task and does not require an extremely large dataset for training. For these reasons, extending a pretrained model is often the first choice for LLM-related research in domain-specific appliances, such as extraction from Dutch EHR data.

Possible implementations of LLMs for classification

For all three approaches described above, the LLMs can subsequently be used for classification tasks in several ways, of which two common methods are described below:

- Direct classification: In direct classification, a model learns from a set of training samples (here, clinical notes) to directly discriminate between a fixed set of predefined classes (here, 'absent', 'present', and 'not reported'). After training, the model can be applied to other samples to predict to which class the new sample would belong (here, to predict for a clinical note if a given sign or symptom is 'present', 'absent', or 'not reported').

- Prompt-based classification: In prompt-based classification, the model is not directly trained for the classification task, but a so-called conversational LLM is used and prompted with a natural language prompt that asks the LLM using human language to classify the text that it is given. For these types of models, one of the following setup types can be applied:

    - Zero-shot classification: The model classifies based solely on a task description.

        Example prompt: "Classify whether the patient in this text has fever or not: [text]"

    - One-shot classification: The model classifies based on a task description and one labelled sample.

        Example prompt: ""While the patient coughs constantly, there is no sign of elevated temperature" => "No fever". Classify whether the patient in this text has a fever or not: [text]"

- Few-shot or multi-shot classification: The model classifies based on a task description and two or more labelled samples.

  Example prompt: ""While the patient coughs constantly, there is no sign of elevated temperature" => "No fever". "Fever was confirmed after measuring a temperature of 39.2 degrees Celsius" => "Fever". Classify whether the patient in this text has a fever or not: [text]"

Our selection, approach, and implementation of LLMs

We first performed a search on Hugging Face (an open-source platform for AI and machine learning models) to find and select suitable pre-trained LLMs. Our selection criteria for LLMs included (1) locally applicability, (2) pre-trained, (3) a sufficient and locally computable number of parameters, (4) compatibility with Dutch language, and (5) preferably familiarized with medical texts (see the entire selection procedure in Appendix A). We selected two models fulfilling the criteria, MedRoBERTA.nl[17] and RobBERT[18] (see further model details in Appendix A).

We approached our task of extracting signs and symptoms from Dutch EHR data by fine-tuning the selected pre-trained LLMs. We implemented the models in two different ways for each model: as a direct classifier and as a prompt-based classifier, leading to the following four model configurations that were evaluated:

- Direct classifier – MedRoBERTa.nl
- Direct classifier – RobBERT
- Prompt-based – MedRoBERTa.nl
- Prompt-based – RobBERT

Implementing LLMs as direct classifiers

For direct classification, each of the models was trained using the manually annotated dataset to discriminate between the three classes of 'present', 'absent', and 'not reported'. We trained each of the models for 5 epochs, a number which we empirically determined beforehand by assessing the training and evaluation loss of MedRoBERTa.nl after 10 epochs. To additionally evaluate the effect of training sample size on the performance of the direct classifier LLMs, we varied the number of training samples by 200, 400, 600, 800, 1,000, 1,200, 1,400, and 1,600. The performance of the models was evaluated by 5-fold cross-validation in which we used fixed folds across each experiment and kept the same 400 samples as the validation set.

Implementing LLMs as prompt-based classifiers

Before we could use the models as prompt-based classifiers, we had to convert them to process Dutch prompts and apply those to Dutch EHR data. The following steps were taken to convert the original models to prompt-based classifiers:

1. First, we transformed the models to be able to also generate text, as the prompt-based classifiers predict labels through next token prediction. Therefore, we combined the encoder structure of the original models with the base multilingual BERT decoder using a Hugging Face Transformers EncoderDecoderModel2 architecture.

2. Second, all model parameters (including the decoder parameters) were fine-tuned for three epochs using a dataset of over 100,000 patient-doctor conversations (the HealthCareMagic-100k dataset[19]) which was translated beforehand to Dutch using the Google Translate API.

3. After fine-tuning, a prompt could be used in which the multiclass prediction task is instructed to the model by providing a sample and an instruction to classify based on some provided samples.

The architecture of our prompts is shown in Box 1 in which the amount of input samples in the prompt for these models would be varied between 1, 2 or 3 samples.

Box 1. The original Dutch prompt and the translated English prompt that was used for the prompt-based models to extract the signs and symptoms from the Dutch EHRs.

Original (in Dutch)	Translated from original (to English)
"Classificeer de volgende teksten op basis van de aanwezigheid van het symptoom 'hoesten' als volgt:	*"Classify the following texts based on the presence of the symptom 'cough' as follows:*
Als 'hoesten' wordt vermeld als voorkomend bij de patiënt, label het als 'Aanwezig.'	*If 'cough' is mentioned as present for the patient, label this as 'Present'*
Als 'hoesten' wordt vermeld als niet voorkomend bij de patiënt, label het als 'Niet aanwezig.'	*If 'cough' is mentioned as not present in the patient, label this as 'Not present'*
Als 'hoesten' helemaal niet wordt vermeld in de tekst, label het als 'Niet vermeld.'	*If 'cough' is not mentioned in the text, label this as 'Not mentioned'.*
Voorbeeld(en): *<insert samples here>*	*Example(s):* *<insert samples here>*
Print exclusief het label bijbehorende aan de volgende tekst: *<insert to-be-classified text here>"*	*Print only the label corresponding to the following text:* *<insert to-be-classified text here>"*

[1]https://huggingface.co/datasets/lavita/ChatDoctor-HealthCareMagic-100k
[2]https://huggingface.co/docs/transformers/v4.48.2/en/model_doc/encoder-decoder#transformers.EncoderDecoderModel

Performance evaluation

In total, separate LLMs were developed and evaluated for each of the 9 target signs/symptoms and for each of the training sample sizes, leading to a total of 2 (direct classifiers) * 9 (signs and symptoms) * 8 (training sample sizes) + 2 (prompt-based classifiers) * 9 (signs and symptoms) * 3 (training sample sizes) = 198 models.

To evaluate and compare the performance of the LLMs, we computed the following metrics and averaged these across the 5-fold cross-validations:

- Precision (P): Precision is the ratio of true positive predictions to the total amount of positive predictions (a combination of true positives (TP) and false positives (FP)), i.e. the relative amount of correctly predicted positive samples, and is calculated by by $P = \frac{TP}{TP+FP}$;

- Recall (R): Recall quantifies how well the model is able to identify positive cases, i.e. all cases where a symptom is present in the text. Recall is calculated using the formula $R = \frac{TP}{TP+FN}$;

- F1-score (F1): The F1-score provides a harmonic mean of precision and recall and functions as a balanced measure of both metrics. The F1-score is calculated using the formula $F1 = 2\frac{P \times R}{P+R}$.

The performance metrics were computed per-class and as macro-average across classes.

Implementation details

The experiments were conducted in Python v3.11.5 on a virtual computer with 32 GB of RAM, an Intel(R) Xeon(R) Platinum 8272CL CPU, using an NVIDIA T4 GPU.

RESULTS

Model comparison: MedRoBERTa.nl versus RobBERT

Direct classification

As direct classifiers, the MedRoBERTa.nl and RobBERT models performed similarly in classifying the presence of LRTI signs and symptoms from the Dutch clinical notes (Fig 1a). For the MedRoBERTa.nl model at the largest number of 1,600 training samples used, the obtained recall in direct classification was on average 0.72 (range 0.54–0.86), the precision 0.76 (range 0.60–0.90), and the F1-score 0.74 (range 0.56–0.87) across the signs and symptoms. For the RobBERT model at the largest number of 1,600 training samples used, the obtained recall in direct classification was on average 0.69 (range 0.45–0.84), the

precision 0.71 (range 0.48–0.89), and the F1-score 0.69 (range 0.46–0.86) across the signs and symptoms (Appendix B Table A3).

Prompt-based classification

When the MedRoBERTa.nl and RobBERT models were implemented as prompt-based classifiers, the performance was significantly lower compared to direct classification, and less similar between the two models (Fig 1b). At the largest number of three samples used, for the MedRoBERTa.nl model the obtained recall in prompt-based classification was on average 0.33 (range 0.32–0.34), precision 0.15 (range 0.01–0.34), and F1-score 0.08 (range 0.02–0.30), while for the RobBERT model the obtained recall in prompt-based was on average 0.33 (range 0.31–0.33), precision 0.06 (range 0.01–0.21), and F1-score 0.08 (range 0.02–0.22) across the signs and symptoms (Appendix B Table A4). For most signs and symptoms, the prompt-based models conflated to predict one outcome for all data points. Hence, the recall values of 1.0 for one of the outcomes and close to 0 for the others, indicating poor model performance (Fig 1b).

Effect of sign and symptom reporting and class imbalance

Direct classification

The level of reporting of a sign or symptom across the clinical notes significantly impacted the MedRoBERTa.nl and RobBERT models as direct classifiers in their performance. For example, in classifying two of the symptoms that were rarely reported, 'chills' (reported in 4.9%) and 'confusion' (reported in 3.1%), the models performed worse compared to more commonly reported signs and symptoms (Fig 1a). Moreover, the degree of class imbalance (i.e., the balance between 'present' and 'absent') also seemed to affect the performance of the models. For example, the extraction of the well-balanced symptoms of 'fever' (43.1% 'present' if recorded), 'ill appearance' (35.3% 'present' if recorded), and 'crackles upon auscultation' (27.5% 'present' if recorded) showed the highest performance with F1-scores up to 0.87 compared to less balanced symptoms (Fig 1a).

Prompt-based classification

In the prompt-based classification, the RobBERT model performed better at classifying 'fever' while the MedRoBERTa.nl model performed better at classifying 'crackles upon auscultation' (Fig 1b). For the other signs and symptoms, there were no marked differences between the two models, but there were some small overall model-independent differences between the signs and symptoms. However, those differences were not related to the level of recording of those signs and symptoms in the clinical notes or to the balance in positive and negative recordings (Fig 1b).

Effect of training sample size

Direct classification

In direct classification the performance of both models generally increased with an increasing number of training samples (Fig 1a). Recall ranged from 0.33–0.49 (MedRoBERTa.nl) and 0.26–0.38 (RobBERT) with 200 training samples to 0.54–0.86 (MedRoBERTa.nl) and 0.45–0.84 (RobBERT) with 1,600 training samples, precision ranged from 0.23–55 (MedRoBERTa.nl) 0.23–0.41 (RobBERT) with 200 training samples to 0.60–0.90 (MedRoBERTa.nl) and 0.48–0.89 (RobBERT) with 1,600 training samples, and F1-score ranged from 0.27–0.47 (MedRoBERTa.nl) and 0.26–0.38 (RobBERT) with 200 training samples to 0.56–0.87 (MedRoBERTa.nl) and 0.46–0.86 (RobBERT) with 1,600 training samples (Appendix B Table A3). For both models, the classification performance for some symptoms, such as 'chills' and 'confusion', was generally low and, especially for 'confusion', did not increase with more training samples. Generally, models' performance reached a plateau by a minimum of 800 training samples (Fig 1a).

Prompt-based classification

When implementing the MedRoBERTa.nl and RobBERT models as prompt-based classifiers, the number of samples in the prompts (i.e., 1, 2, or 3) did not affect the performance of the models (Fig 1b). This was similar for both models and across all signs and symptoms.

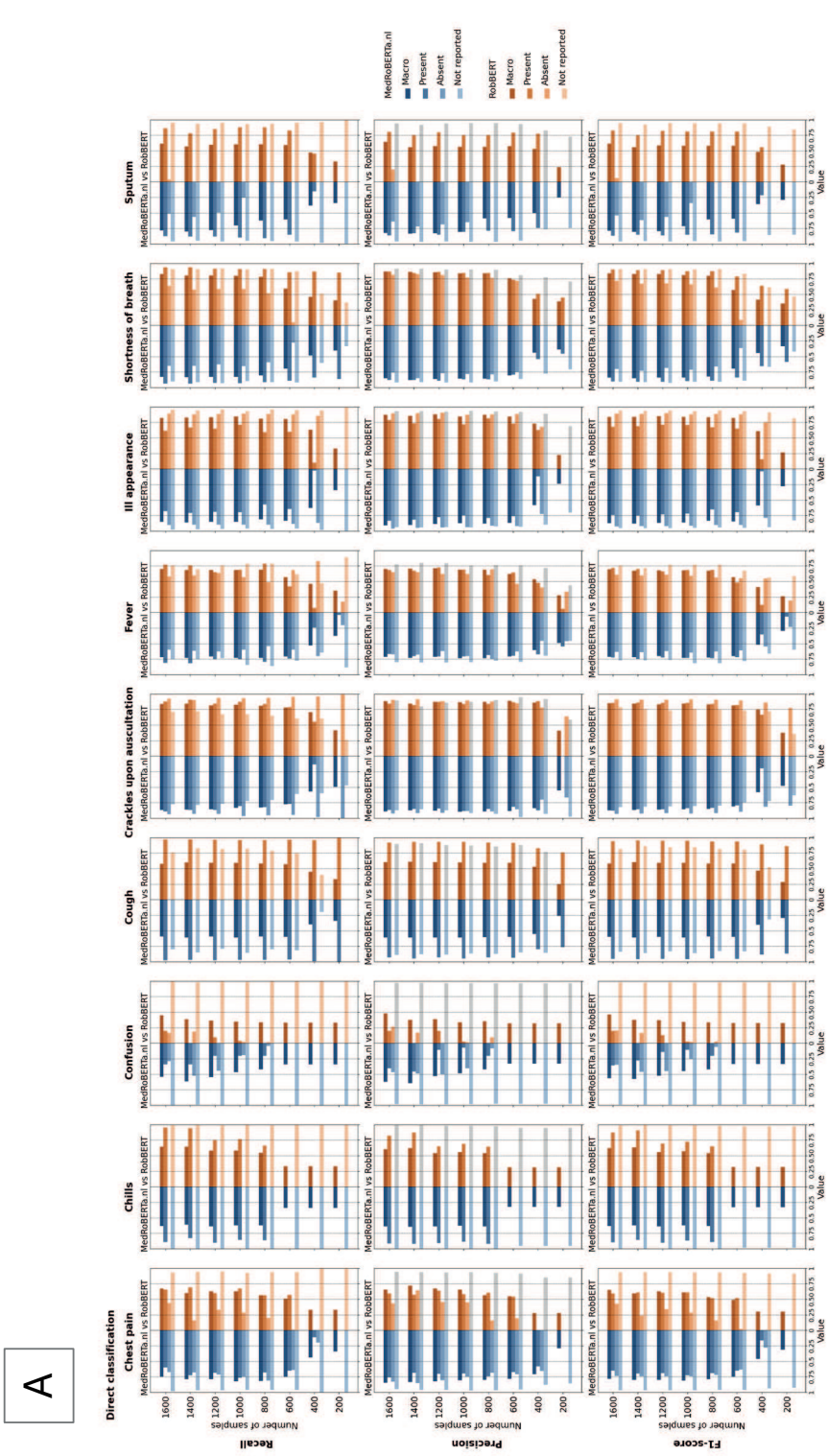
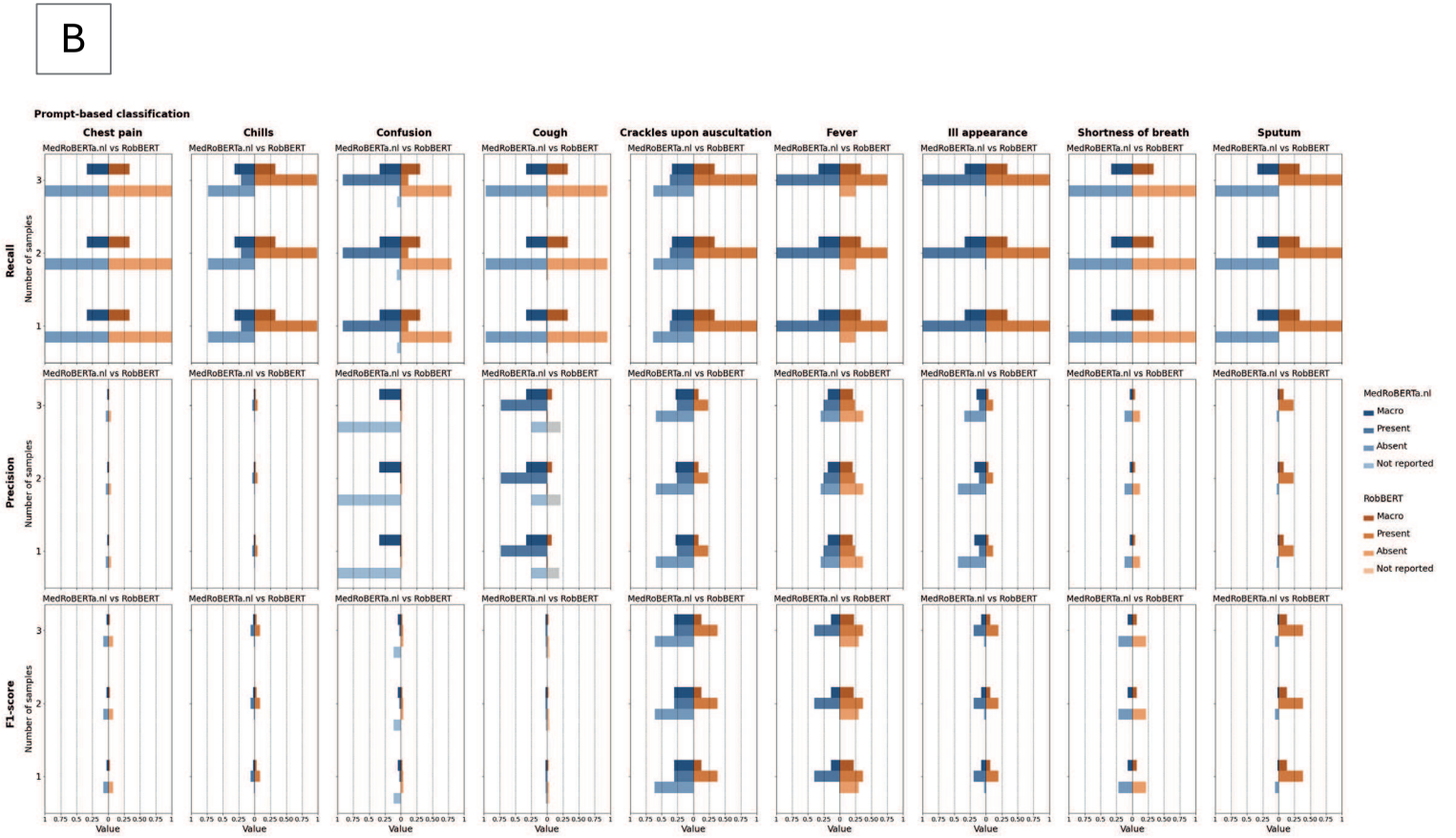


Figure 1. The performance of the MedRoBERTa.nl and RobBERT models in classifying signs and symptoms related to LRTIs based on clinical notes in Dutch primary care EHR.



The performance is expressed in terms of the macro-averaged and per-class ('present', 'absent', or 'not reported') recall, precision, and F1-score. The two models were implemented as (a) direct classifier trained on 200, 400, 600, 800, 1,000, 1,200, 1,400 or 1,600 training samples and (b) prompt-based classifier trained with 1, 2, or 3 samples.

LRTI, lower respiratory tract infection; EHR, electronic health records.

DISCUSSION

We evaluated the performance of two LLMs for the Dutch language, RobBERT and MedRoBERTa.nl, for extracting signs and symptoms of LRTI patients from Dutch EHR data. Overall, the models performed similarly, but variation was observed across implementation methods (direct classification outperformed prompt-based classification), across signs and symptoms that were extracted (higher performance for signs or symptoms with limited missingness and class-imbalance), and across the sample size of the training data. Both models showed good performance when implemented as direct classifiers and trained with sufficient data (at least 800 samples) with F1-scores reaching up to 0.87.

The MedRoBERTa.nl and RobBERT models showed similar overall performance although we anticipated that the MedRoBERTa.nl model – which was pre-trained on domain-specific medical data – would outperform RobBERT for the domain-specific task of this study.[20] However, the RobBERT model was originally trained on larger collections of data[18], which may explain why this model – not specifically trained on medical data – performed similarly as MedRoBERTa.nl. This may imply that LLMs pre-trained with large amounts of general domain data may equally suffice in extracting medical information from EHR data compared to LLMs trained with smaller amounts of medical domain data. Contrary, in the study by *Muizelaar et al*.[21] MedRoBERTa outperformed all other tested models including RobBERT in classifying reports, suggesting that the performance of models is also context or task dependent.

For both LLMs, implementation as prompt-based classifier showed significantly worse performance compared to the direct classifier method. Although conceptually interesting as they only require an instruction and thus avoid the need for costly manually annotated data, the performance of the prompt-based classifiers in this study was low. This is possibly due to the limited size of the prompt and the numbers and/or content of the examples in the prompt, or to the fact that these models were not trained as an encoder-decoder model from the ground up. In a similar clinical note classification task, incorporation of a reasoning step as a part of the prompt, in which the model is explicitly asked to output its stepwise reasoning, did increase performance.[22] Further engineering to optimize prompts may thus be needed to improve the performance of the prompt-based classifiers.[23,24]

Substantial difference in model performance was also observed across the nine signs and symptoms that were extracted by the LLMs. For some, such as 'crackles upon auscultation', 'ill appearance', and 'shortness of breath', the direct classifiers showed promising performance, while for others the LLMs performed significantly worse. The differences in performance across signs and symptoms were likely related to class imbalance and missingness (non-reporting), with worse performance for more imbalanced and/or less often

reported signs and symptoms. This, however, will remain a problem inherent to EHR data, despite mitigation methods.[25] In addition, the variation in performance across the signs and symptoms may also be partly explained by the amount of variation in descriptions of signs or symptoms that clinicians use. For example, 'ill appearance' and 'confusion' may be reported in a wider variety of descriptions than other signs and symptoms, although we did not further quantify this nor explored the potential effect on model performance.

As expected, for most signs and symptoms the performance increased with increasing sample sizes of training data for direct classifiers, while the prompt-based classifiers showed no difference in performance between the number of training samples. We found that generally a minimum of 800 training samples is needed to obtain good model performance, with F1-scores up to 0.87 for some of the signs or symptoms for the MedRoBERTa.nl model. This implies that, with manual data annotation being laborious and resource-consuming, a relatively limited amount of training samples may suffice to train LLMs for an EHR extraction task.

Strengths and limitations

This study is one of the first to evaluate and compare the performance of LLM in extracting signs and symptoms from Dutch unstructured EHR data. Further strengths of this study include the thorough search for locally applicable Dutch LLMs, the comparison between direct and prompt-based classifying models, and the large amount of manually annotated training samples which allowed us to determine the minimally required sample size of training data. Nevertheless, some limitations should be considered when interpreting these results. First, although we performed a careful selection with reasonable selection criteria, we only tested two models in our study. The performance of other models, especially non-local models that could not be implemented in our EHR setting, could not be determined. Also, larger models, which could not be run in our environment, hold promise to significantly improve performance.[26,27] Second, we did not extensively engineer our prompts, which likely explains why the prompt-based classifiers underperformed. We did not further elaborate on optimal prompt engineering in this study as our focus was on the exploration and comparison of currently available models for Dutch EHR data extraction. Lastly, given the large variability in performance between the signs and symptoms, the insights of the current study may be highly context specific, and the performance of the models may not be generalizable to other data or clinical settings.

CONCLUSION

After a thorough search for currently available LLMs for the extraction of signs and symptoms from Dutch EHR data, we selected the MedRoBERTa.nl and RobBERT models and compared their performance in a case study. Both models demonstrated good

performance when implemented as direct classifiers and are therefore promising for the extraction of signs and symptoms from Dutch unstructured EHR data, while prompt-based classifiers consistently showed worse performance in our analyses. This study also found that a minimal training sample size of 800 samples was generally sufficient to train the models as direct classifiers for extraction of most signs and symptoms. These findings support the feasibility of using LLMs for scalable and automated symptom extraction in Dutch clinical research settings. However, model performance varied substantially across signs and symptoms, with better results for clinically structured and consistently reported findings, and lower performance for more sparsely documented symptoms. This highlights the need for clinical awareness when interpreting model outputs and suggests that symptom-specific strategies may be required. In the context of prediction research, future studies should assess whether including signs and symptoms that are automatically extracted from Dutch EHR data by LLMs leads to improvement of clinical prediction model performance and, ultimately, improves prediction of individual patient outcomes' risks. Moreover, further development in prompt engineering may enhance performance and thereby reduce the need for costly manual annotations of clinical notes. These steps are essential to move from technical feasibility to meaningful clinical impact.

REFERENCES

1. Gao Y, Dligach D, Christensen L, et al. A scoping review of publicly available language tasks in clinical natural language processing. *J Am Med Inform Assoc*. 2022;**29**(10):1797–806.
2. Sun W, Cai Z, Li Y, Liu F, Fang S, Wang G. Data processing and text mining technologies on electronic medical records: a review. *J Healthc Eng*. 2018;**2018**(1):4302425.
3. Li L, Zhou J, Gao Z, et al. A scoping review of using large language models (LLMs) to investigate electronic health records (EHRs). *arXiv* (preprint). 2024. Available from: http://arxiv.org/abs/2405.03066.
4. Hadi MU, Tashi QA, Qureshi R, et al. A survey on large language models: applications, challenges, limitations, and practical usage. *TechRxiv* (preprint). 2023. Available from: https://www.authorea.com/doi/full/10.36227/techrxiv.23589741.v1.
5. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J Am Med Inform Assoc*. 2016;**23**(5):1007–15.
6. Sheikhalishahi S, Miotto R, Dudley JT, et al. Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform*. 2019;**7**(2):e12239.
7. Koleck TA, Dreisbach C, Bourne PE, Bakken S. Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *J Am Med Inform Assoc*. 2019;**26**(4):364–79.
8. Patra BG, Sharma MM, Vekaria V, et al. Extracting social determinants of health from electronic health records using natural language processing: a systematic review. *J Am Med Inform Assoc*. 2021;**28**(12):2716–27.
9. Seinen TM, Fridgeirsson EA, Ioannou S, et al. Use of unstructured text in prognostic clinical prediction models: a systematic review. *J Am Med Inform Assoc*. 2022;**29**(7):1292–302.
10. Tayefi M, Ngo P, Chomutare T, et al. Challenges and opportunities beyond structured data in analysis of electronic health records. *WIREs Comp Stat*. 2021;**13**(6):e1549.
11. Kim MK, Rouphael C, McMichael J, Welch N, Dasarathy S. Challenges in and opportunities for electronic health record-based data analysis and interpretation. *Gut Liver*. 2024;**18**(2):201–8.
12. Minaee S, Mikolov T, Nikzad N, et al. Large language models: a survey. *arXiv* (preprint). 2025. Available from: http://arxiv.org/abs/2402.06196.
13. Névéol A, Dalianis H, Velupillai S, Savova G, Zweigenbaum P. Clinical natural language processing in languages other than English: opportunities and challenges. *J Biomed Semantics*. 2018;**9**(1):12.
14. Rijk MH, Platteel TN, Geersing GJ, et al. 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. *Diagn Progn Res*. 2023;**7**(1):23.
15. Rijk MH, Platteel TN, Mulder MMM, et al. Incomplete and possibly selective recording of signs, symptoms, and measurements in free text fields of primary care electronic health records of adults with lower respiratory tract infections. *J Clin Epidemiol*. 2024;**166**:111240.
16. Smeets HM, Kortekaas MF, Rutten FH, 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.
17. Verkijk S, Vossen P. MedRoBERTa.nl: a language model for Dutch electronic health records. *Comp Linguistics Neth J*. 2021;**11**:141–59.
18. Delobelle P, Winters T, Berendt B. RobBERT: a Dutch RoBERTa-based language model. *arXiv* (preprint). 2020. Available from: http://arxiv.org/abs/2001.06286.

19. Li Y, Li Z, Zhang K, Dan R, Jiang S, Zhang Y. ChatDoctor: a medical chat model fine-tuned on a large language model meta-AI (LLaMA) using medical domain knowledge. *Cureus*. 2023;**15**(6):e40895.
20. Kerner T. Domain-specific pretraining of language models: a comparative study in the medical field. *arXiv* (preprint). 2024. Available from: http://arxiv.org/abs/2407.14076.
21. Muizelaar H, Haas M, van Dortmont K, van der Putten P, Spruit M. Extracting patient lifestyle characteristics from Dutch clinical text with BERT models. *BMC Med Inf Decis Mak*. 2024;**24**(1):151.
22. Zhang X, Talukdar N, Vemulapalli S, et al. Comparison of prompt engineering and fine-tuning strategies in large language models in the classification of clinical notes. *AMIA Jt Summits Transl Sci Proc*. 2024;**2024**:478–87.
23. Marvin G, Hellen N, Jjingo D, Nakatumba-Nabende J. Prompt engineering in large language models. In: Jacob IJ, Piramuthu S, Falkowski-Gilski P, editors. Data Intelligence and Cognitive Informatics. Singapore: Springer Nature; 2024. p. 387–402.
24. Wang J, Shi E, Yu S, et al. Prompt engineering for healthcare: methodologies and applications. *arXiv* (preprint). 2024. Available from: http://arxiv.org/abs/2304.14670.
25. Ren W, Liu Z, Wu Y, et al. Moving beyond medical statistics: a systematic review on missing data handling in electronic health records. *Health Data Sci*. 2024;**4**:0176.
26. Yang X, Chen A, PourNejatian N, et al. A large language model for electronic health records. *npj Digit Med*. 2022;**5**(1):194.
27. Zhang J, Sun K, Jagadeesh A, et al. The potential and pitfalls of using a large language model such as ChatGPT, GPT-4, or LLaMA as a clinical assistant. *J Am Med Inform Assoc*. 2024;**31**(9):1884–91.

SUPPLEMENTARY MATERIALS FOR CHAPTER 6

A. SELECTION OF LARGE LANGUAGE MODELS

Selection criteria

We defined a set of selection criteria that indicate the suitability of a large language model (LLM) for the purpose of our study which targets the extraction of data from Dutch electronic health records (EHRs):

- *Local applicability*: Since clinical notes and EHRs in general contain a multitude of privacy sensitive data, it is mandatory that the model is able to be completely implemented on local hardware, as no EHR data is allowed to be transmitted from the environment it is stored on. For this reason, only open-source, locally applicable LLMs could be selected.

- *Availability of a pre-trained model*: The model must be pre-trained, as we do not have the computational power and amount of data to train a model from scratch. We do however want to be able to further fine-tune our models on our own data and convert them into prompt-based classifiers as well.

- *Parameter count of a model*: The number of parameters in a model represents the size and complexity of a model and thus correlates with the computational power required to run it. As computational power is a limiting factor due to local application of the model, we needed to balance between having enough parameters for sufficient LLM performance and the ability to run the LLM run on a virtual desktop.

- *Dutch language competency*: It is vital for the model to be able to handle Dutch language data, either through optimization for the Dutch language or being completely trained on a Dutch dataset.

- *Medical text familiarity*: Models trained or fine-tuned on medical texts are more likely to be able to correctly process our domain-specific knowledge and jargon and are thus preferred.

Model search

To retrieve a list of relevant models, a literature search was performed and combined with a direct search for models on the Hugging Face website. These websites provide comprehensive filtering options, enabling us to find models that can be run offline rather easily. To find the models via HuggingFace, the following filters were applied during a search performed on 18 October 2023:

- *Tasks:* Text Classification, Zero-Shot Classification, Conversational, Text Generation, Fill-Mask

- *Libraries:* PyTorch, TensorFlow, Transformers, Sentence Transformers, Adapter Transformers.

- *Datasets:* Not specified

- *Languages:* Dutch

- *Licences:* Not specified

- *Other:* Not specified

In the context of this research, we focused our search of models on LLMs that show promise in classification tasks. The final list of potential models for those tasks and their relevant features is presented in Table A1.

Final selection

After our search for available models, two models were selected as most suitable for our task (see computation details of each model in Table A2):

- MedRoBERTa.nl[1]

    - Description: A Dutch medical model created by researchers at Vrije Universiteit Amsterdam, who used 13GB of text data from Dutch hospital notes to train an altered version of the RoBERTa model. This model has shown to perform better on this data in odd-one-out tasks when compared to general Dutch LLMs and is thus likely to perform well in our research.

    - Motivation: Of all chosen models, MedRoBERTa.nl is the only one directly trained on Dutch medical data. Consequently, this model is expected to perform very well on the dataset. One limitation is that the parameter count of the model is rather low by modern standards, which, while being a positive factor in terms of computational efficiency, might reduce its performance.

Table A1. The large language models that were retrieved by our search including the characteristics on suitability for extraction of information from Dutch electronic health records.

Model	Base Model	Number of parameters	Prompt-based model available	Local	Pretrained	Dutch	Medical	Open-source	Relevant paper	Model link	Suitable
BERTje	BERT	109M	No	Yes	Yes	Yes	No	Yes	Link	GH | HF	Yes
GPT-2 (recycled for Dutch)	GPT-2	129M or 369M	No	Yes	Yes	Yes	No	Yes	Link	HF small, medium	Yes
GPT-2 XL	GPT-2	1.5B	No	Yes	Yes	Multilingual	No	Yes	Link	HF	Yes
DistilBERT-nl	BERT	69M	No	Yes	Yes	Yes	No	Yes	Link	HF	Yes
GPT-3.5/4	GPT-3.5/4	154B / 1.76T	Yes	No	Yes	Multilingual	No	No	Link3.5 / Link4	NA	No
Legal BERT	BERT	295M	No	Yes	Yes	Yes	No	Yes	Link	HF	No
medRoBERTa.nl	BERT	117M	Yes	Yes	No (possibly via contact)	Yes	Yes	Yes	Link	GH	Yes
Google Bard	Lambda	137B	Yes	No	Yes	Multilingual	No	No	NA	NA	No
LLaMa	LLaMa	65B	Yes	Yes	Yes	Multilingual	No	No	Link	HF	Yes
LLaMa 2	LLaMa2	7B to 70B	Yes	Yes	Yes	Multilingual	No	No	Link	HF	Yes
ChatDoctor	LLaMa-7B	7B	Yes	Yes	No	Multilingual	Yes	No	Link	GH	Yes
RoBERTa	BERT	355M	Yes	Yes	Yes	Multilingual	No	Yes	Link	HF	Yes
RobBERT-2023	BERT	117M	Yes	Yes	Yes	Yes	No	Yes	Link	HF	Yes
DialoGPT	None	117M, 345M or 762M	Yes	Yes	Yes	Multilingual	No	Yes	Link	GH	Yes, but not used

- RobBERT[2]

    - Description: RobBERT is a model built upon the foundation of BERT, serving as the core architecture. BERT was released as a multilingual model, and RobBERT was fine-tuned as a language-specific model, as research points out that doing so generally results in higher performance.

    - Motivation: RobBERT's strong performance on various NLP tasks and proven ability to handle Dutch text make it a suitable candidate for our investigation.

Table A2. The number of parameters of the selected MedRoBERTa.nl and RobBERT models.

Classifier	Model	Number of parameters
Direct	*MedRoBERTa.nl*	125,980,419
	RobBERT	116,764,419
Prompt-based	*MedRoBERTa.nl*	263,859,258
	RobBERT	254,643,258

Both models were implemented as direct classifier and as prompt-based classifier. Prompt-based classifiers had a larger number of parameters compared to the direct classifiers due to extra layers being added by the EncoderDecoder model (with the base multilingual BERT model as decoder).

1. Verkijk S, Vossen P. MedRoBERTa.nl: a language model for Dutch electronic health records. Comp Linguistics Neth J. 2021;**11**:141–59.

2. Delobelle P, Winters T, Berendt B. RobBERT: a Dutch RoBERTa-based language model. arXiv (preprint). 2020. Available from: http://arxiv.org/abs/2001.06286.

B. Performance metrics of the LLMs in extracting signs and symptoms from Dutch EHRs

Table A3. The performance of the MedRoBERTa.nl and RobBERT models as direct classifier in terms of recall, precision, and F1-score for different training sample sizes.

						Per class metrics								
			Macro-averaged metrics			'Present'			'Absent'			'Not reported'		
Model	Sign or symptom	Number of samples	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
MedRoBERTa.nl	Chest pain	200	0,28	0,33	0,31	0,00	0,00	0,00	0,00	0,00	0,00	0,85	1,00	0,92
MedRoBERTa.nl	Chest pain	400	0,70	0,43	0,45	0,58	0,10	0,16	0,65	0,19	0,27	0,87	0,99	0,93
MedRoBERTa.nl	Chest pain	600	0,77	0,74	0,74	0,67	0,65	0,65	0,70	0,63	0,63	0,95	0,95	0,95
MedRoBERTa.nl	Chest pain	800	0,79	0,81	0,78	0,74	0,68	0,68	0,68	0,80	0,72	0,96	0,95	0,95
MedRoBERTa.nl	Chest pain	1000	0,80	0,82	0,80	0,69	0,76	0,72	0,74	0,74	0,73	0,96	0,95	0,96
MedRoBERTa.nl	Chest pain	1200	0,82	0,78	0,79	0,70	0,68	0,68	0,80	0,71	0,74	0,96	0,96	0,96
MedRoBERTa.nl	Chest pain	1400	0,83	0,78	0,80	0,68	0,72	0,69	0,84	0,67	0,74	0,96	0,96	0,96
MedRoBERTa.nl	Chest pain	1600	0,84	0,74	0,78	0,75	0,59	0,65	0,82	0,67	0,73	0,94	0,97	0,96
MedRoBERTa.nl	Chills	200	0,32	0,33	0,32	0,00	0,00	0,00	0,00	0,00	0,00	0,95	1,00	0,97
MedRoBERTa.nl	Chills	400	0,32	0,33	0,32	0,00	0,00	0,00	0,00	0,00	0,00	0,95	1,00	0,97
MedRoBERTa.nl	Chills	600	0,32	0,33	0,32	0,00	0,00	0,00	0,00	0,00	0,00	0,95	1,00	0,97
MedRoBERTa.nl	Chills	800	0,64	0,62	0,63	0,91	0,86	0,88	0,00	0,00	0,00	0,99	1,00	1,00
MedRoBERTa.nl	Chills	1000	0,62	0,62	0,62	0,88	0,85	0,86	0,00	0,00	0,00	0,99	1,00	0,99
MedRoBERTa.nl	Chills	1200	0,63	0,63	0,63	0,90	0,90	0,90	0,00	0,00	0,00	0,99	1,00	1,00
MedRoBERTa.nl	Chills	1400	0,64	0,61	0,62	0,92	0,82	0,85	0,00	0,00	0,00	0,99	1,00	1,00
MedRoBERTa.nl	Chills	1600	0,63	0,63	0,63	0,91	0,89	0,89	0,00	0,00	0,00	0,99	1,00	1,00
MedRoBERTa.nl	Confusion	200	0,32	0,33	0,33	0,00	0,00	0,00	0,00	0,00	0,00	0,97	1,00	0,98
MedRoBERTa.nl	Confusion	400	0,32	0,33	0,33	0,00	0,00	0,00	0,00	0,00	0,00	0,97	1,00	0,98
MedRoBERTa.nl	Confusion	600	0,32	0,33	0,33	0,00	0,00	0,00	0,00	0,00	0,00	0,97	1,00	0,98
MedRoBERTa.nl	Confusion	800	0,42	0,41	0,41	0,20	0,20	0,20	0,08	0,04	0,05	0,97	1,00	0,99
MedRoBERTa.nl	Confusion	1000	0,48	0,46	0,44	0,07	0,20	0,10	0,40	0,19	0,25	0,98	1,00	0,99
MedRoBERTa.nl	Confusion	1200	0,52	0,54	0,52	0,10	0,20	0,13	0,49	0,44	0,44	0,98	0,99	0,99

Table A3. Continued

			Macro-averaged metrics			Per class metrics								
						'Present'			'Absent'			'Not reported'		
Model	Sign or symptom	Number of samples	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
MedRoBERTa.nl	Confusion	1400	0,64	0,61	0,57	0,45	0,33	0,27	0,48	0,52	0,45	0,98	0,99	0,99
MedRoBERTa.nl	Confusion	1600	0,61	0,54	0,56	0,40	0,33	0,35	0,46	0,29	0,34	0,98	0,99	0,99
MedRoBERTa.nl	Cough	200	0,25	0,33	0,29	0,76	1,00	0,86	0,00	0,00	0,00	0,00	0,00	0,00
MedRoBERTa.nl	Cough	400	0,55	0,39	0,40	0,79	0,99	0,88	0,00	0,00	0,00	0,85	0,19	0,31
MedRoBERTa.nl	Cough	600	0,59	0,59	0,59	0,93	0,96	0,95	0,00	0,00	0,00	0,85	0,81	0,83
MedRoBERTa.nl	Cough	800	0,60	0,59	0,59	0,93	0,96	0,94	0,00	0,00	0,00	0,87	0,79	0,82
MedRoBERTa.nl	Cough	1000	0,60	0,60	0,60	0,94	0,96	0,95	0,00	0,00	0,00	0,87	0,85	0,86
MedRoBERTa.nl	Cough	1200	0,60	0,59	0,59	0,93	0,97	0,95	0,00	0,00	0,00	0,88	0,79	0,83
MedRoBERTa.nl	Cough	1400	0,60	0,60	0,60	0,94	0,96	0,95	0,00	0,00	0,00	0,87	0,85	0,86
MedRoBERTa.nl	Cough	1600	0,60	0,59	0,59	0,93	0,97	0,95	0,00	0,00	0,00	0,88	0,79	0,83
MedRoBERTa.nl	Crackles upon auscultation	200	0,55	0,49	0,47	0,00	0,00	0,00	0,66	1,00	0,80	0,97	0,47	0,63
MedRoBERTa.nl	Crackles upon auscultation	400	0,83	0,57	0,57	0,87	0,13	0,20	0,70	0,98	0,81	0,93	0,59	0,72
MedRoBERTa.nl	Crackles upon auscultation	600	0,88	0,77	0,81	0,81	0,77	0,79	0,85	0,94	0,89	0,97	0,61	0,74
MedRoBERTa.nl	Crackles upon auscultation	800	0,88	0,82	0,85	0,84	0,81	0,83	0,88	0,95	0,91	0,92	0,70	0,79
MedRoBERTa.nl	Crackles upon auscultation	1000	0,89	0,83	0,85	0,88	0,80	0,84	0,88	0,96	0,92	0,92	0,72	0,80
MedRoBERTa.nl	Crackles upon auscultation	1200	0,87	0,85	0,86	0,84	0,86	0,85	0,91	0,92	0,92	0,87	0,76	0,80
MedRoBERTa.nl	Crackles upon auscultation	1400	0,87	0,86	0,86	0,85	0,86	0,85	0,91	0,92	0,92	0,85	0,78	0,81
MedRoBERTa.nl	Crackles upon auscultation	1600	0,88	0,86	0,87	0,86	0,88	0,87	0,92	0,93	0,93	0,87	0,77	0,82
MedRoBERTa.nl	Fever	200	0,48	0,37	0,29	0,54	0,03	0,06	0,46	0,20	0,23	0,45	0,88	0,59
MedRoBERTa.nl	Fever	400	0,60	0,52	0,51	0,67	0,24	0,35	0,45	0,70	0,54	0,69	0,64	0,66
MedRoBERTa.nl	Fever	600	0,70	0,70	0,70	0,69	0,74	0,71	0,62	0,60	0,61	0,79	0,77	0,78
MedRoBERTa.nl	Fever	800	0,73	0,73	0,72	0,68	0,79	0,73	0,74	0,54	0,62	0,77	0,85	0,81
MedRoBERTa.nl	Fever	1000	0,73	0,73	0,72	0,70	0,75	0,72	0,69	0,59	0,64	0,79	0,84	0,81
MedRoBERTa.nl	Fever	1200	0,71	0,70	0,70	0,70	0,75	0,72	0,62	0,62	0,62	0,80	0,73	0,76
MedRoBERTa.nl	Fever	1400	0,72	0,72	0,72	0,70	0,81	0,74	0,67	0,61	0,64	0,80	0,75	0,77
MedRoBERTa.nl	Fever	1600	0,71	0,72	0,71	0,66	0,81	0,72	0,67	0,60	0,63	0,79	0,75	0,77
MedRoBERTa.nl	Ill appearance	200	0,23	0,33	0,27	0,00	0,00	0,00	0,00	0,00	0,00	0,69	1,00	0,82

Table A3. Continued

			Macro-averaged metrics			Per class metrics								
						'Present'			'Absent'			'Not reported'		
Model	Sign or symptom	Number of samples	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
MedRoBERTa.nl	Ill appearance	400	0,58	0,62	0,58	0,11	0,02	0,03	0,72	0,86	0,78	0,89	0,98	0,93
MedRoBERTa.nl	Ill appearance	600	0,86	0,83	0,84	0,75	0,64	0,69	0,91	0,88	0,90	0,93	0,96	0,94
MedRoBERTa.nl	Ill appearance	800	0,87	0,81	0,83	0,78	0,57	0,64	0,91	0,89	0,90	0,92	0,96	0,94
MedRoBERTa.nl	Ill appearance	1000	0,87	0,85	0,86	0,74	0,69	0,71	0,93	0,89	0,91	0,93	0,95	0,94
MedRoBERTa.nl	Ill appearance	1200	0,88	0,85	0,86	0,77	0,69	0,73	0,94	0,89	0,91	0,93	0,96	0,95
MedRoBERTa.nl	Ill appearance	1400	0,90	0,85	0,87	0,81	0,70	0,75	0,94	0,89	0,92	0,94	0,97	0,95
MedRoBERTa.nl	Ill appearance	1600	0,90	0,85	0,87	0,83	0,67	0,73	0,95	0,90	0,92	0,93	0,97	0,95
MedRoBERTa.nl	Shortness of breath	200	0,38	0,40	0,33	0,45	0,86	0,58	0,00	0,00	0,00	0,71	0,33	0,42
MedRoBERTa.nl	Shortness of breath	400	0,43	0,48	0,44	0,53	0,84	0,65	0,00	0,00	0,00	0,77	0,60	0,66
MedRoBERTa.nl	Shortness of breath	600	0,80	0,69	0,69	0,79	0,89	0,84	0,75	0,27	0,36	0,86	0,91	0,88
MedRoBERTa.nl	Shortness of breath	800	0,85	0,80	0,81	0,86	0,91	0,88	0,78	0,59	0,66	0,90	0,91	0,90
MedRoBERTa.nl	Shortness of breath	1000	0,85	0,82	0,83	0,86	0,93	0,89	0,78	0,65	0,71	0,92	0,89	0,90
MedRoBERTa.nl	Shortness of breath	1200	0,86	0,82	0,83	0,89	0,92	0,90	0,77	0,62	0,68	0,91	0,93	0,92
MedRoBERTa.nl	Shortness of breath	1400	0,88	0,83	0,85	0,87	0,93	0,90	0,85	0,64	0,73	0,91	0,91	0,91
MedRoBERTa.nl	Shortness of breath	1600	0,85	0,82	0,83	0,87	0,93	0,90	0,75	0,65	0,69	0,92	0,90	0,91
MedRoBERTa.nl	Sputum	200	0,24	0,33	0,28	0,00	0,00	0,00	0,00	0,00	0,00	0,73	1,00	0,85
MedRoBERTa.nl	Sputum	400	0,50	0,37	0,36	0,74	0,15	0,21	0,00	0,00	0,00	0,76	0,98	0,85
MedRoBERTa.nl	Sputum	600	0,58	0,60	0,59	0,79	0,85	0,81	0,00	0,00	0,00	0,94	0,95	0,95
MedRoBERTa.nl	Sputum	800	0,58	0,62	0,60	0,78	0,90	0,84	0,00	0,00	0,00	0,96	0,95	0,95
MedRoBERTa.nl	Sputum	1000	0,80	0,70	0,71	0,80	0,89	0,84	0,64	0,25	0,33	0,96	0,95	0,95
MedRoBERTa.nl	Sputum	1200	0,83	0,77	0,79	0,84	0,87	0,86	0,68	0,49	0,56	0,95	0,95	0,95
MedRoBERTa.nl	Sputum	1400	0,83	0,79	0,80	0,82	0,88	0,85	0,71	0,56	0,61	0,96	0,94	0,95
MedRoBERTa.nl	Sputum	1600	0,81	0,78	0,78	0,86	0,87	0,86	0,63	0,51	0,54	0,95	0,95	0,95
RobBERT	Chest pain	200	0,28	0,33	0,31	0,00	0,00	0,00	0,00	0,00	0,00	0,85	1,00	0,92
RobBERT	Chest pain	400	0,28	0,33	0,31	0,00	0,00	0,00	0,00	0,00	0,00	0,85	1,00	0,92
RobBERT	Chest pain	600	0,56	0,51	0,49	0,54	0,58	0,53	0,20	0,01	0,03	0,92	0,94	0,93
RobBERT	Chest pain	800	0,57	0,57	0,54	0,61	0,57	0,52	0,16	0,20	0,16	0,93	0,94	0,94

Table A3. Continued

			Macro-averaged metrics			Per class metrics								
						'Present'			'Absent'			'Not reported'		
Model	Sign or symptom	Number of samples	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
RobBERT	Chest pain	1000	0,66	0,63	0,61	0,59	0,68	0,62	0,46	0,29	0,29	0,94	0,93	0,94
RobBERT	Chest pain	1200	0,68	0,63	0,63	0,64	0,60	0,61	0,46	0,33	0,35	0,93	0,95	0,94
RobBERT	Chest pain	1400	0,72	0,60	0,60	0,58	0,70	0,62	0,65	0,16	0,24	0,94	0,95	0,94
RobBERT	Chest pain	1600	0,66	0,68	0,66	0,60	0,66	0,60	0,44	0,44	0,43	0,95	0,94	0,94
RobBERT	Chills	200	0,32	0,33	0,32	0,00	0,00	0,00	0,00	0,00	0,00	0,95	1,00	0,97
RobBERT	Chills	400	0,32	0,33	0,32	0,00	0,00	0,00	0,00	0,00	0,00	0,95	1,00	0,97
RobBERT	Chills	600	0,32	0,33	0,32	0,00	0,00	0,00	0,00	0,00	0,00	0,95	1,00	0,97
RobBERT	Chills	800	0,54	0,55	0,55	0,65	0,67	0,65	0,00	0,00	0,00	0,98	0,99	0,99
RobBERT	Chills	1000	0,56	0,59	0,57	0,70	0,77	0,73	0,00	0,00	0,00	0,99	1,00	0,99
RobBERT	Chills	1200	0,55	0,58	0,56	0,65	0,76	0,70	0,00	0,00	0,00	0,99	0,99	0,99
RobBERT	Chills	1400	0,63	0,65	0,64	0,88	0,95	0,91	0,00	0,00	0,00	1,00	1,00	1,00
RobBERT	Chills	1600	0,61	0,65	0,63	0,82	0,96	0,88	0,00	0,00	0,00	1,00	0,99	0,99
RobBERT	Confusion	200	0,32	0,33	0,33	0,00	0,00	0,00	0,00	0,00	0,00	0,97	1,00	0,98
RobBERT	Confusion	400	0,32	0,33	0,33	0,00	0,00	0,00	0,00	0,00	0,00	0,97	1,00	0,98
RobBERT	Confusion	600	0,32	0,33	0,33	0,00	0,00	0,00	0,00	0,00	0,00	0,97	1,00	0,98
RobBERT	Confusion	800	0,36	0,34	0,34	0,00	0,00	0,00	0,10	0,02	0,03	0,97	1,00	0,98
RobBERT	Confusion	1000	0,34	0,35	0,35	0,03	0,04	0,03	0,02	0,03	0,02	0,97	0,99	0,98
RobBERT	Confusion	1200	0,39	0,37	0,37	0,20	0,10	0,13	0,00	0,00	0,00	0,97	1,00	0,98
RobBERT	Confusion	1400	0,38	0,39	0,38	0,00	0,00	0,00	0,17	0,19	0,17	0,97	0,99	0,98
RobBERT	Confusion	1600	0,48	0,45	0,46	0,20	0,20	0,20	0,27	0,17	0,21	0,97	1,00	0,98
RobBERT	Cough	200	0,25	0,33	0,29	0,76	1,00	0,86	0,00	0,00	0,00	0,00	0,00	0,00
RobBERT	Cough	400	0,54	0,46	0,47	0,83	0,97	0,89	0,00	0,00	0,00	0,77	0,40	0,53
RobBERT	Cough	600	0,60	0,58	0,58	0,92	0,97	0,94	0,00	0,00	0,00	0,89	0,75	0,81
RobBERT	Cough	800	0,60	0,59	0,59	0,93	0,96	0,94	0,00	0,00	0,00	0,86	0,79	0,82
RobBERT	Cough	1000	0,61	0,60	0,60	0,94	0,97	0,95	0,00	0,00	0,00	0,88	0,83	0,85
RobBERT	Cough	1200	0,61	0,60	0,60	0,93	0,97	0,95	0,00	0,00	0,00	0,89	0,82	0,85
RobBERT	Cough	1400	0,62	0,60	0,61	0,94	0,98	0,96	0,00	0,00	0,00	0,91	0,83	0,87

Table A3. Continued

			Macro-averaged metrics			Per class metrics								
						'Present'			'Absent'			'Not reported'		
Model	Sign or symptom	Number of samples	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
RobBERT	Cough	1600	0,61	0,58	0,59	0,92	0,98	0,95	0,00	0,00	0,00	0,90	0,77	0,82
RobBERT	Crackles upon auscultation	200	0,41	0,42	0,38	0,00	0,00	0,00	0,64	1,00	0,78	0,59	0,27	0,36
RobBERT	Crackles upon auscultation	400	0,87	0,71	0,76	0,89	0,56	0,67	0,79	0,97	0,87	0,93	0,61	0,73
RobBERT	Crackles upon auscultation	600	0,89	0,79	0,82	0,87	0,80	0,83	0,86	0,96	0,90	0,95	0,61	0,74
RobBERT	Crackles upon auscultation	800	0,88	0,81	0,84	0,85	0,85	0,84	0,88	0,94	0,91	0,91	0,65	0,76
RobBERT	Crackles upon auscultation	1000	0,87	0,83	0,85	0,84	0,88	0,86	0,90	0,93	0,92	0,88	0,68	0,76
RobBERT	Crackles upon auscultation	1200	0,88	0,82	0,84	0,88	0,85	0,86	0,89	0,94	0,91	0,86	0,68	0,74
RobBERT	Crackles upon auscultation	1400	0,85	0,85	0,85	0,83	0,91	0,87	0,92	0,91	0,91	0,80	0,73	0,76
RobBERT	Crackles upon auscultation	1600	0,89	0,84	0,86	0,85	0,88	0,86	0,91	0,93	0,92	0,90	0,72	0,79
RobBERT	Fever	200	0,28	0,36	0,26	0,07	0,00	0,00	0,34	0,18	0,20	0,44	0,89	0,59
RobBERT	Fever	400	0,54	0,46	0,42	0,48	0,08	0,13	0,41	0,84	0,55	0,72	0,48	0,57
RobBERT	Fever	600	0,63	0,58	0,58	0,65	0,42	0,49	0,47	0,69	0,56	0,76	0,62	0,68
RobBERT	Fever	800	0,69	0,69	0,68	0,61	0,80	0,69	0,70	0,49	0,57	0,77	0,79	0,78
RobBERT	Fever	1000	0,69	0,69	0,69	0,70	0,70	0,69	0,63	0,58	0,60	0,76	0,79	0,77
RobBERT	Fever	1200	0,72	0,69	0,68	0,73	0,66	0,66	0,63	0,64	0,61	0,80	0,77	0,77
RobBERT	Fever	1400	0,71	0,69	0,68	0,68	0,77	0,71	0,66	0,61	0,61	0,80	0,70	0,72
RobBERT	Fever	1600	0,71	0,71	0,70	0,69	0,78	0,73	0,66	0,58	0,61	0,78	0,76	0,77
RobBERT	Ill appearance	200	0,23	0,33	0,27	0,00	0,00	0,00	0,00	0,00	0,00	0,69	1,00	0,82
RobBERT	Ill appearance	400	0,74	0,64	0,61	0,63	0,10	0,16	0,68	0,86	0,76	0,90	0,94	0,92
RobBERT	Ill appearance	600	0,85	0,81	0,83	0,73	0,60	0,65	0,89	0,88	0,88	0,92	0,95	0,94
RobBERT	Ill appearance	800	0,88	0,82	0,84	0,82	0,59	0,68	0,89	0,89	0,89	0,93	0,96	0,94
RobBERT	Ill appearance	1000	0,85	0,85	0,85	0,72	0,72	0,71	0,88	0,89	0,89	0,94	0,93	0,94
RobBERT	Ill appearance	1200	0,89	0,84	0,85	0,81	0,65	0,72	0,91	0,89	0,90	0,93	0,96	0,95
RobBERT	Ill appearance	1400	0,86	0,84	0,84	0,74	0,67	0,69	0,90	0,88	0,89	0,94	0,95	0,94
RobBERT	Ill appearance	1600	0,88	0,83	0,84	0,79	0,62	0,68	0,90	0,90	0,90	0,93	0,96	0,94
RobBERT	Shortness of breath	200	0,39	0,41	0,36	0,46	0,85	0,59	0,00	0,00	0,00	0,71	0,37	0,47
RobBERT	Shortness of breath	400	0,43	0,46	0,42	0,51	0,87	0,64	0,00	0,00	0,00	0,78	0,52	0,62

Table A3. Continued

			Macro-averaged metrics			Per class metrics								
						'Present'			'Absent'			'Not reported'		
Model	Sign or symptom	Number of samples	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
RobBERT	Shortness of breath	600	0,76	0,60	0,57	0,74	0,86	0,79	0,73	0,05	0,09	0,82	0,88	0,84
RobBERT	Shortness of breath	800	0,84	0,79	0,80	0,85	0,91	0,88	0,77	0,52	0,61	0,90	0,92	0,91
RobBERT	Shortness of breath	1000	0,84	0,80	0,82	0,85	0,91	0,88	0,77	0,59	0,67	0,91	0,91	0,91
RobBERT	Shortness of breath	1200	0,86	0,81	0,83	0,87	0,92	0,89	0,81	0,59	0,68	0,91	0,93	0,91
RobBERT	Shortness of breath	1400	0,87	0,81	0,83	0,85	0,94	0,89	0,83	0,58	0,68	0,92	0,91	0,91
RobBERT	Shortness of breath	1600	0,87	0,83	0,85	0,87	0,94	0,90	0,82	0,64	0,72	0,92	0,91	0,91
RobBERT	Sputum	200	0,24	0,33	0,28	0,00	0,00	0,00	0,00	0,00	0,00	0,73	1,00	0,85
RobBERT	Sputum	400	0,54	0,48	0,49	0,78	0,46	0,56	0,00	0,00	0,00	0,83	0,97	0,89
RobBERT	Sputum	600	0,58	0,60	0,59	0,80	0,83	0,81	0,00	0,00	0,00	0,94	0,96	0,95
RobBERT	Sputum	800	0,57	0,61	0,59	0,76	0,89	0,82	0,00	0,00	0,00	0,95	0,94	0,95
RobBERT	Sputum	1000	0,57	0,61	0,59	0,76	0,88	0,81	0,00	0,00	0,00	0,95	0,94	0,95
RobBERT	Sputum	1200	0,58	0,61	0,59	0,80	0,86	0,83	0,00	0,00	0,00	0,94	0,96	0,95
RobBERT	Sputum	1400	0,56	0,58	0,56	0,76	0,79	0,76	0,00	0,00	0,00	0,92	0,94	0,93
RobBERT	Sputum	1600	0,65	0,62	0,62	0,81	0,87	0,83	0,20	0,04	0,07	0,94	0,96	0,95

Table A4. The performance of the MedRoBERTa.nl and RobBERT models as prompt-based classifier in terms of recall, precision, and F1-score for different sample sizes.

						Per class metrics								
			Macro-averaged metrics			'Present'			'Absent'			'Not reported'		
Model	Sign or symptom	Number of samples	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
MedRoBERTa.nl	Chest pain	1	0,01	0,33	0,03	0,00	0,00	0,00	0,04	1,00	0,08	0,00	0,00	0,00
MedRoBERTa.nl	Chest pain	2	0,01	0,33	0,03	0,00	0,00	0,00	0,04	1,00	0,08	0,00	0,00	0,00
MedRoBERTa.nl	Chest pain	3	0,01	0,33	0,03	0,00	0,00	0,00	0,04	1,00	0,08	0,00	0,00	0,00
MedRoBERTa.nl	Chills	1	0,01	0,32	0,02	0,04	0,21	0,06	0,00	0,73	0,01	0,00	0,00	0,00
MedRoBERTa.nl	Chills	2	0,01	0,32	0,02	0,04	0,21	0,06	0,00	0,73	0,01	0,00	0,00	0,00
MedRoBERTa.nl	Chills	3	0,01	0,32	0,02	0,04	0,21	0,06	0,00	0,73	0,01	0,00	0,00	0,00
MedRoBERTa.nl	Confusion	1	0,34	0,33	0,05	0,01	0,92	0,02	0,00	0,02	0,00	1,00	0,06	0,11
MedRoBERTa.nl	Confusion	2	0,34	0,33	0,05	0,01	0,92	0,02	0,00	0,02	0,00	1,00	0,06	0,11
MedRoBERTa.nl	Confusion	3	0,34	0,33	0,05	0,01	0,92	0,02	0,00	0,02	0,00	1,00	0,06	0,11
MedRoBERTa.nl	Cough	1	0,33	0,33	0,02	0,73	0,01	0,02	0,01	0,97	0,03	0,25	0,01	0,02
MedRoBERTa.nl	Cough	2	0,33	0,33	0,02	0,73	0,01	0,02	0,01	0,97	0,03	0,25	0,01	0,02
MedRoBERTa.nl	Cough	3	0,33	0,33	0,02	0,73	0,01	0,02	0,01	0,97	0,03	0,25	0,01	0,02
MedRoBERTa.nl	Crackles upon auscultation	1	0,28	0,33	0,30	0,26	0,37	0,30	0,59	0,63	0,61	0,00	0,00	0,00
MedRoBERTa.nl	Crackles upon auscultation	2	0,28	0,33	0,30	0,26	0,37	0,30	0,59	0,63	0,61	0,00	0,00	0,00
MedRoBERTa.nl	Crackles upon auscultation	3	0,28	0,33	0,30	0,26	0,37	0,30	0,59	0,63	0,61	0,00	0,00	0,00
MedRoBERTa.nl	Fever	1	0,18	0,33	0,14	0,25	1,00	0,40	0,30	0,00	0,01	0,00	0,00	0,00
MedRoBERTa.nl	Fever	2	0,18	0,33	0,14	0,25	1,00	0,40	0,30	0,00	0,01	0,00	0,00	0,00
MedRoBERTa.nl	Fever	3	0,18	0,33	0,14	0,25	1,00	0,40	0,30	0,00	0,01	0,00	0,00	0,00
MedRoBERTa.nl	Ill appearance	1	0,18	0,34	0,07	0,11	1,00	0,20	0,44	0,01	0,03	0,00	0,00	0,00

Table A4. Continued

			Macro-averaged metrics			Per class metrics								
						'Present'			'Absent'			'Not reported'		
Model	Sign or symptom	Number of samples	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
MedRoBERTa.nl	Sputum	2	0,01	0,33	0,02	0,00	0,00	0,00	0,03	1,00	0,05	0,00	0,00	0,00
MedRoBERTa.nl	Sputum	3	0,01	0,33	0,02	0,00	0,00	0,00	0,03	1,00	0,05	0,00	0,00	0,00
RobBERT	Chest pain	1	0,01	0,33	0,03	0,00	0,00	0,00	0,04	1,00	0,08	0,00	0,00	0,00
RobBERT	Chest pain	2	0,01	0,33	0,03	0,00	0,00	0,00	0,04	1,00	0,08	0,00	0,00	0,00
RobBERT	Chest pain	3	0,01	0,33	0,03	0,00	0,00	0,00	0,04	1,00	0,08	0,00	0,00	0,00
RobBERT	Chills	1	0,02	0,33	0,03	0,05	0,99	0,09	0,00	0,00	0,00	0,00	0,00	0,00
RobBERT	Chills	2	0,02	0,33	0,03	0,05	0,99	0,09	0,00	0,00	0,00	0,00	0,00	0,00
RobBERT	Chills	3	0,02	0,33	0,03	0,05	0,99	0,09	0,00	0,00	0,00	0,00	0,00	0,00
RobBERT	Confusion	1	0,01	0,31	0,02	0,02	0,12	0,03	0,02	0,80	0,04	0,00	0,00	0,00
RobBERT	Confusion	2	0,01	0,31	0,02	0,02	0,12	0,03	0,02	0,80	0,04	0,00	0,00	0,00
RobBERT	Confusion	3	0,01	0,31	0,02	0,02	0,12	0,03	0,02	0,80	0,04	0,00	0,00	0,00
RobBERT	Cough	1	0,07	0,32	0,02	0,00	0,00	0,00	0,01	0,95	0,03	0,19	0,02	0,03
RobBERT	Cough	2	0,07	0,32	0,02	0,00	0,00	0,00	0,01	0,95	0,03	0,21	0,02	0,03
RobBERT	Cough	3	0,07	0,32	0,02	0,00	0,00	0,00	0,01	0,95	0,03	0,21	0,02	0,03
RobBERT	Crackles upon auscultation	1	0,08	0,33	0,13	0,23	1,00	0,38	0,00	0,00	0,00	0,00	0,00	0,00
RobBERT	Crackles upon auscultation	2	0,08	0,33	0,13	0,23	1,00	0,38	0,00	0,00	0,00	0,00	0,00	0,00
RobBERT	Crackles upon auscultation	3	0,08	0,33	0,13	0,23	1,00	0,38	0,00	0,00	0,00	0,00	0,00	0,00
RobBERT	Fever	1	0,20	0,33	0,22	0,24	0,75	0,37	0,37	0,25	0,30	0,00	0,00	0,00
RobBERT	Fever	2	0,20	0,33	0,22	0,24	0,75	0,37	0,37	0,25	0,30	0,00	0,00	0,00
RobBERT	Fever	3	0,21	0,33	0,22	0,24	0,75	0,37	0,37	0,25	0,30	0,00	0,00	0,00
RobBERT	Ill appearance	1	0,04	0,33	0,06	0,11	1,00	0,19	0,00	0,00	0,00	0,00	0,00	0,00
RobBERT	Ill appearance	2	0,04	0,33	0,06	0,11	1,00	0,19	0,00	0,00	0,00	0,00	0,00	0,00
RobBERT	Ill appearance	3	0,04	0,33	0,06	0,11	1,00	0,19	0,00	0,00	0,00	0,00	0,00	0,00
RobBERT	Shortness of breath	1	0,04	0,33	0,07	0,00	0,00	0,00	0,12	1,00	0,22	0,00	0,00	0,00
RobBERT	Shortness of breath	2	0,04	0,33	0,07	0,00	0,00	0,00	0,12	1,00	0,22	0,00	0,00	0,00
RobBERT	Shortness of breath	3	0,04	0,33	0,07	0,00	0,00	0,00	0,12	1,00	0,22	0,00	0,00	0,00
RobBERT	Sputum	1	0,08	0,33	0,13	0,24	1,00	0,38	0,00	0,00	0,00	0,00	0,00	0,00

Table A4. Continued

			Macro-averaged metrics			Per class metrics								
						'Present'			'Absent'			'Not reported'		
Model	Sign or symptom	Number of samples	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
RobBERT	Sputum	2	0,08	0,33	0,13	0,24	1,00	0,38	0,00	0,00	0,00	0,00	0,00	0,00
RobBERT	Sputum	3	0,08	0,33	0,13	0,24	1,00	0,38	0,00	0,00	0,00	0,00	0,00	0,00

7

The incremental predictive value of signs and symptoms derived from clinical notes using large language models for prediction of hospitalisation or mortality in patients presenting to general practice with a lower respiratory tract infection

Merijn H. Rijk*
Isa Spiero*
Tamara N. Platteel
Frans H. Rutten
Monika Hollander
Geert-Jan Geersing
Artuur M. Leeuwernberg
Roderick P. Venekamp

* Authors contributed equally
Prepared for submission

Objectives To evaluate the incremental predictive value of natural language processing (NLP)-derived signs and symptoms from clinical notes beyond structured electronic health records (EHR) data for predicting 30-day hospitalisation or mortality in patients presenting to general practice with a lower respiratory tract infection (LRTI).

Study design and setting Prediction model development and validation study based on two general practice cohorts of LRTI patients aged ≥40 years derived from Dutch EHR registries: a 2016–2019 cohort from Julius General Practitioners' Network (JGPN; development) and a 2022–2023 cohort from Academic Network of General Practice Amsterdam (ANHA; external validation) from the Utrecht and Amsterdam regions, respectively. Nine presenting signs and symptoms were retrieved from EHR clinical notes by a large language model specifically fine-tuned and validated for this task. Four elastic net logistic regression models predicting individual risk of 30-day all-cause hospitalisation or mortality were developed in JGPN: (i) a basic model (age and sex); (ii) the basic model plus NLP-derived signs and symptoms; (iii) an extended model of structured EHR-based predictors (age, sex, medical history, medication use, and LRTI diagnosis); and (iv) this extended model plus NLP-derived signs and symptoms. All models were internally validated by bootstrapping and externally validated in ANHA. Model performance was evaluated based on discrimination, calibration, and decision curve analysis.

Results The development and validation cohorts comprised 11,065 patients (outcome prevalence: 7.8%) and 6,284 patients (outcome prevalence: 11.4%), respectively. Discrimination of the basic model showed modest improvement after adding NLP-derived predictors (c-statistic from 0.61 (0.59–0.63) to 0.67 (0.65–0.69)). For the extended model, discrimination only slightly improved after addition of NLP-derived predictors (c-statistic 0.71 (0.69–0.73) to 0.73 (0.71–0.75)) without substantial improvement of model calibration. Decision curve analysis of the extended models revealed limited net benefit of adding NLP-derived predictors.

Conclusion Although signs and symptoms improved model performance beyond age and sex, their incremental value for predicting 30-day hospitalisation or mortality beyond structured EHR-based predictors was limited. This implies that signs and symptoms do not add substantial predictive information in LRTI patients, provided that GPs adequately incorporate all relevant structured information available from patients' EHR.

Abstract

INTRODUCTION

Electronic health records (EHR) of large numbers of individuals hold promise for a range of research purposes, including prediction research.[1] EHR cover both structured (e.g., code-based demographics, medical history, prescriptions) and unstructured (e.g., presenting signs and symptoms captured in clinical notes) data. Data from clinical notes may be affected by incomplete and selective reporting and manual annotation is typically labour intensive and costly.[2] As a result, most prediction studies using EHR data mainly rely on structured EHR data, albeit the potential of unstructured EHR data is increasingly recognised as a novel and rich source of information.[3–5] Recent advances in the field of natural language processing (NLP), specifically large language models (LLMs), allow for efficient and automatic extraction of relevant unstructured data from EHR clinical notes.[6] Whether such NLP-derived data could improve the predictive performance of models consisting of structured EHR data is, however, yet to be determined.

A field that may benefit from NLP-derived signs and symptoms is that of prognostic research on patients presenting to their general practitioner (GP) with a lower respiratory tract infection (LRTI). In patients suspected of LRTI, presenting signs and symptoms provide important diagnostic information.[7,8] Furthermore, signs and symptoms such as confusion, sputum, pulmonary crackles, and ill appearance have previously been identified as potentially relevant predictors for adverse outcomes in LRTI patients, and are typically captured in clinical notes.[9–12] A recently developed and externally validated EHR-based prediction model, predicting individual risk of 30-day hospitalisation or mortality in patients with GP-diagnosed LRTI, includes predictors based on structured EHR data, i.e., demographics, medical history, current medication use, and LRTI diagnosis. This model may thus benefit from adding signs and symptoms.

In this study, we therefore aimed to evaluate the incremental value of NLP-derived predictors beyond structured EHR data for predicting 30-day all-cause hospitalisation or mortality in patients presenting to general practice with an LRTI.

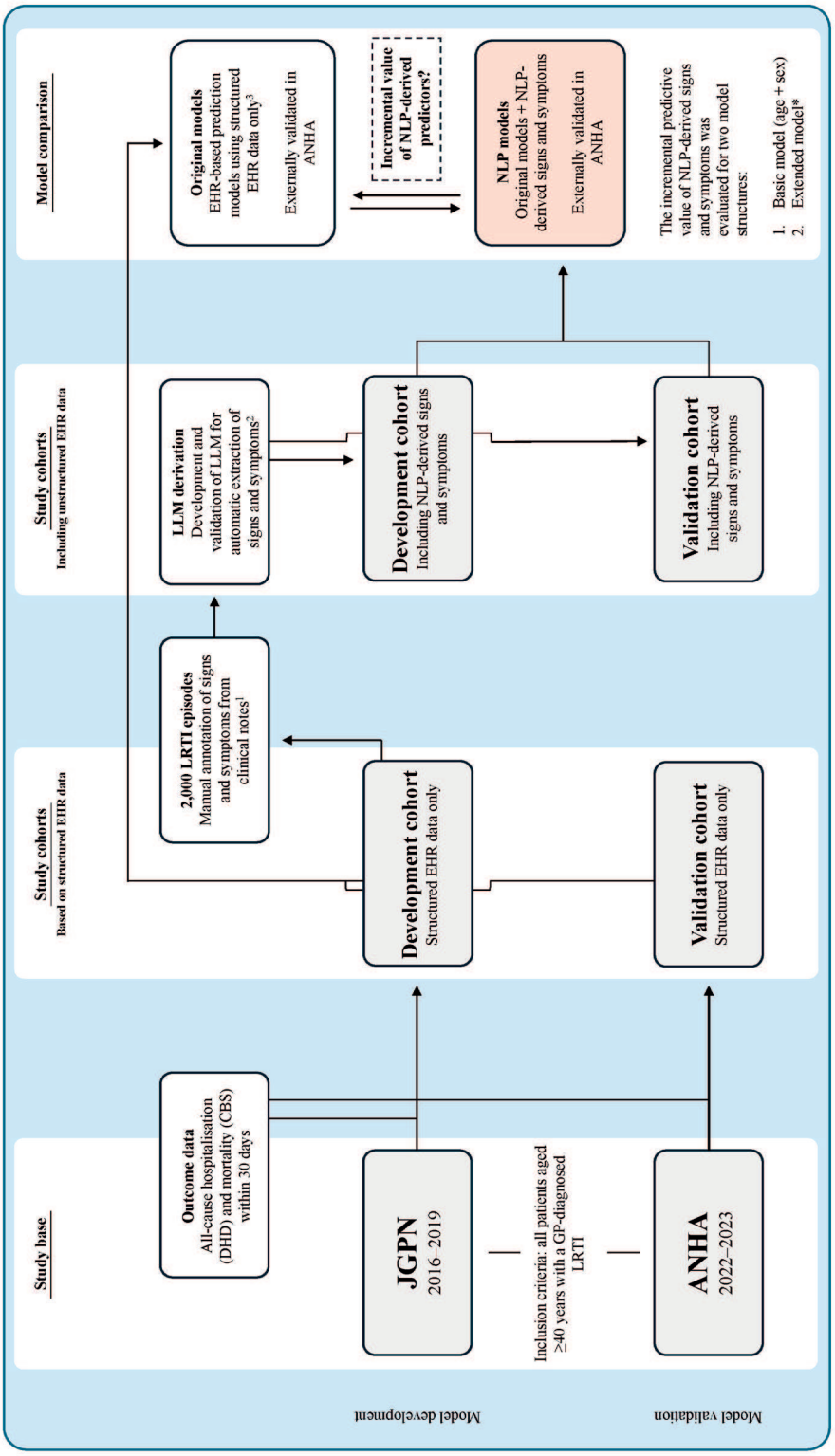


Figure 1. Graphical overview of study structure embedded in previous research

[1] Rijk MH et al. Incomplete and possibly selective recording of signs, symptoms, and measurements in free text fields of primary care electronic health records of adults with lower respiratory tract infections. *J Clin Epidemiol*. 2024 Feb;**166**:111240.

[2] Chapter 6 of this thesis, submitted for publication.

[3] Chapter 5 of this thesis, submitted for publication.

* Predictors of the extended model based on structured EHR data include: age, sex, number of cardiometabolic diseases (0, 1, ≥2), medical history (pneumonia, malignancy, lung disease, dementia, and hospitalisation <1 year), influenza vaccination <1 year, antibiotic prescription <1 month, current medication use (immunosuppressants, inhalation medication, and antidepressants), and diagnosis of pneumonia.

EHR, electronic health records; JGPN, Julius General Practitioners' Network; ANHA, Academic Network of General Practice Amsterdam; GP, general practitioner, LRTI, lower respiratory tract infection; LLM, large language model; NLP, natural language processing.

METHODS

While reporting this study, we adhered to the prediction model risk of bias assessment tool (PROBAST+AI) criteria and transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD+AI) statement.[13,14]

Study design, databases, and participants

This study builds on results of our previous prediction model development and external validation study which solely relied on structured EHR data (Figure 1). In the current prediction model development and external validation study, we systematically evaluated the incremental predictive value of NLP-derived signs and symptoms retrieved from clinical notes beyond structured EHR data. Details on the rationale, cohort construction, and sample size considerations have been reported elsewhere.[15]

In short, cohorts for model development and external validation were derived from the Julius General Practitioners' Network (JGPN)[16] and the Academic Network of General Practice Amsterdam (ANHA)[17], respectively. These general practice registries hold routine EHR data on patient demographics, coded medical history and new diagnoses (based on the International Classification of Primary Care (ICPC)), prescriptions (based on Anatomical Therapeutic Chemical (ATC) codes), and consultations, including clinical notes. JGPN covers data on approximately 450,000 patients from around 70 participating practices from the region of Utrecht, the Netherlands, whereas ANHA holds data on approximately 600,000 patients from 125 participating practices from the region of Amsterdam, the Netherlands.

For model development, all participants aged ≥40 years enlisted in JGPN participating practices between 1 January 2016 and 31 December 2019 with a GP-diagnosis of LRTI – defined as either acute bronchitis (ICPC code R78) or pneumonia (R81) – were included in the development cohort. For model validation, all participants aged ≥40 years enlisted in ANHA participating practices between 1 January 2022 and 31 December 2023 (post-COVID-19-pandemic) with a GP-diagnosis of LRTI were included in the external validation cohort.

In both cohorts, an LRTI-related consultation after a period of 28 days or more without such consultation was considered a new LRTI episode. In case of multiple LRTI episodes per individual patient, only the first episode during the study period was included. The years of the COVID-19 pandemic were deliberately not included in the study period since LRTI aetiology was dominated by SARS-COV-2 and presenting signs and symptoms as well as hospitalisation and mortality rates may not resemble those of the current (post-pandemic) era.[18]

General practice data of the included JGPN and ANHA participants were linked to data on hospitalisations (Dutch Hospital Data (DHD)[19]) and the national mortality registry (Statistics Netherlands (CBS)) on the patient level.

Candidate predictors

Candidate predictors were measured on the day of first GP consultation within an LRTI episode (i.e., index date). Predictors of our previously validated model of structured EHR data included demographics (age and sex), cardiometabolic diseases (diabetes, heart failure, ischaemic heart disease, stroke, transient ischaemic attack, pulmonary embolism, deep venous thrombosis, atrial fibrillation, and peripheral artery disease), other medical history (history of pneumonia, non-dermatological malignancies, chronic obstructive pulmonary disease, asthma, dementia, and hospitalisation <1 year, influenza vaccination <1 year), current medication use (antibiotic prescription <1 month, immunosuppressants, inhalation medication, and antidepressants), and LRTI diagnosis (either acute bronchitis or pneumonia).

Relevant signs and symptoms from clinical notes – selected based on a systematic literature review and with recording patterns previously reported[2,9] – included patient-reported symptoms (chest pain, chills, cough, fever, shortness of breath, and sputum), patient- or GP-reported confusion, and GP-reported signs (pulmonary crackles, ill appearance). Information on these signs and symptoms was retrieved from clinical notes on the index date by NLP. To this end, an LLM based on MedRoBERTa.nl[20], specifically fine-tuned and validated for this task on data of the development cohort, classified each sign or symptom as either 'present', 'absent', or 'not reported'.[21] An overview of all predictors and their code-based definitions is included in the Supplementary material (Table S1).

Outcome

The outcome was defined as all-cause hospitalisation or mortality within 30 days from LRTI diagnosis and was considered binary. Outcome data was derived from the DHD (hospitalisation) and CBS databases (mortality).

Missing data

For structured EHR data, absence of a registered ICPC (for medical history) or ATC (for prescriptions) code was classified as absent. As a result, data of structured EHR-based predictors were fully complete. Given the complex recording patterns of the NLP-derived candidate predictors – due to both incomplete and selective recording[2] – the missing at random assumption for signs and symptoms was likely violated and multiple imputation techniques may therefore not be appropriate.[22] Missing data on NLP-derived candidate predictors was therefore explicitly coded as a separate category, i.e. 'not reported'.

Statistical analysis

Characteristics of the model development and validation cohorts were descriptively summarised. The incremental predictive value of NLP-derived candidate predictors was evaluated based on comparison of four models: a basic model (including age, sex, and an interaction term; model 1), the basic model plus NLP-derived candidate predictors (model 2), an extended model of structured EHR-based predictors of demographics, medical history, current medication use, and LRTI diagnosis (model 3), and the extended model plus NLP-derived candidate predictors (model 4) (Table 1). Age was modelled as linear and all other candidate predictors as categorical candidate predictors. The prediction models were developed using elastic net penalised multivariable logistic regression to account for predictor selection and minimise potential overfitting. The optimal regularisation penalty was determined by hyperparameter tuning with 10-fold cross-validation. To further account for model optimism, all models were internally validated by 1,000 bootstrapping samples. Lastly, the models were externally validated in the post-pandemic validation cohort. Upon internal and external validation, model performance was compared by measures of discrimination (c-statistic), calibration (intercept, slope, and flexible calibration curve), and the range of predicted risks, with accompanying 95% confidence intervals (CI) where appropriate.

For models 3 and 4, predicted risks were compared within individual patients to evaluate shifts in predicted risks. Differences in predicted risks by the models with and without NLP-derived predictors were further explored across age, sex, and LRTI diagnosis. Ultimately, to evaluate whether NLP-derived predictors improved the models' clinical utility, we performed decision curve analyses for model 1 versus model 2 and model 3 versus model 4. In the absence of a single established risk threshold for close monitoring in GP-diagnosed LRTI patients, decision curve analyses were performed for a reasonable range of thresholds probabilities for a close monitoring strategy, i.e., from 20% to 50%.[23]

To explore the hypothesis that signs and symptoms are selectively reported, i.e., not reported signs and symptoms are most likely absent, we conducted a sensitivity analysis in which we combined 'absent' and 'not reported' values into a single category and subsequently evaluated the impact on model performance.

All analysis were performed in R version 4.2.2(24) using the *pmsampsize* (sample size calculations), *glmnet* (model fitting), *CalibrationCurves* (model performance assessment), and *dcurves* (decision curve analysis) packages.

Table 1. Overview of developed and externally validated models

Model	Candidate predictors
Model 1	Age and sex (basic)
Model 2	Basic plus NLP-derived signs and symptoms[1]
Model 3	Age, sex, medical history, current medication use, and LRTI diagnosis[2] (extended)
Model 4	Extended plus NLP-derived signs and symptoms[1]

[1] NLP-derived signs and symptoms include patient reported (chest pain, chills, cough, fever, shortness of breath, sputum), GP-reported (ill appearance, pulmonary crackles), and patient or GP-reported (confusion) signs and symptoms categorised as either present, absent, or not reported.

[2] Additional predictors based on structured EHR data included medical history (number of cardiometabolic diseases, history of pneumonia, non-dermatological malignancies, chronic lung disease, dementia, and hospitalisation <1 year, influenza vaccination <1 year), current medication use (antibiotic prescription <1 year, immunosuppressants, inhalation medication, and antidepressants), and LRTI diagnosis (either pneumonia or acute bronchitis).

NLP, natural language processing; LRTI, lower respiratory tract infection.

RESULTS

Study participants

Within JGPN, 11,702 patients had at least one GP-diagnosis of LRTI between 2016–2019, of whom 11,065 (94.6%) could be linked to hospitalisation and mortality data and were included in the development cohort (Table 2). Pneumonia was diagnosed in 6,518 (58.9%) patients and 6,206 (56.1%) were female. A total of 859 (7.8%) were hospitalised and 21 (0.2%) died within 30 days from LRTI diagnosis, accounting for 863 (7.8%) composite outcomes.

Within ANHA, 6,513 patients had at least one GP-diagnosed LRTI episode between 2022–2023. Of these, 6,284 (96.5%) could be linked to hospitalisation and mortality data and were included in the validation cohort. Pneumonia was diagnosed in 4,426 (70.4%) patients and 3,575 (56.9%) were female. A total of 679 (10.8%) patients of the validation cohort were hospitalised and 73 (1.2%) died within 30 days from LRTI diagnosis, accounting for 717 (11.4%) composite outcomes.

Table 2. Characteristics of model development and validation cohort at index date

	Development cohort		Validation cohort	
Characteristic	**Total**	**With outcome**	**Total**	**With outcome**
n (%)	11065	863 (7.8)	6284	717 (11.4)
Year, n (%)				
2016	2720 (24.6)	196 (22.7)	-	-
2017	2555 (23.1)	186 (21.6)	-	-
2018	3065 (27.7)	237 (27.5)	-	-
2019	2725 (24.6)	244 (28.3)	-	-
2022	-	-	3089 (49.2)	367 (51.2)
2023	-	-	3195 (50.8)	350 (48.8)
Structured EHR data				
Median age (IQR)	64 (53 – 74)	70 (61 – 78)	66 (55 – 76)	72 (62 – 81)
Female sex, n (%)	6206 (56.1)	451 (52.3)	3575 (56.9)	395 (55.1)
LRTI diagnosis, n (%)				
Acute bronchitis	547 (41.1)	164 (19.0)	1858 (29.6)	79 (11.0)
Pneumonia	6518 (58.9)	699 (81.0)	4426 (70.4)	638 (89.0)
Comorbidities, n (%)				
Diabetes	578 (5.2)	67 (7.8)	246 (3.9)	37 (5.2)
Malignancy	1073 (9.7)	154 (17.8)	756 (12.0)	133 (18.5)
Dementia	526 (4.8)	68 (7.9)	498 (7.9)	84 (11.7)
COPD or asthma	2279 (20.6)	239 (27.7)	1414 (22.5)	188 (26.2)
Number of CMD, n (%)				
0	7859 (71.0)	498 (57.7)	4199 (66.8)	348 (48.5)
1	2161 (19.5)	211 (24.4)	1366 (21.7)	232 (32.4)
≥2	1045 (9.4)	154 (17.8)	719 (11.4)	137 (19.1)
Medical history, n (%)				
Pneumonia	3462 (31.3)	340 (39.4)	2079 (33.1)	282 (39.3)
Hospitalisation <1 year	2387 (21.6)	388 (45.0)	1377 (21.9)	291 (40.6)
Median number of GP-consultations <1 year (IQR)	8 (3 – 15)	12 (6 – 21)	5 (1 – 12)	8 (2 – 17)
Medication use, n (%)				
Immunosuppressant	118 (1.1)	16 (1.9)	125 (2.0)	18 (2.5)
Inhalation medication	806 (7.3)	102 (11.8)	358 (5.7)	39 (5.4)
Antidepressant	649 (5.9)	68 (7.9)	363 (5.8)	52 (7.3)
Antibiotics <1 month	804 (7.3)	116 (13.4)	482 (7.7)	88 (12.3)
Influenza vaccination <1 year, n (%)	3327 (30.1)	328 (38.0)	1880 (29.9)	252 (35.1)
NLP-derived signs and symptoms				
Patient reported				
Chest pain				
Present	1704 (15.4)	142 (16.5)	1151 (18.3)	129 (18.0)
Absent	806 (7.3)	62 (7.2)	685 (10.9)	90 (12.6)
Not reported	8555 (77.3)	659 (76.4)	4448 (70.8)	498 (69.5)

Table 2. Continued

Characteristic	Development cohort		Validation cohort	
	Total	With outcome	Total	With outcome
Chills				
Present	741 (6.7)	75 (8.7)	482 (7.7)	61 (8.5)
Absent	<10 (*)	<10 (*)	<10 (*)	<10 (*)
Not reported	<10324 (<93.3*)	<788 (*)	<5802 (<92.3)	<656 (<91.5)
Cough				
Present	8865 (80.1)	573 (66.4)	5230 (83.2)	520 (72.5)
Absent	45 (0.4)	10 (1.2)	20 (0.3)	<10 (*)
Not reported	2155 (19.5)	280 (32.4)	1034 (16.5)	<197 (<27.5)
Fever				
Present	3237 (29.3)	226 (26.2)	2028 (32.3)	203 (28.3)
Absent	4129 (37.3)	271 (31.4)	2568 (40.9)	263 (36.7)
Not reported	3699 (33.4)	366 (42.4)	1688 (26.9)	251 (35.0)
Shortness of breath				
Present	5481 (49.5)	432 (50.1)	3543 (56.4)	416 (58.0)
Absent	1331 (12.0)	91 (10.5)	852 (13.6)	92 (12.8)
Not reported	4253 (38.4)	340 (39.4)	1889 (30.1)	209 (29.1)
Sputum				
Present	3431 (31.0)	223 (25.8)	2256 (35.9)	187 (26.1)
Absent	406 (3.7)	22 (2.5)	227 (3.6)	29 (4.0)
Not reported	7228 (65.3)	618 (71.6)	3801 (60.5)	501 (69.9)
GP reported				
Ill appearance				
Present	1029 (9.3)	86 (10.0)	701 (11.2)	107 (14.9)
Absent	3001 (27.1)	204 (23.6)	1922 (30.6)	171 (23.8)
Not reported	7035 (63.6)	573 (66.4)	3661 (58.3)	439 (61.2)
Pulmonary crackles				
Present	2747 (24.8)	257 (29.8)	1988 (31.6)	262 (36.5)
Absent	6589 (59.5)	393 (45.5)	3435 (54.7)	315 (43.9)
Not reported	1729 (15.6)	213 (24.7)	861 (13.7)	140 (19.5)
Patient or GP reported				
Confusion				
Present	32 (0.3)	11 (1.3)	24 (0.4)	<10 (*)
Absent	277 (2.5)	32 (3.7)	168 (2.7)	31 (4.3)
Not reported	10756 (97.2)	820 (95.0)	6092 (96.9)	<686 (<95.7)

EHR, electronic health record; IQR, interquartile range; LRTI, lower respiratory tract infection; COPD, chronic obstructive pulmonary disease; CMD, cardiometabolic disease; GP, general practitioner.

* As per output restrictions of Statistics Netherlands, it is not allowed to present data based on fewer than 10 observations. Observations based on fewer than 10 observations are therefore indicated with '<10'.

NLP-derived signs and symptoms

In general, signs and symptoms reporting rates yielded by NLP were similar or slightly higher in the validation cohort compared to the development cohort. Frequently reported signs and symptoms included pulmonary crackles (84.4% and 86.3% in the development and validation cohort, respectively), cough (80.5% and 83.5%), fever (66.6% and 73.1%), and shortness of breath (61.6% and 69.9%), whereas confusion (2.8% and 3.1%) and chills (>6.7% and >7.7%) were rarely reported. Signs and symptoms that were frequently classified as 'present' included cough (80.1% and 83.2%), shortness of breath (49.5% and 56.4%), and sputum (31.0% and 35.9%), whereas confusion (2.5% and 2.7%) and ill appearance (27.1% and 30.6%) were mostly classified as 'absent' when reported.

Model development and internal validation

Apparent model performances and linear predictors of all four models are presented in Supplementary materials (Table S2, Table S3). Model performance remained stable upon internal validation (Table 3). For the model 1 versus model 2, discrimination improved after adding NLP-derived predictors (c-statistic 0.61 (95% CI 0.59–0.63) versus 0.67 (0.65–0.69)), whereas model calibration was stable. Upon adding NLP-derived predictors, the range of predicted risks increased from 3.8%–20.8% (median 7.4%) to 1.6%–73.4% (median 6.5%). For model 3 versus model 4, discrimination only slightly increased after including NLP-derived predictors (c-statistic 0.73 (95% CI 0.71–0.74) versus 0.75 (0.74–0.77)) and model calibration remained stable. The range of predicted risks slightly increased after including NLP-derived predictors, i.e., from 1.5%–53.4% (median 5.9%) to 1.1%–69.0% (5.6%) (Figure 2). For model 3 versus model 4, differences between individually predicted risks with and without NLP increased with age, but no substantial differences were observed across sex and LRTI diagnosis (Figure S4).

Table 3. Comparisons of model performance upon internal validation (1,000 bootstrapping samples) in the development cohort.

Model	Candidate predictors	#RC	Model performance (95% CI)			Range of predicted risks (median)
			c-statistic	Calibration intercept	Calibration slope	
Model 1 (Basic)	Age, sex, and interaction term	4	0.61 (0.59 – 0.63)	0.00 (-0.07 – 0.07)	1.04 (0.86 – 1.23)	3.8% – 20.8% (7.4%)
Model 2 (Basic plus NLP)	Model 1 + NLP-derived signs and symptoms[1]	22	0.67 (0.65 – 0.69)	0.00 (-0.07 – 0.07)	1.03 (0.92 – 1.15)	1.6% – 73.4% (6.5%)
Model 3 (Extended)	Model 1 + LRTI diagnosis, medical history, and current medication use[2]	17	0.73 (0.71 – 0.74)	0.00 (-0.07 – 0.07)	1.03 (0.94 – 1.12)	1.5% – 53.4% (5.9%)

Table 3. Continued

Model	Candidate predictors	#RC	Model performance (95% CI)			Range of predicted risks (median)
			c-statistic	Calibration intercept	Calibration slope	
Model 4 (Extended plus NLP)	Model 3 + NLP-derived signs and symptoms[1]	35	0.75 (0.74 – 0.77)	0.00 (-0.07 – 0.07)	1.04 (0.95 – 1.12)	1.1% – 69.0% (5.6%)

[1] NLP-derived signs and symptoms include patient reported (chest pain, chills, cough, fever, shortness of breath, sputum), GP-reported (ill appearance, pulmonary crackles), and patient or GP-reported (confusion) signs and symptoms categorised as either present, absent, or not reported.

[2] Additional predictors based on structured EHR data included LRTI diagnosis (either pneumonia or acute bronchitis), medical history (number of cardiometabolic diseases, history of pneumonia, non-dermatological malignancies, chronic lung disease, dementia, and hospitalisation <1 year, influenza vaccination <1 year), and current medication use (antibiotic prescription <1 year, immunosuppressants, inhalation medication, and antidepressants).

CI, confidence interval; #RC, number of regression coefficients; NLP, natural language processing; LRTI, lower respiratory tract infection; GP, general practitioner; EHR, electronic health records.

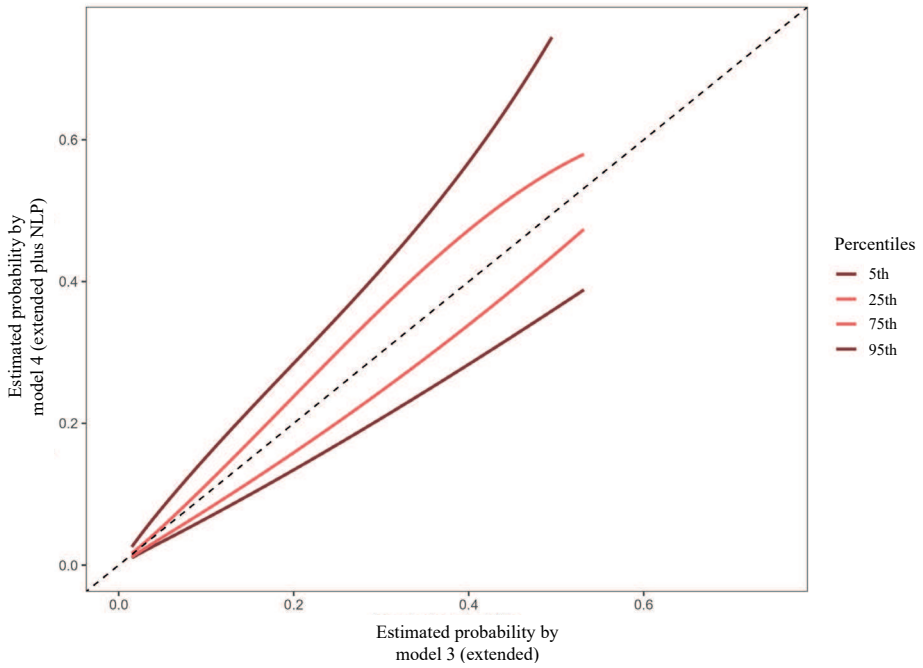


Figure 2. Individually predicted risks by model 3 versus model 4 in the model development cohort

NLP, natural language processing

External validation

Upon external validation, model discrimination was relatively stable. External validation of the models 1 and 2 yielded a c-statistic of 0.63 (95% CI 0.60–0.65) for the basic model and 0.67 (0.65–0.69) for the basic model with NLP-derived predictors (Figure 3). In terms of calibration, intercepts were similar but the basic model including NLP-derived predictors yielded a favourable slope (1.16 (0.95–1.37) versus 1.01 (0.87–1.14), respectively). In addition, including NLP-derived predictors had a minor impact on the distribution of predicted risks (Figure 4).

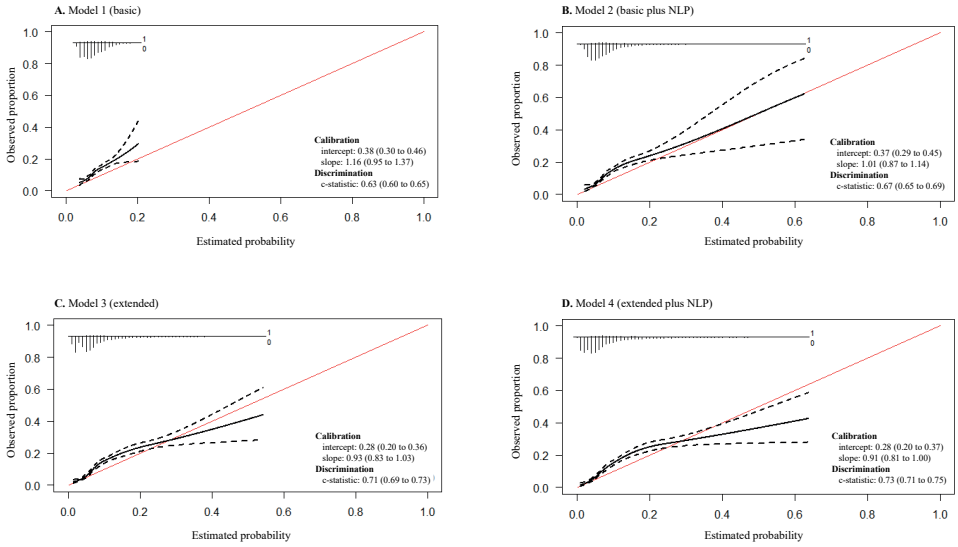


Figure 3. Flexible calibration plots of all four models upon external calibration

NLP, natural language processing

For model 3 versus model 4, model discrimination was slightly favourable with incorporation of NLP-derived predictors (c-statistic 0.71 (0.69–0.73) versus 0.73 (0.71–0.75)). Both models yielded a similar calibration intercept and slope without substantial shifts in the distribution of predicted risks (1.5%–51.3% (median 5.9%) versus 1.2%–63.3% (6.9%)).

Visual inspection of flexible calibration plots in the validation cohort revealed modest underprediction up to observed risks of around 20% for all four models (Figure 3). Calibration of the basic model including NLP-derived predictors and the extended model based on structured EHR data alone was favourable for patients with a higher predicted risk.

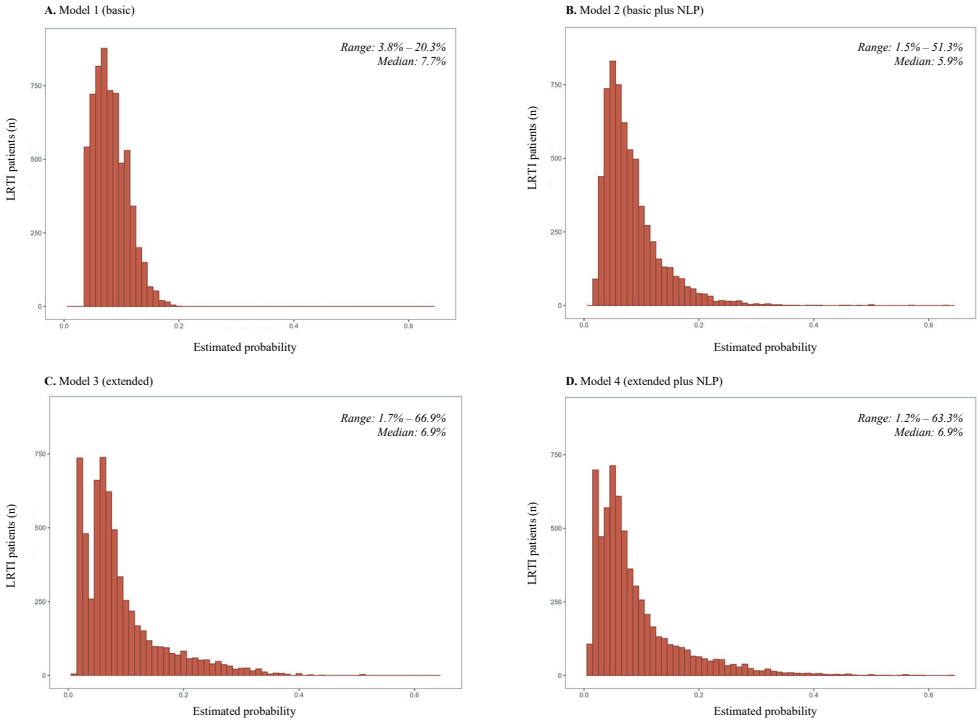


Figure 4. Distribution of predicted risks of all four models upon external validation

NLP, natural language processing

Decision curve analysis

Decision curve analyses for threshold probabilities ranging from 20% to 50% for all models revealed a higher net benefit of the extended models compared to the basic models up to a probability threshold of approximately 30% (Figure 5). Net benefit of model 3 and model 4 was generally similar over the range of threshold probabilities.

Sensitivity analysis

Upon reclassifying 'not reported' signs and symptoms to 'not present', model performance did not substantially alter (Table S5).

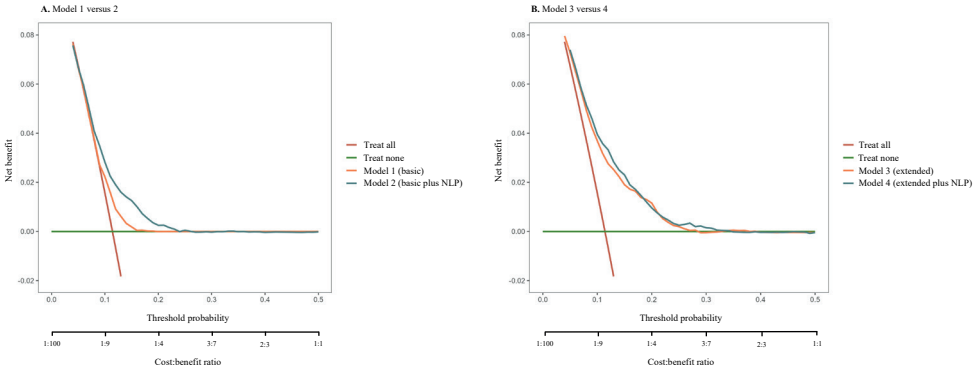


Figure 5. Decision curve analysis of all developed models upon external validation
NLP, natural language processing

DISCUSSION

In this study, we evaluated the incremental predictive value of signs and symptoms retrieved from unstructured clinical notes by using an LLM specifically fine-tuned and validated for this task beyond information based on structured EHR data for predicting 30-day hospitalisation or mortality in LRTI patients presenting to general practice. Albeit adding NLP-derived signs and symptoms improved model performance beyond a basic model consisting of age and sex, the incremental value of these NLP-derived predictors – as measured by discrimination, calibration, distribution of predicted risks, and net benefit – beyond structured EHR-based predictors was limited, and likely not clinically relevant.

Our findings imply that both structured data and unstructured signs and symptoms data hold valuable predictive information, but with low degree of complementarity in the context of this case study. The incremental value of predictors from unstructured EHR data is likely case dependent, with a systematic review showing variation across care settings and text processing methods.[5] For general practice, the quality of EHR clinical notes – evaluated in a previous study[2] – may particularly suffer from incomplete and selective reporting due to generally low prior probabilities of adverse outcomes, which could limit its predictive value.[25] On the contrary, our finding that reclassifying 'not reported' signs and symptoms as 'absent' did not alter model performance suggests that both classes share similar predictive information, and consequences of selective reporting may thus be limited. Nevertheless, our results are in line with previous work comparing structured and unstructured EHR-based models in general practice, with models including unstructured data yielding similar or at best slightly better performance.[26–29] Of note, in contrast to the present study, previous studies implemented a word embedding approach in

which clinical notes were represented by a vector of numbers. Using NLP to classify individual predictors from clinical notes is, therefore, a novelty in general practice research.

Strengths of this study include the development of an LLM fine-tuned and validated specifically for this study using manually annotated EHR data from 2,000 clinical notes and the thorough assessment of the incremental value of NLP-derived predictors through external validation – not often performed in prediction modelling studies on unstructured EHR data[5] – and decision curve analysis. In addition, our study explored opportunities of NLP methods on Dutch language, whereas most studies to date have focussed on English unstructured text.[5]

Some limitations deserve consideration. First, performance of the LLM in directly classifying candidate predictors from clinical notes decreased with lower reporting rates and increasing class imbalance. As a result, infrequently reported or imbalanced predictors, such as confusion and chills, may have suffered from erroneous LLM classifications, and their predictive value may subsequently be underestimated. In addition, low positivity rates may have deflated the predictive value of these potentially relevant predictors.[30] Second, as stated before, the predictive value of candidate predictors retrieved from unstructured data varies across NLP methods.[5] By implementing our LLM for classification of individual predictors, other potentially predictive information from clinical notes may have partly been omitted. Data registry restrictions, however, did not allow us to link free text clinical notes directly to outcomes on a patient level, hence our focus on classification of selected predictors. Third, since we fine-tuned the LLM to specifically classify signs and symptoms we did not consider relevant vital sign measurements such as peripheral oxygen saturation, pulse rate, blood pressure and respiratory rate captured in clinical notes, although these are likely to capture valuable predictive information from a clinical perspective.[9] Lastly, although we performed external validation of all models in this study, we did not separately validate the LLM – developed in JGPN – on ANHA data to evaluate its performance as direct-classifier in this cohort. Prediction model performances as observed in both JGPN and ANHA in this study were, however, stable, suggesting similar performance of the LLM in both cohorts.

Albeit the predictive value of NLP-derived signs and symptoms beyond age and sex was evident when predicting 30-day all-cause hospitalisation or mortality in LRTI patients presenting to general practice, the incremental value of these predictors beyond a model including all relevant structured EHR-based predictors was limited. This implies that signs and symptoms do not add substantial predictive information in LRTI patients, provided that GPs adequately incorporate all relevant information available from patients' EHR. In addition, given the complexity of implementing an EHR-based prediction model with integrated NLP methods and the negligible improvement in model performance and net benefit, a model based on structured EHR data alone is likely to be preferred in the clini-

cal setting. Nevertheless, our findings underline the predictive value of unstructured EHR data, which thus deserve consideration when developing EHR-based prediction models, certainly in a context where a full patient history is not readily available, e.g., in out-of-hours primary care. Hence, future research in the field of NLP should acknowledge this contextual need to evaluate the impact of integrated NLP approaches on predictive performance of EHR-based models, naturally in the context of continued advancements in the field of NLP. In addition, future prediction modelling studies should consider exploring whether vital sign measurements derived from clinical notes could further improve prognostication of LRTI patients in general practice.

REFERENCES

1. Bots SH, Groenwold RHH, Dekkers OM. Using electronic health record data for clinical research: a quick guide. *Eur J Endocrinol*. 2022;**186**(4):E1–6.
2. Rijk MH, Platteel TN, Mulder AMM, Geersing GJ, Rutten FH, van Smeden M, et al. Incomplete and possibly selective recording of signs, symptoms and measurements in free text fields of primary care electronic health records of adults with lower respiratory tract infections. *J Clin Epidemiol*. 2024:**166**;111240.
3. Seinen TM, Kors JA, van Mulligen EM, Rijnbeek PR. Using structured codes and free-text notes to measure information complementarity in electronic health records: feasibility and validation study. *J Med Internet Res*. 2025;**27**:e66910.
4. Tayefi M, Ngo P, Chomutare T, Dalianis H, et al. Challenges and opportunities beyond structured data in analysis of electronic health records. *Wiley Interdiscip Rev Comput Stat*. 2021;**13**:1695907.
5. Seinen TM, Fridgeirsson EA, Ioannou S, et al. Use of unstructured text in prognostic clinical prediction models: a systematic review. *J Am Med Inform Assoc*. 2022;**29**(7):1292–1302.
6. Ntinopoulos V, Rodriguez Cetina Biefer H, Tudorache I, et al. Large language models for data extraction from unstructured and semi-structured electronic health records: a multiple model performance evaluation. *BMJ Health Care Inform*. 2025;**32**(1):e101139.
7. Moore M, Stuart B, Little P, et al. Predictors of pneumonia in lower respiratory tract infections: 3C prospective cough complication cohort study. *Eur Respir J*. 2017;**50**(5):1700434.
8. Van Vugt SF, Broekhuizen BDL, Lammens C, et al. Use of serum C reactive protein and procalcitonin concentrations in addition to symptoms and signs to predict pneumonia in patients presenting to primary care with acute cough: diagnostic study. *BMJ*. 2013;**346**(7909):f2450.
9. Rijk MH, Platteel TN, van den Berg TMC, et al. Prognostic factors and prediction models for hospitalisation and all-cause mortality in adults presenting to primary care with a lower respiratory tract infection: a systematic review. *BMJ Open*. 2024;**14**(3):e075475.
10. Lim WS, Van Der Eerden MM, Laing R, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax*. 2003;**58**(5):377–82.
11. Moore M, Stuart B, Lown M, et al. Predictors of adverse outcomes in uncomplicated lower respiratory tract infections. *Ann Fam Med*. 2019;**17**(3):231–8.
12. Bruyndonckx R, Hens N, Verheij TJM, 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.
13. Moons KGM, Damen JAA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ*. 2025;**388**:e082505.
14. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;**385**:e078378.
15. Rijk MH, Platteel TN, Geersing GJ, et al. 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. *Diagn Progn Res*. 2023;**7**(1):23.
16. Smeets H, Kortekaas M, Rutten F, 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.

17. Amsterdam UMC. *Academisch Netwerk Huisartsgeneeskunde Amsterdam (ANHA)*. Available from: https://www.vumc.nl/research/overzicht/academisch-netwerk-huisartsgeneeskunde-amsterdam-umc-anha.html.
18. Chow EJ, Uyeki TM, Chu HY. The effects of the COVID-19 pandemic on community respiratory virus activity. *Nat Rev Microbiol*. 2023;**21**(3):195–210.
19. Dutch Hospital Data (DHD). *Landelijke Basisregistratie Ziekenhuiszorg (LBZ) - years: 2016-2021*. Available from: https://www.dhd.nl/producten-diensten/lbz/paginas/dataverzameling-lbz.aspx.
20. Verkijk S, Vossen P. MedRoBERTa.nl: a language model for Dutch electronic health records. *Comp Linguistics Neth J*. 2021;**11**:141–59.
21. Spiero I, Rijk MH, Scheeres MA, et al. Comparison of local large language models for extraction of signs and symptoms data from electronic health records. *medRxiv* (preprint). 2025. Available from: https://www.medrxiv.org/content/10.64898/2025.12.11.25341954v1.full.
22. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;**339**(7713):157–60.
23. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*. 2016;**352**:i6.
24. R Core Team. *R: A Language and Environment for Statistical Computing* [Internet]. R Foundation for Statistical Computing, Vienna, Austria; 2022. Available from: https://www.R-project.org/.
25. Groenwold RHH. Informative missingness in electronic health record systems: the curse of knowing. *Diagn Progn Res*. 2020;**4**:8.
26. Seinen TM, Kors JA, van Mulligen EM, Fridgeirsson E, Rijnbeek PR. The added value of text from Dutch general practitioner notes in predictive modeling. *J Am Med Inform Assoc*. 2023;**30**(12):1973–1984.
27. Song J, Hobensack M, Bowles KH, et al. Clinical notes: an untapped opportunity for improving risk prediction for hospitalization and emergency department visit during home health care. *J Biomed Inform*. 2022:**128**:104039.
28. Luik TT, Abu-Hanna A, van Weert HCPM, Schut MC. Early detection of colorectal cancer by leveraging Dutch primary care consultation notes with free text embeddings. *Sci Rep*. 2023;**13**(1):10760.
29. Schut MC, Luik TT, Vagliano I, et al. Artificial intelligence for early detection of lung cancer in GPs' clinical notes: a retrospective observational cohort study. *Br J Gen Pract*. 2025;**75**(754):e316–22.
30. Schummers L, Himes KP, Bodnar LM, Hutcheon JA. Predictor characteristics necessary for building a clinically useful risk prediction model: a simulation study. *BMC Med Res Methodol*. 2016;**16**(1):123.

SUPPLEMENTARY MATERIALS FOR CHAPTER 7

Table S1. Details on candidate predictors and definitions

Candidate predictor	Definition or underlying registration(s)	Codes, if applicable		Data source		Values
		ICPC	ATC	Development	Validation	
Demographics						
Age (years)	Age in years at day of LRTI diagnosis			JGPN	ANHA	40 – infinity (integer)
Sex (male, female)	Biological sex			JGPN	ANHA	Male, female
Cardiometabolic comorbidities						
Diabetes	Diabetes mellitus type 1 and 2	T90		JGPN	ANHA	Yes, no
Atrial fibrillation	Atrial fibrillation/flutter	K78		JGPN	ANHA	Yes, no
Cerebrovascular disease	CVA TIA	K90 K89		JGPN	ANHA	Yes, no
Coronary artery disease	Acute myocardial infarction Angina pectoris Other chronic/ischemic heart disease	K75 K74 K76		JGPN	ANHA	Yes, no
Heart failure	Congestive heart failure	K77		JGPN	ANHA	Yes, no
Venous thromboembolism	Pulmonary embolism Deep venous thrombosis	K93 K94.01		JGPN	ANHA	Yes, no
Peripheral artery disease	Intermittent claudication	K92.01		JGPN	ANHA	Yes, no
Health care use						
Number of GP consultations ≤12 months	Number of GP consultations in the year prior to LRTI diagnosis			JGPN	ANHA	0 – infinity (integer)
Hospitalization ≤12 months	Hospitalisation in the year prior to LRTI diagnosis			DHD	DHD	Yes, no

Table S1. Continued

Candidate predictor	Definition or underlying registration(s)	Codes, if applicable		Data source		Values
		ICPC	ATC	Development	Validation	
Vaccination status						
Influenza vaccination in previous year	Immunisation/preventive medicine of airways	R44	J07AL01 (<1 year)	JGPN	ANHA	Yes, no
Other comorbidities						
History of pneumonia	Pneumonia	R81		JGPN	ANHA	Yes, no
COPD or asthma	Emphysema/COPD	R95		JGPN	ANHA	Yes, no
	Chronic bronchitis/bronchiectasis	R91				
	Asthma	R96				
Dementia	Cognitive disorders	P20		JGPN	ANHA	Yes, no
	Dementia	P70				
Non-dermatological malignancy	Hodgkin lymphoma	B72.01		JGPN	ANHA	Yes, no
	Non-Hodgkin lymphoma	B72.02				
	Leukaemia	B73				
	Hematologic/lymphatic malignancy NOS	B74				
	Benign hematologic/lymphatic Neoplasm	B75				
	Stomach malignancy	D74				
	Colon/rectum malignancy	D75				
	Pancreatic malignancy	D76				
	Digestive tract malignancies NOS	D77				
	Malignancy of bronchus/lung	R84				
	Other airway malignancy	R85				
	Malignancy of kidney	U75				
	Malignancy of bladder	U76				
	Other malignancy of urinary tract	U77				

Table S1. Continued

Candidate predictor	Definition or underlying registration(s)	Codes, if applicable		Data source		Values
		ICPC	ATC	Development	Validation	
	Malignancy prostate	Y77				
	Other urologic malignancy	Y78				
	Malignancy of cervix uteri	X75				
	Malignancy of breast (female)	X76				
	Other gynaecological Malignancies	X77				
	Malignancy of unknown primary origin	A79				
	Malignancy of eye/adnexa	F74.01				
	Malignancy of ear	H75.01				
	Cardiovascular malignancy	K72.01				
	Musculoskeletal malignancy	L71.01				
	Neurologic malignancy	N74				
	Malignancy of thyroid	T71				
Medication use[1]						
Antibiotic prescription within one month prior to LRTI diagnosis	Prescription within one month prior to LRTI diagnosis		J01	JGPN	ANHA	Yes, no
Immunosuppressants	Systemic glucocorticoids Predniso(lo)n ≥7.5mg/day Immunosuppressants		H02AB L04A	JGPN	ANHA	Yes, no
Antidepressants	Antidepressants		N06A	JGPN	ANHA	Yes, no
Inhalation medication	Long-acting beta antagonists (LABA):			JGPN	ANHA	Yes, no
	Formoterol		R03AC13			
	Indacaterol		R03AC18			
	Olodaterol		R03AC19			
	Salmeterol		R03AC12			

Table S1. Continued

Candidate predictor	Definition or underlying registration(s)	Codes, if applicable		Data source		Values
		ICPC	ATC	Development	Validation	
	Long-acting muscarine antagonists (LAMA):					
	Aclidinium		R03BB05			
	Glycopyrronium		R03BB06			
	Tiotropium		R03BB04			
	Umeclidinium		R03BB07			
	LABA+LAMA:					
	Aclidinium/formoterol		R03AL05			
	Glycopyrronium/formoterol		R03AL07			
	Indacaterol/glycopyrronium		R03AL04			
	Tiotropium/olodaterol		R03AL06			
	Umeclidinium/vilanterol		R03AL03			
	Inhalation corticosteroids (ICS):					
	Beclomethasone		R03BA01			
	Budesonide		R03BA02			
	Fluticasone		R03BA05			
	ICS+LABA:					
	Formoterol/beclomethasone		R03AK08			
	Formoterol/budesonide Salmeterol/fluticasone propionate		R03AK07			
	Vilanterol/fluticasone furoate		R03AK06			

Table S1. Continued

Candidate predictor	Definition or underlying registration(s)	Codes, if applicable		Data source		Values
		ICPC	ATC	Development	Validation	
	ICS+LABA+LAMA					
	Beclomethasone/formoterol/glycopyrronium		R03AL09			
	Fluticasone/umeclidinium/vilanterol		R03AL08			
	Formoterol/glycopyrronium/budesonide		R03AL11			
Signs and symptoms						
Patient reported						
Chest pain	Classified from clinical notes from first LRTI consultation within episode by modified MedRoBERTa.nl large language model.[2,3]			JGPN	ANHA	Present, absent, not reported
Chills	Classified from clinical notes from first LRTI consultation within episode by modified MedRoBERTa.nl large language model.[2,3]			JGPN	ANHA	Present, absent, not reported
Cough	Classified from clinical notes from first LRTI consultation within episode by modified MedRoBERTa.nl large language model.[2,3]			JGPN	ANHA	Present, absent, not reported
Fever	Classified from clinical notes from first LRTI consultation within episode by modified MedRoBERTa.nl large language model.[2,3]			JGPN	ANHA	Present, absent, not reported
Shortness of breath	Classified from clinical notes from first LRTI consultation within episode by modified MedRoBERTa.nl large language model.[2,3]			JGPN	ANHA	Present, absent, not reported
Sputum	Classified from clinical notes from first LRTI consultation within episode by modified MedRoBERTa.nl large language model.[2,3]			JGPN	ANHA	Present, absent, not reported

Table S1. Continued

Candidate predictor	Definition or underlying registration(s)	Codes, if applicable		Data source		Values
		ICPC	ATC	Development	Validation	
Patient or GP-reported						
Confusion	Classified from clinical notes from first LRTI consultation within episode by modified MedRoBERTa.nl large language model.[2,3]			JGPN	ANHA	Present, absent, not reported
GP-reported						
Pulmonary crackles	Classified from clinical notes from first LRTI consultation within episode by modified MedRoBERTa.nl large language model.[2,3]			JGPN	ANHA	Present, absent, not reported
Ill appearance	Classified from clinical notes from first LRTI consultation within episode by modified MedRoBERTa.nl large language model.[2,3]			JGPN	ANHA	Present, absent, not reported

Abbreviations: ICPC, International Classification of Primary Care; ATC, anatomical therapeutic chemical; CMD, cardiometabolic disease; LRTI, lower respiratory tract infection; JGPN, Julius General Practitioners' Network; ANHA, Academic Network of General Practitioners Amsterdam; GP, general practitioner; DHD, Dutch Hospital Data; COPD, chronic obstructive pulmonary disease; NOS, not otherwise specified; mg, milligram; LABA, long-acting beta antagonist; LAMA, long-acting muscarine antagonist; ICS, inhalation corticosteroids.

[1] Medication was selected only when commonly used in a primary care setting and is based upon guidelines issued by the Dutch College of General Practitioners and the Dutch pharmacotherapeutic compass.

[2] Verkijk, S, Vossen, P. MedRoBERTa. nl: a language model for Dutch electronic health records. Computational Linguistics in the Netherlands. 221 Dec; 11: 141–59.

[3] Chapter 6 of this thesis

Table S2. Modelling strategy and apparent model performance in development cohort of developed models

Model	Candidate predictors	#RC	Model performance (95% CI)			Range of predicted risks (median)
			c-statistic	Calibration intercept	Calibration slope	
Model 1 (Basic)	Age, sex, and interaction term	4	0.61 (0.59 – 0.63)	0.00 (-0.07 – 0.07)	1.04 (0.86 – 1.23)	3.8% – 20.8% (7.4%)
Model 2 (Basic plus NLP)	Model 1 + NLP-derived signs and symptoms[1]	22	0.67 (0.65 – 0.69)	0.00 (-0.07 – 0.07)	1.01 (0.90 – 1.13)	1.6% – 73.4% (6.5%)
Model 3 (Extended)	Model 1 + LRTI diagnosis, medical history, and current medication use[2]	17	0.73 (0.71 – 0.74)	0.00 (-0.07 – 0.07)	1.02 (0.93 – 1.11)	1.5% – 53.4% (5.9%)
Model 4 (Extended plus NLP)	Model 3 + NLP-derived signs and symptoms[1]	35	0.74 (0.73 – 0.76)	0.00 (-0.07 – 0.07)	1.05 (0.96 – 1.13)	1.1% – 69.0% (5.6%)

[1] NLP-derived signs and symptoms include patient reported (chest pain, chills, cough, fever, shortness of breath, sputum), GP-reported (ill appearance, pulmonary crackles), and patient or GP-reported (confusion) signs and symptoms categorised as either present, absent, or not reported.

[2] Additional predictors based on structured EHR data included LRTI diagnosis (either pneumonia or acute bronchitis), medical history (number of cardiometabolic diseases, history of pneumonia, non-dermatological malignancies, chronic lung disease, dementia, and hospitalisation <1 year, influenza vaccination <1 year), and current medication use (antibiotic prescription <1 year, immunosuppressants, inhalation medication, and antidepressants).

CI, confidence interval; #RC, number of regression coefficients; NLP, natural language processing; LRTI, lower respiratory tract infection; GP, general practitioner; EHR, electronic health records.

Table S3. Linear predictors and regression coefficients of all developed models

Predictor	Regression coefficients			
	Model 1	**Model 2**	**Model 3**	**Model 4**
Intercept	-4.298	-4.007	-4.618	-4.130
Age (linear)	0.029	0.034	0.015	0.015
Female sex	0	0.721	0	0
*Age (linear)*female sex*	-0.003	-0.013	-0.001	-0.001
No. of CMD = 1	-	-	0.115	0.108
No of. CMD ≥ 2	-	-	0.344	0.313
LRTI diagnosis = pneumonia	-	-	1.015	0.928
Hospitalisation <1 year	-	-	0.922	0.874
History of pneumonia	-	-	-0.137	-0.115
Malignancy	-	-	0.393	0.412
COPD or asthma	-	-	0.112	0.038
Dementia	-	-	0.222	0.166
Influenza vaccination <1 year	-	-	0.076	0.056
Immunosuppressant use	-	-	0.182	0.177
Inhalation medication use	-	-	0.082	0.037
Antibiotics <1 month	-	-	0.335	0.241
Antidepressant use	-	-	0.254	0.241
Dyspnoea = present	-	0.205	-	0.171
Dyspnoea = not reported	-	-0.153	-	-0.109
Fever = present	-	0.074	-	0
Fever = not reported	-	0.140	-	00.060
Cough = present	-	-1.148	-	-0.596
Cough = not reported	-	-0.442	-	0
Confusion = present	-	1.087	-	0.969
Confusion = not reported	-	-0.309	-	-0.145
Chest pain = present	-	0.285	-	0.173
Cheast pain = not reported	-	0.003	-	0
Chills = present	-	0.323	-	0.125
Chilles = not reported	-	0	-	-0.067
Ill apprearance = present	-	0.213	-	0.097
Ill appearance = not reported	-	0.034	-	0.065
Pulmonary crackles = present	-	0.340	-	0.061
Pulmonary crackles = not reported	-	0.433	-	0.176
Sputum = present	-	0.134	-	0
Sputum = not reported	-	0.248	-	0.094

Linear predictors of all developed models with regression coefficients per candidate predictor, rounded on three decimal places. Reference category for sex is 'male', No. of CMD is '0', and smoking status is 'current'. For medical history and medication use, the reference category is 'absent'.

CMD, cardiometabolic disease; EHR, electronic health records; RCS, restricted cubic splines; CVA, cerebrovascular accident; TIA, transient ischaemic attack; PE, pulmonary embolism; DVT, deep venous thrombosis; LRTI, lower respiratory tract infection; COPD, chronic obstructive pulmonary disease; GP, general practitioner.

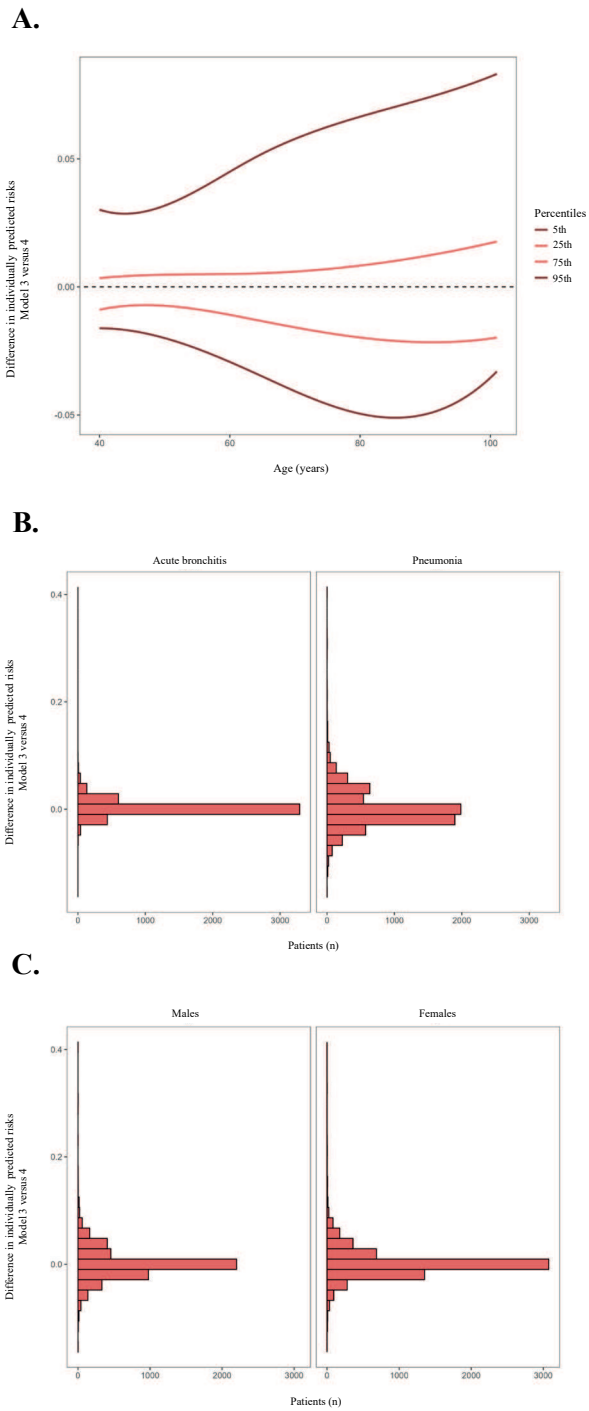


Figure S4. Differences in individually predicted risks by model 3 versus model 4

Differences between predictions of both models were stratified according to **(a)** age, **(b)** LRTI diagnosis, and **(c)** sex.

Table S5. Apparent model performance in development cohort of models from sensitivity analysis, with 'not reported' signs and symptoms reclassified as 'absent' for both model 2 and model 4

Model	Candidate predictors	#RC	Model performance (95% CI)			
			c-statistic	Calibration intercept	Calibration slope	Range of predicted risks (median)
Model 2	Age and sex + NLP-derived signs and symptoms[1]	22	0.67 (0.65 – 0.69)	0.00 (-0.07 – 0.07)	1.01 (0.90 – 1.13)	1.6% – 73.4% (6.5%)
Model 2 *sensitivity*	Age and sex + NLP-derived signs and symptoms[2]	13	0.66 (0.64 – 0.68)	0.00 (-0.07 – 0.07)	1.01 (0.89 – 1.13)	2.0% – 75.3% (6.6%)
Model 4	Demographics, medical history, medication use and LRTI diagnosis + NLP-derived signs and symptoms[1]	35	0.74 (0.73 – 0.76)	0.00 (-0.07 – 0.07)	1.05 (0.96 – 1.13)	1.1% – 69.0% (5.6%)
Model 4 *sensitivity*	Demographics, medical history, medication use and LRTI diagnosis + NLP-derived signs and symptoms[2]	26	0.74 (0.72 – 0.76)	0.00 (-0.07 – 0.07)	1.05 (0.96 – 1.13)	1.2% – 69.3% (5.6%)

[1] NLP-derived signs and symptoms include patient reported (chest pain, chills, cough, fever, shortness of breath, sputum), GP-reported (ill appearance, pulmonary crackles), and patient or GP-reported (confusion) signs and symptoms categorised as either present or absent.

[2] NLP-derived signs and symptoms were classified as either present or absent.

CI, confidence interval; #RC, number of regression coefficients; NLP, natural language processing; LRTI, lower respiratory tract infection; GP, general practitioner; EHR, electronic health records.

8

Excess cardiovascular risk associated with community-acquired lower respiratory tract infections: a population-based nested self-controlled case-series

Merijn H. Rijk
Lonneke A. van Vught
Tamara N. Platteel
Florien S. van Royen
Paul Little
Geert-Jan Geersing
Frans H. Rutten
Patricia C.J.L. Bruijning-Verhagen
Roderick P. Venekamp

Accepted for publication in BMC Medicine

Abstract

Background Lower respiratory tract infections (LRTI) are common and can trigger acute cardiovascular diseases (CVD). Previous studies focussing on this association only reported relative effect measures, leaving the population impact largely unknown.

Methods This population-based study was based on routine health care data of the Julius General Practitioners' Network in the Netherlands and included patients aged ≥40 years presenting with an LRTI to general practice between 2016–2019. The burden of CVD following LRTI was assessed in three consecutive steps. First, 90-day incidence rates of CVD following LRTI were mapped for major adverse cardiac and cerebrovascular events (MACCE), venous thromboembolism (VTE), and new-onset atrial fibrillation (nAF). Second, incidence rate ratios (IRR) within 90 days following LRTI were estimated per CVD type, using a self-controlled case-series analysis. Third, excess CVD cases associated with LRTI were calculated.

Findings In total, 11,065 patients contributed 16,497 LRTI episodes. The 90-day incidence rate per 1,000 LRTI patients was 4.2 (95% CI 3.3–5.3, n=69) for MACCE, 2.7 (1.9–3.6, n=44) for VTE, and 6.6 (5.4–8.1, n=96) for nAF. IRRs of CVD within 0–90 days following LRTI were 1.3 (1.0–1.8) for MACCE, 6.8 (4.4–10.4) for VTE, and 2.9 (2.3–3.7) for nAF. For all CVD types, IRRs were highest in the first week following LRTI. Per 1,000 LRTI patients, 0.5 (-0.5–1.6) to 1.1 (0.1–2.4) excess MACCE, 2.1 (1.1–3.6) to 3.0 (1.8–4.9) excess VTE, and 2.3 (1.1–3.7) to 4.4 (2.9–6.2) excess nAF diagnoses were associated with LRTI.

Interpretation The temporal increase in CVD-risk following LRTI culminates into 5–9 excess acute CVD per 1,000 LRTI patients aged ≥40 years. Given the large number of LRTI patients during seasonal outbreaks, this implies a need for targeted preventative interventions and patient-focussed campaigns to increase seasonal vaccine uptake.

INTRODUCTION

Lower respiratory tract infections (LRTI) are among the most common infections encountered in general practice, with an annual incidence of approximately three percent.[1] Evidence is accumulating that LRTIs increase risk of acute cardiovascular disease (CVD), such as myocardial infarction (MI), stroke, venous thromboembolism (VTE), and atrial fibrillation.[2–5] This risk is particularly increased during the first two weeks following infection, but remains elevated until at least 90 days after LRTI diagnosis.[2,3] Specific pathogens, notably S. *pneumoniae*, influenza and SARS-CoV-2 virus, have particularly been associated with cardiovascular sequelae, and clinical evidence from the COVID-19 pandemic further underlined the association between viral LRTI and CVD.[6–9]

Previous studies focussing on the association between LRTIs and CVD in general practice mainly relied on relative effect measures, with landmark studies published some 20 years ago.[2–4] However, relative measures do not provide insight into absolute excess CVD risk associated with LRTI, leaving the population impact and need for risk mitigating strategies largely unknown.

To unravel the population impact and guide preventative interventions, evidence on absolute excess CVD risk across CVD types and patient characteristics is pivotal. To this end, we first calculated the absolute 90-day incidence of major adverse cardiac and cerebrovascular events (MACCE), VTE, and new-onset atrial fibrillation (nAF) in adults presenting to their general practitioner (GP) with LRTI across various patient characteristics such age, sex, and co-morbidities. Second, we estimated incidence rate ratios (IRR) for a 90-day risk period following LRTI diagnosis per CVD type using a self-controlled case-series (SCCS) design. Finally, we estimated excess CVD risk associated with LRTI for a broad range of CVD types.

METHODS

Study design

This retrospective cohort study based on routine electronic health records (EHR) data was designed in three consecutive steps. First, we summarised 90-day CVD incidence (MACCE, VTE, and nAF) in the general population (population cohort) and in patients following an LRTI diagnosis (LRTI cohort) (step 1). Next, we performed a nested SCCS analysis to estimate incidence rate ratios (IRR) of these three CVD types within 90 days following LRTI compared to a control (non-LRTI) period (step 2). By design, SCCS studies rule-out all time-invariant confounding through within-patient comparisons, thus allowing us to estimate robust IRRs.[10] Finally, based on the results from step 1 and 2, we calculated excess CVD risk associated with LRTI (step 3).

While reporting, we adhered to the STROBE[11] and ROBINSE[12] checklists.

Participants

Study cohorts were derived from the Julius General Practitioners' Network (JGPN) and have been described in detail elsewhere.[13] In short, JGPN holds longitudinal routine primary care data on approximately 450,000 patients annually from approximately 70 participating general practices in the Netherlands, representative of the Dutch population.[14] It contains structured data on patient demographics, medical history and diagnoses (using International Classification of Primary Care (ICPC) codes), consultations (including clinical notes), and medication use. JGPN data was linked on a patient level to hospitalisation (Dutch Hospital Data (DHD)) and mortality data (National Mortality Registry).

From JGPN, we included all patients aged ≥40 years between 1 January 2016 and 31 December 2019 in the population cohort, irrespective of prior CVD history (Figure 1A). The years of the COVID-19 pandemic (2020–2023) were deliberately not included in the study period, since LRTI aetiology was dominated by SARS-CoV-2 and LRTI incidence and severity may not resemble those of the current post-pandemic period.[15] We then selected all patients with a GP-diagnosis of LRTI during the study period to derive the nested LRTI cohort. For nAF analyses specifically, we selected LRTI patients without a prior AF diagnosis. Finally, from the LRTI cohort, we included all patients with at least one MACCE, VTE, or nAF diagnosis during the study period in the respective nested SCCS cohorts.

Procedures

For the population cohort, baseline characteristics were collected on demographics (age, sex), relevant comorbidities (diabetes, hypertension, rheumatoid arthritis, prior MI, stroke, heart failure, and AF)[16], CVD outcomes during follow-up, and date of death (where applicable). For the LRTI cohort, additional data were collected on date(s) of LRTI diagnosis and influenza vaccinations during study conduct. Based on available LRTI ICPC codes, an LRTI episode was defined as GP-diagnosis of either pneumonia (R81) or acute bronchitis (R78). An LRTI-related consultation after a period of 28 days without LRTI consultations was considered a new LRTI episode. Where applicable, multiple LRTI episodes of individual patients were included.

To calculate 90-day incidence rates (IR) of CVD, we collected the follow-up time in days, defined as the time from start of follow-up or JGPN registration (whatever came last) to end of registration, end of follow-up, or death (whatever came first).

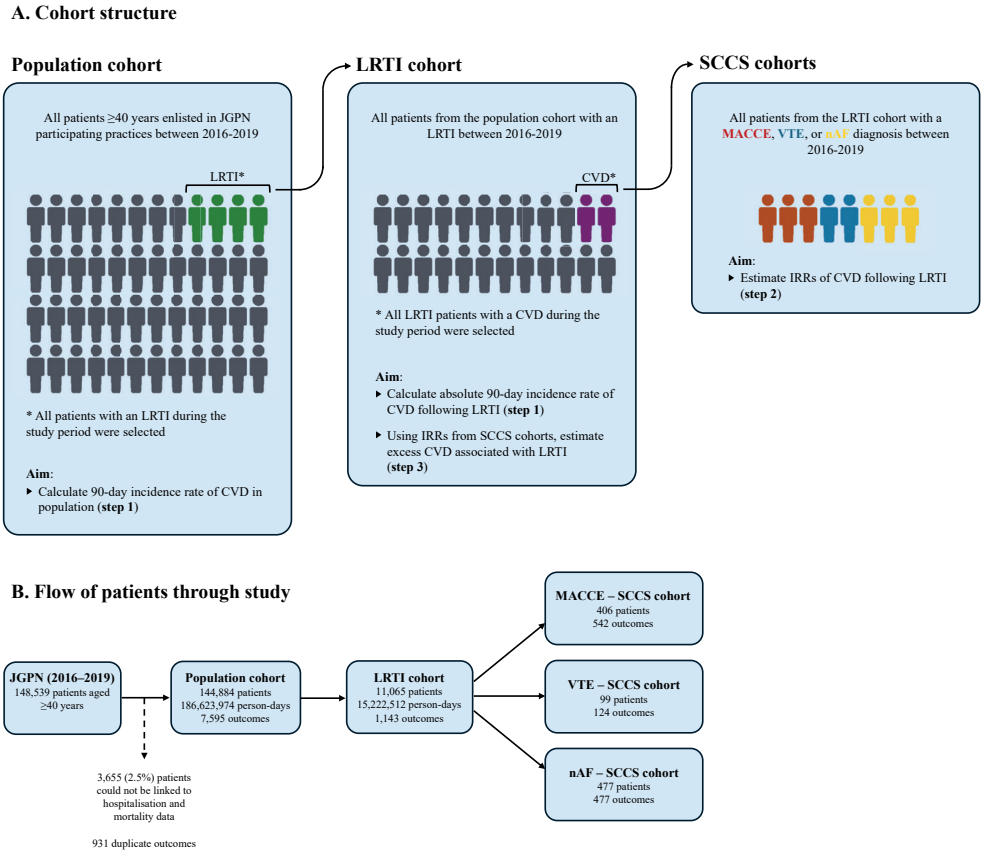


Figure 1. (a) Nested cohort construction and aim of respective cohorts, and **(b)** flow of patients through study.

JGPN, Julius General Practitioners' Network; LRTI, lower respiratory tract infection; CVD, cardiovascular disease; IRR, incidence rate ratio; nAF, new-onset atrial fibrillation; MACCE, major adverse cardiac and cerebrovascular event; VTE, venous thromboembolism; SCCS, self-controlled case-series.

We included three CVD types as outcomes of interest: (1) MACCE (acute coronary syndrome (i.e., MI and unstable angina), stroke, and transient ischaemic attack (TIA)), (2) VTE (pulmonary embolism (PE) and deep venous thrombosis (DVT)), and (3) nAF (i.e., first AF diagnosis). Outcomes were collected from JGPN (nAF, TIA and DVT only – since these may be diagnosed and treated in general practice) and DHD (for all CVD, based on ICD-10 codes of primary and secondary hospital diagnoses). Within JGPN, first ever ICPC-registrations of nAF, TIA, or DVT without prior registration were considered as outcomes of interest. For patients with a history of TIA or DVT and in whom a new TIA or DVT consultation was registered during the study period, clinical notes were assessed by MHR and TNP independently, to determine whether registration was attributable to a new event or to a follow-up visit after a prior event. A new event was defined as a consultation for new signs or symptoms registered by the GP as TIA or DVT.

We collected data, including date of diagnosis, on all CVD that occurred during the study period for all patients of both the population and LRTI cohorts. Consecutive registration within 7 days of identical diagnoses in JGPN and DHD data were considered duplicates. Code-based definitions of all parameters and outcomes are presented in the Supplementary materials (Table S1).

Statistical analysis

For SCCS sample size calculations, expected IRRs of CVD within 90 days following LRTI were based on prior research – approximately 1.8 for MACCE[2], 2.5 for VTE[5], and 2.0 for nAF[4,17] – with an estimated mean follow-up per participant of 3.5 years, 80% power and alpha at 0.05. This resulted in a required sample size of 230 LRTI patients with MACCE, 82 with VTE, and 158 with nAF during the complete study period.

For step 1, IRs for all three CVD types were calculated in both the population cohort (overall 90-day IR) and LRTI cohort (90-day IR following LRTI diagnosis) with 95% confidence intervals (CI). IRs were stratified according to age, sex, LRTI diagnosis and comorbidities at baseline.

For step 2, SCCS analyses were conducted for all three CVD types. MACCE and VTE were considered (potentially) recurrent events, but nAF is non-recurrent by definition. However, although developed for recurrent events, SCCS methods also accommodate relatively rare non-recurrent events – such as nAF – provided that cumulative incidence is below 10%.[18]

For the SCCS analyses, all patients from the LRTI cohort with a MACCE, VTE, or nAF diagnosis during the study period were included in the respective analysis, and follow-up time per individual patient was labelled as either risk or control period (Figure 2). As primary analysis, post-exposure risk periods for which IRRs were determined were 0–7 days, 8–30 days, and 31–90 days following LRTI diagnosis. In addition, 0–90 day IRRs were estimated to calculate excess CVD within 90 days following LRTI. A separate pre-exposure period (14–1 days prior to LRTI diagnosis) was included to account for potential event-dependent exposure bias, i.e., CVD temporarily increases risk of LRTI.[10] We applied a wash-out period of 90 days following the risk period to minimize carry-over effects between risk periods and control periods. All remaining person-time during the study period served as control period. We excluded person-time serving as control period during the first 90 days of the study period to avoid misclassification, since we had no information on LRTI diagnoses prior to the start of the study.

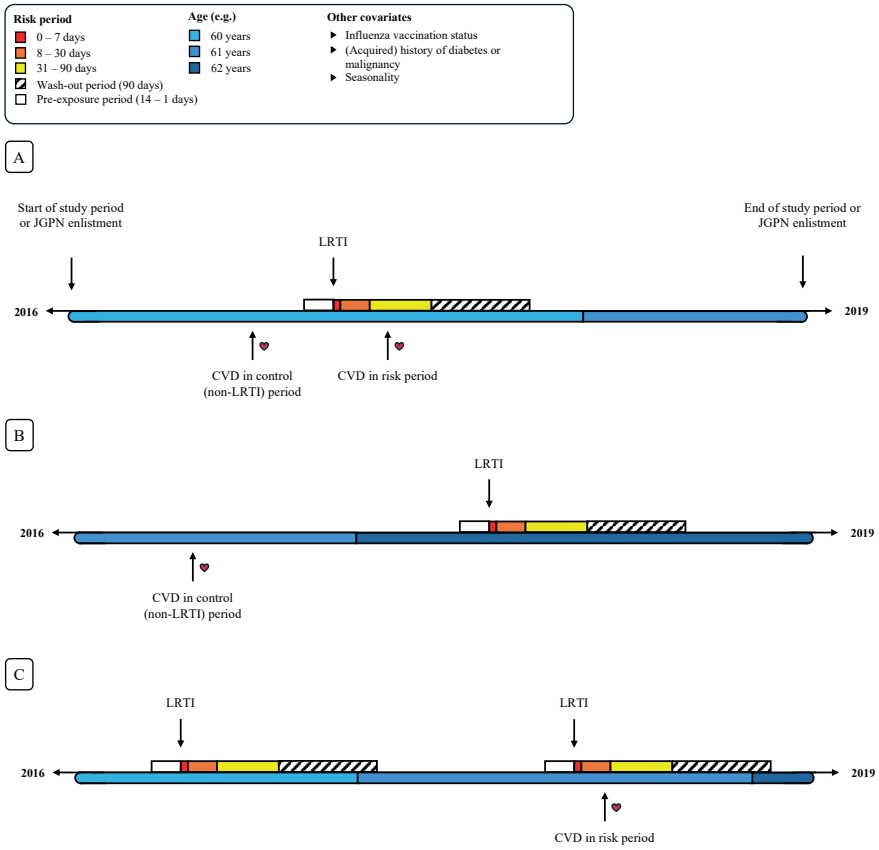


Figure 2. Self-controlled case-series analysis structure and illustrative examples of timeline for individual patients

Three illustrative examples of an individual patient's timeline: **A**. one LRTI episode and two CVD outcomes (in control and risk period), **B**. one LRTI episode and one CVD outcome (in control period), **C**. two LRTI episodes and one CVD outcome (in risk period).

DM, diabetes mellitus; JGPN, Julius General Practitioners' Network; LRTI, lower respiratory tract infection; CVD, cardiovascular disease.

For each risk period, IRRs were estimated by fitting a conditional Poisson model using full Poisson likelihood, with profile-likelihood estimation of 95% CIs. We adjusted for time-varying confounding by adding age (continuous), seasonality (3-month periods), comorbidities diagnosed during study period (diabetes and malignancies), and influenza vaccination status (no indication, vaccinated, or unvaccinated) to the model.

To explore the robustness of our findings, several sensitivity analyses were performed. First, to account for potential diagnostic misclassification of LRTI (i.e., CVD-related symptoms initially misdiagnosed as LRTI, while progression of symptoms results in a subsequent CVD diagnosis) we applied varying definitions of the risk period directly fol-

lowing LRTI diagnosis for the primary analysis. In addition to the 0–7 days risk period, we excluded person-time and outcomes (1) on the day of LRTI diagnosis (adjusted risk period: 1–7 days), (2) on the first two days following LRTI diagnosis (adjusted risk period: 2–7 days), and (3) on the first three days following LRTI diagnosis (adjusted risk period: 3–7 days). Second, a key assumption of SCCS is that outcome events do not affect subsequent exposures. Since CVD may increase the probability of death, follow-up following outcomes may be shorter, thereby decreasing the probability of subsequent exposures. To assess the impact of mortality on our results, we performed two sensitivity analysis in which we (1) excluded all patients who died during the study period[10], and (2) applied the end of the study period as end of follow-up, had death not occurred.[6] Third, prior CVD may increase the probability of a recurrent event, possibly violating assumptions regarding independent recurrences of events. We therefore conducted two sensitivity analyses (1) by including only first CVD diagnoses and (2) by adjusting for CVD recurrence.[18] Finally, we explored the impact of extending the pre-exposure risk period from 14 to 28 days.

For step 3, we calculated the excess CVD cases associated with LRTI based on observed CVD incidences in the control period and estimated IRRs for the risk period obtained from the SCCS analyses (Supplementary file S2). Since the SCCS cohorts were nested in the LRTI cohort, we could subsequently calculate the excess CVD cases per 1,000 LRTI episodes.

All analyses were performed in R version 4.4.2, using the *SCCS* (sample size calculations), *epitools* (incidence calculations), and *gnm* (conditional Poisson modelling) packages.

RESULTS

Between 2016–2019, 148,539 participants aged ≥40 years were enlisted in JGPN (Figure 1B). Of these, 144,884 (97.5%) participants could be linked to hospitalisation data and were included in the population cohort, contributing a total follow-up of 186,623,974 person-days (Table 1). Among participants of the population cohort, 11,065 were diagnosed with one or more LRTI during the study period and thus included in the LRTI cohort, contributing a total of 16,497 LRTI episodes of which 9,384 (56.9%) were registered as pneumonia. The LRTI cohort had a median age of 62 years (interquartile range (IQR) 52–72) and 6,206 (56.1%) patients were female. Among LRTI patients, 10,788 had no prior history of AF (Table S3).

Table 1. Baseline characteristics of population cohort and LRTI cohort at start of follow-up

	Population cohort	LRTI cohort
Total, n	144 884	11 065
Total follow-up time (days)	186 623 974	15 222 512
Median age at inclusion, years (IQR)	52 (44 – 63)	62 (52 – 72)
Female sex, n (%)	74 596 (51.5)	6 206 (56.1)
Mean follow-up time, years (SD)	3.5 (1.0)	3.8 (0.8)
LRTI episodes during study period, n (%)		
0	133 819 (92.4)	-
1	7 808 (5.4)	7 808 (70.6)
2	2 054 (1.4)	2 054 (18.6)
3	693 (0.5)	693 (6.3)
4	299 (0.2)	299 (2.7)
≥5	211 (0.1)	211 (1.9)
Medical history, n (%)		
Hypertension	42 699 (29.5)	4 418 (39.9)
Atrial fibrillation	3 349 (2.3)	277 (2.5)
Diabetes	3 835 (2.6)	579 (5.2)
Myocardial infarction	4 071 (2.8)	554 (5.0)
Heart failure	2 512 (1.7)	477 (4.3)
Stroke	3 534 (2.4)	425 (3.8)
Rheumatoid arthritis	813 (0.6)	150 (1.4)
Per unique LRTI episode		
Number of LRTI episodes	-	16 497
LRTI diagnosis, (%)		
Acute bronchitis	-	7 113 (43.1)
Pneumonia	-	9 384 (56.9)

LRTI, lower respiratory tract infection; IQR, interquartile range; SD, standard deviation.

Among participants of the population cohort, 7,595 outcomes were registered during the study period: 3,800 MACCE, 835 VTE, and 2,960 nAF diagnoses (Table S4). The observed 90-day IR per 1,000 participants was 1.8 (95% CI 1.8–1.9) for MACCE, 0.4 (0.4–0.4) for VTE, and 1.5 (1.4–1.5) for nAF (Table S5).

Among patients with an LRTI episode during the study period, 1,143 outcomes were registered: 542 MACCE, 124 VTE, and 477 nAF diagnoses (Table S6). Of these, 69 MACCE, 44 VTE, and 96 nAF diagnoses were registered within 90 days following LRTI diagnosis, resulting in IRs of 4.2 (3.3–5.3) for MACCE, 2.7 (1.9–3.6) for VTE, and 6.6 (5.4–8.1) for nAF within 90 days following LRTI per 1,000 LRTI patients (Table 2). The median time to outcome in days from LRTI diagnosis was 45 (IQR 6–66) for MACCE, 9.5 (2–42.5) for VTE, and 10.5 (1–29) for nAF (Figure S7).

Incidence of all CVD types within 90 days following LRTI diagnosis were higher in patients diagnosed with pneumonia: 4.7 (95% CI 3.4–6.3) for MACCE, 4.4 (3.1–5.9) for VTE, and 8.4 (6.6–10.7) for nAF (Table S5). In addition, 90-day MACCE incidence following LRTI diagnosis was particularly high in male patients (6.1 (4.4–8.2)), patients with heart failure (8.9 (3.8–17.5)), diabetes (7.5 (3.0–15.5)), and hypertension (6.7 (4.9–9.0)). 90-day VTE incidence following LRTI was most pronounced in the lower age categories, i.e., between 40–49 years (3.0 (1.3–6.0)) and 60–69 years (4.0 (2.3–6.4)). nAF incidence within 90 days following LRTI diagnosis increased with age and was especially high in those with heart failure (26.2 (14.7–43.2)), diabetes (10.9 (4.7–21.5)) and hypertension (9.7 (7.3–12.6)).

Table 2. 90-day incidence rate following LRTI diagnoses per cardiovascular disease type

	90-day incidence rate following LRTI diagnosis per 1 000 LRTI patients (95% CI)		
	MACCE	**VTE**	**New-onset AF[1]**
Overall	4.2 (3.3 – 5.3)	2.7 (1.9 – 3.6)	6.6 (5.4 – 8.1)
LRTI diagnosis			
Pneumonia	4.7 (3.4 – 6.3)	4.4 (3.1 – 5.9)	8.4 (6.6 – 10.7)
Age			
40 – 49	1.1 (0.2 – 3.3)	3.0 (1.3 – 6.0)	0.4 (0.0 – 2.2)
50 – 59	2.4 (1.0 – 4.8)	1.8 (0.7 – 4.0)	1.6 (0.5 – 3.8)
60 – 69	5.5 (3.4 – 8.3)	4.0 (2.3 – 6.4)	7.7 (5.1 – 11.1)
70 – 79	4.8 (2.9 – 7.5)	2.8 (1.4 – 5.0)	10.6 (7.4 – 14.7)
≥80	6.6 (3.9 – 10.6)	1.2 (0.2 – 3.4)	14.0 (9.2 – 20.4)
Sex			
Female	2.7 (1.7 – 4.0)	2.2 (1.3 – 3.3)	5.3 (3.9 – 7.1)
Male	6.1 (4.4 – 8.2)	3.3 (2.1 – 5.0)	8.4 (6.2 – 11.0)
Medical history			
Diabetes	7.5 (3.0 – 15.5)	2.1 (0.3 – 7.8)	10.9 (4.7 – 21.5)
Hypertension	6.7 (4.9 – 9.0)	1.8 (0.9 – 3.1)	9.7 (7.3 – 12.6)
Rheumatoid arthritis	3.9 (0.1 – 21.7)	3.9 (0.1 – 21.7)	9.3 (1.1 – 33.5)

Table 2. 90-day incidence rate following LRTI diagnoses per cardiovascular disease type

	90-day incidence rate following LRTI diagnosis per 1 000 LRTI patients (95% CI)		
	MACCE	**VTE**	**New-onset AF[1]**
Myocardial infarction	19.2 (11.2 – 30.8)	2.3 (0.3 – 8.2	15.7 (7.8 – 28.0)
Stroke	25.7 (15.2 – 40.6)	2.9 (0.3 – 10.3)	22.7 (12.1 – 38.9)
Heart failure	8.9 (3.8 – 17.5)	0.0 (0.0 – 4.1)	26.2 (14.7 – 43.2)
Atrial fibrillation	4.4 (1.6 – 9.6)	0.7 (0.0 – 4.1)	-

[1] Incidence rates of new-onset atrial fibrillation were calculated among LRTI patients without a history of prior atrial fibrillation.

CI, confidence interval; MACCE, major adverse cardiac and cerebrovascular event; VTE, venous thromboembolism; AF, atrial fibrillation; LRTI, lower respiratory tract infection.

From the LRTI cohort, 406, 99, and 477 individual patients with at least one MACCE, VTE, and nAF diagnosis were included in the respective SCCS analyses, of whom 18 (4.4%), 5 (5.0%), and 24 (5.0%) died during the study period, respectively (Table S8). For MACCE, IRRs were 3.7 (95% CI 2.2–5.8) within 0–7 days, 0.6 (0.3–1.2) within 8–30 days, and 1.3 (0.9–1.8) within 31–90 days from LRTI diagnosis (Table 3), with a 0–90 days IRR of 1.3 (1.0–1.8). For VTE, IRRs were 30.3 (17.0–52.0) within 0–7 days, 8.5 (4.5–15.2) within 8–30 days, and 2.4 (1.1–4.6) within 31–90 days from LRTI diagnosis, with a 0–90 days IRR of 6.8 (4.4–10.4). For nAF, IRRs were 14.6 (10.4–20.0) within 0–7 days, 3.3 (2.2–4.8) within 8–31 days, and 1.0 (0.6–1.6) within 31–90 days following LRTI diagnosis, with a 0–90 days IRR of 2.9 (2.3–3.7).

Upon excluding days directly following LRTI diagnosis from the risk period, IRRs decreased to 2.2 (0.9–4.3) for MACCE, 14.2 (5.4–31.4) for VTE, and 6.7 (3.6–11.4) for nAF within 3–7 days following LRTI. Results obtained by additional sensitivity analyses did not substantially differ from the primary analysis (Table S9).

Depending on risk period definition in the primary and sensitivity analyses, i.e., ranging from 0–90 days to 3–90 days, the number of excess CVD cases within 90 days from LRTI diagnosis per 1,000 LRTI episodes ranged from 1.1 (95% CI 0.1–2.4) to 0.5 (-0.5–1.6) for MACCE, from 3.0 (1.8–4.9) to 2.1 (1.1–3.6) for VTE, and from 4.4 (2.9–6.2) to 2.3 (1.1–3.7) for nAF, respectively (Table 4, Table S10).

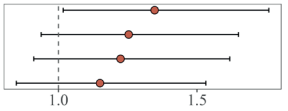
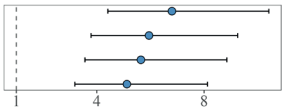
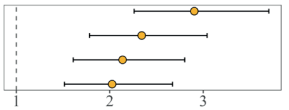
Table 3. Estimated incidence rate ratio per acute cardiovascular disease type and per risk period following LRTI diagnosis

Risk period	**Incidence rate ratio (95% CI)**		
Days following LRTI diagnosis	**MACCE**	**VTE**	**New-onset AF**
0–7 days	3.7 (2.2 – 5.8)	30.3 (17.0 – 52.0)	14.6 (10.4 – 20.0)
1–7 days	3.1 (1.7 – 5.2)	21.1 (10.5 – 39.4)	9.2 (5.9 – 13.7)
2–7 days	2.9 (1.5 – 5.1)	18.1 (8.2 – 36.0)	7.4 (4.3 – 11.7)
3–7 days	2.2 (0.9 – 4.3)	14.2 (5.4 – 31.4)	6.7 (3.6 – 11.4)
8–30 days	0.6 (0.3 – 1.2)	8.5 (4.5 – 15.2)	3.3 (2.2 – 4.8)
31–90 days	1.3 (0.9 – 1.8)	2.4 (1.1 – 4.6)	1.0 (0.6 – 1.6)

In addition to the primary analysis (**bold**), sensitivity analyses were conducted that excluded follow-up time and outcomes on the day(s) of, and directly following, LRTI diagnosis to explore the impact of potential diagnostic misclassification.

CI, confidence interval; LRTI, lower respiratory tract infection; MACCE, major adverse cardiac and cerebrovascular event; VTE, venous thromboembolism; AF, atrial fibrillation.

Table 4. Estimated excess acute cardiovascular diseases associated with LRTI within 90 days following LRTI diagnosis

Risk period	Estimated IRR (95% CI)	Estimated IRR (95% CI) Forest plot	Estimated excess CVD in this study (95% CI)	Estimated excess CVD per 1,000 LRTI patients (95% CI)
Days following LRTI diagnosis		**MACCE**		
0 – 90 days	1.3 (1.0 – 1.8)		18.0 (0.9 – 39.3)	1.1 (0.1 – 2.4)
1 – 90 days	1.3 (0.9 – 1.6)		13.0 (-3.2 – 33.2)	0.8 (-0.2 – 2.0)
2 – 90 days	1.2 (0.9 – 1.6)		11.4 (-4.5 – 31.3)	0.7 (-0.3 – 1.9)
3 – 90 days	1.1 (0.8 – 1.5)		7.5 (-7.6 – 26.7)	0.5 (-0.5 – 1.6)
		VTE		
0 – 90 days	6.8 (4.4 – 10.4)		49.6 (29.2 – 80.5)	3.0 (1.8 – 4.9)
1 – 90 days	5.9 (3.8 – 9.3)		41.9 (23.6 – 69.8)	2.5 (1.4 – 4.2)
2 – 90 days	5.6 (3.6 – 8.8)		38.9 (21.4 – 65.7)	2.4 (1.3 – 4.0)
3 – 90 days	5.1 (3.2 – 8.1)		34.1 (18.1 – 59.0)	2.1 (1.1 – 3.6)
		New-onset AF		
0 – 90 days	2.9 (2.3 – 3.7)		70.8 (46.8 – 100.4)	4.4 (2.9 – 6.2)
1 – 90 days	2.3 (1.8 – 3.0)		49.3 (28.8 – 75.0)	3.1 (1.8 – 4.7)
2 – 90 days	2.1 (1.6 – 2.8)		41.3 (22.1 – 65.5)	2.6 (1.4 – 4.1)
3 – 90 days	2.0 (1.5 – 2.7)		36.8 (18.5 – 60.1)	2.3 (1.1 – 3.7)

In addition to the primary analysis (**bold**), sensitivity analyses were conducted that excluded follow-up time and outcomes on the day(s) of, and directly following, LRTI diagnosis to explore the impact of potential diagnostic misclassification. Underlying calculations are presented in the Supplementary materials (Supplementary File S2 and Table S10).

IRR, incidence rate ratio; CI, confidence interval; IR, incidence rate; LRTI, lower respiratory tract infection; MACCE, major adverse cardiac and cerebrovascular event; VTE, venous thromboembolism; AF, atrial fibrillation.

DISCUSSION

This large population-based study with nested self-controlled case-series allowed for robust estimations of excess CVD risk within 90 days following LRTI diagnosis across a broad range of CVD types. For every 1,000 patients with a GP-diagnosis of LRTI, 4.2 and 2.7 are diagnosed with MACCE and VTE within 90 days, respectively, and per 1,000 LRTI patients without prior AF diagnosis 6.6 are diagnosed with nAF. Of these events, and depending on risk period definition, an estimated 0.5 to 1.1 MACCE, 2.1 to 3.0 VTE, and 2.3 to 4.4 nAF diagnoses are associated with LRTI. CVD risk was consistently highest in the first week following LRTI diagnosis and remained elevated during the first month for VTE and nAF. For MACCE and nAF, observed 90-day incidence rates following LRTI diagnosis increased with age and were particularly high in patients with heart failure, diabetes and hypertension. Contrary, for VTE the 90-day incidence rate following LRTI diagnosis was most pronounced among younger patients. CVD incidence following LRTI diagnosis was generally higher among patients diagnosed with pneumonia compared to acute bronchitis.

In sensitivity analyses, we explored the impact of diagnostic misclassification potentially introduced by using routine EHR data. IRRs within 0–7 days from LRTI diagnosis were particularly inflated by PE and nAF diagnoses shortly following LRTI diagnosis, an issue also observed in previous studies.[3,5] Diagnostic misclassification is likely most relevant for PE, with diagnostic delay being a well-known issue.[19] Whether CVD-related symptoms were initially misdiagnosed as LRTI, or CVD did indeed occur directly after LRTI diagnosis in these cases cannot be accurately determined retrospectively. Our results, however, provide robust evidence that CVD risk remains substantially increased even when accounting for this potential misclassification.

The association between RTI and CVD has been established by previous studies, which reported similar findings. A systematic review and meta-analysis on causal effects of RTIs on CVD found a 2.9-fold increase of MI risk and 2.6-fold increase of stroke risk within 30 days from RTI diagnosis.[4] Risk of DVT within 30 days increased 1.9 to 2.6-fold, and risk of AF 1.2-fold. In addition, a population-based case-control study reported an 8.1 higher odds of PE for one year following pneumonia.[20] The magnitude of these effects decreased with increasing follow-up duration, which aligns with our findings that the vast majority of CVD occur in the first week following LRTI diagnosis. Effect sizes also differed across RTI type, i.e., relative risks were generally lower for upper RTIs than LRTIs. Our finding that CVD incidence is generally higher among patients diagnosed with pneumonia supports the hypothesis that CVD risk may be related to RTI severity. In line with this, the 28-day incidence of a composite outcome including MI, stroke, TIA, and cardiovascular death in a recent prediction modelling study among RTI patients was lower than in our pneumonia population (0.3% versus 0.5%), with RTI type – notably LRTI or pneumonia – as important

predictor.[21] Other predictors included in the model also align with the subset of patients with higher CVD risk identified in our study: increasing age, history of heart failure and diabetes. In addition, our finding that CVD incidence is especially high among LRTI patients with diabetes or hypertension is consistent with results of another study.[22]

In LRTI patients, CVD are suggested to be triggered by physiological adaptive changes, such as increased cardiac demand, systemic inflammation, and hypercoagulation.[23] Studies on potential preventative interventions to mitigate CVD risk therefore targeted these pathways, but results are inconsistent. An observational study among community-acquired pneumonia patients found that aspirin use substantially reduced risk of MI and stroke.[24] Contrary, a trial on community patients with COVID-19 found no effect of aspirin to prevent MI, stroke, or PE.[25]

To the best of our knowledge, this study is the first to provide data on absolute excess risk of CVD associated with LRTI across a broad range of CVD types and stratified according to patient characteristics. Strengths include the comprehensive population-based coverage of CVD following LRTI, both in terms of methods (mapping of incidence, estimating causal effects and absolute excess risk) and CVD endpoints. The nested cohort structure allowed us to calculate excess CVD risk where previous studies could not.[26]

Some limitations, however, deserve consideration. First, since we estimated causal effects with observational data, these may suffer from residual confounding. Although SCCS analyses are designed to rule-out time-invariant confounding and several important time-varying confounders were adjusted for, some residual time-varying confounding may exist. In particular, data on smoking status and medication use (e.g., anticoagulants, statins, non-steroidal anti-inflammatory drugs) were lacking in our study. Whether smoking status should be considered as time-varying confounder during our four-year study period could, however, be debated, given the low cessation success rates.[27] Moreover, by adjusting for history of MACCE and VTE in a sensitivity analysis, confounding by anticoagulant use has been partly accounted for. Second, calculations of excess CVD were made under the assumption that these events would not have occurred in the counterfactual, that is, would LRTI not have occurred. Third, our definition of LRTI did not include exacerbations of chronic obstructive pulmonary disease (COPD), since these could not be distinguished from routine COPD follow-up consultations in our data. This may have led to an underestimation of the observed causal inferences, since COPD exacerbations, often triggered by RTIs, are now included in control period whereas they have been linked to increased CVD risk.[28] We may have even further underestimated the burden of CVD following LRTI since we only included medically attended LRTIs and some CVD may have remained clinically undetected.[29]

Albeit the absolute excess CVD risk associated with LRTI may seem modest, approximately 1 out of 30 adults are diagnosed with LRTI annually of whom an estimated 0.5–0.9% will develop LRTI-associated CVD. Moreover, CVD incidence following LRTI varies across patient characteristics which open venues for tailored preventative interventions. The finding that CVD risk increases with RTI severity warrants policy aimed at increasing seasonal vaccine uptake. Communication strategies could, for example, increase population awareness of the potential benefit of vaccinations in reducing CVD risk among those at highest CVD risk. For these patients, tailored interventions – e.g., low-cost and low-risk interventions to reduce LRTI severity[30], antithrombotic treatment to prevent thromboembolic events or targeted case-finding strategies for nAF – deserve further investigation, as well as strategies aimed at increasing clinicians' awareness of CVD risk associated with LRTI. Future interventions should be developed in co-creation with all relevant stakeholders, including patients, who need to carefully weigh pros (benefits) and cons (risk of side-effects, costs) of any strategies prior to large-scale evaluation of clinical and cost-effectiveness.

REFERENCES

1. Hak E, Rovers MM, Kuyvenhoven MM, Schellevis FG, Verheij TJM. Incidence of GP-diagnosed respiratory tract infections according to age, gender and high-risk co-morbidity: The second Dutch national survey of general practice. *Fam Pract*. 2006;**23**(3):291–4.
2. Smeeth L, Thomas SL, Hall AJ, et al. Risk of myocardial infarction and stroke after acute infection or vaccination. *N Engl J Med*. 2004;**351**(25):2611–8.
3. Smeeth L, Cook C, Thomas S, et al. Risk of deep vein thrombosis and pulmonary embolism after acute infection in a community setting. *Lancet*. 2006;**367**:1075–9.
4. La Roi-Teeuw HM, Van Smeden M, Bos M, et al. Estimated causal effects of common respiratory infections on cardiovascular risk: a meta-analysis. *Open Heart*. 2023;**10**(2):e002501.
5. Clayton TC, Gaskin M, Meade TW. Recent respiratory infection and risk of venous thromboembolism: case-control study through a general practice database. *Int J Epidemiol*. 2011;**40**(3):819–27.
6. de Boer AR, Riezebos-Brilman A, van Hout D, et al. Influenza infection and acute myocardial infarction. *NEJM Evid*. 2024;**3**(7):EVIDoa2300361.
7. Katsoularis I, Fonseca-Rodríguez O, Farrington P, Lindmark K, Fors Connolly AM. Risk of acute myocardial infarction and ischaemic stroke following COVID-19 in Sweden: a self-controlled case series and matched cohort study. *Lancet*. 2021;**398**(10300):599–607.
8. Katsoularis I, Fonseca-Rodríguez O, Farrington P, et al. Risks of deep vein thrombosis, pulmonary embolism, and bleeding after covid-19: nationwide self-controlled cases series and matched cohort study. *BMJ*. 2022;**377**:e069590.
9. Warren-Gash C, Blackburn R, Whitaker H, McMenamin J, Hayward AC. Laboratory-confirmed respiratory infections as triggers for acute myocardial infarction and stroke: a self-controlled case series analysis of national linked datasets from Scotland. *Eur Respir J*. 2018;**51**(3):1701794.
10. Petersen I, Douglas I, Whitaker H. Self controlled case series methods: an alternative to standard epidemiological study designs. *BMJ*. 2016;**354**:i4515.
11. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 2008;**61**(4):344–9.
12. Higgins JPT, Morgan RL, Rooney AA, et al. A tool to assess risk of bias in non-randomized follow-up studies of exposure effects (ROBINS-E). *Environ Int*. 2024;**186**:108602.
13. Rijk MH, Platteel TN, Geersing GJ, et al. 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. *Diagn Progn Res*. 2023;**7**(1):23.
14. Smeets H, Kortekaas M, Rutten F, 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.
15. 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.
16. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ*. 2008;**336**(7659):1475–82.
17. Corica B, Tartaglia F, Oliva A, et al. Prevalence of new-onset atrial fibrillation in hospitalized patients with community-acquired pneumonia: a systematic review and meta-analysis. *Intern Emerg Med*. 2023;**18**(1):127–35.
18. Whitaker HJ, Steer CD, Farrington CP. Self-controlled case series studies: Just how rare does a rare non-recurrent outcome need to be? *Biom J*. 2018;**60**(6):1110–1120.

19. van Maanen R, Trinks-Roerdink EM, Rutten FH, Geersing GJ. A systematic review and meta-analysis of diagnostic delay in pulmonary embolism. *Eur J Gen Pract*. 2022;**28**(1):165–172.
20. Ribeiro DD, Lijfering WM, Van Hylckama Vlieg A, Rosendaal FR, Cannegieter SC. Pneumonia and risk of venous thrombosis: results from the MEGA study. *J Thromb Haemos*. 2012;**10:**1179–82.
21. Lee JJ, Koshiaris C, Wright-Drakesmith C, et al. Development and external validation of a risk prediction score (DASHI) for cardiovascular events following acute respiratory infections: derivation and validation retrospective cohort study. *eClinicalMedicine*. 2025;**84**:103273.
22. Davidson JA, Banerjee A, Smeeth L, 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.
23. Stark K, Massberg S. Interplay between inflammation and thrombosis in cardiovascular pathology. *Nat Rev Cardiol*. 2021:**18**;666–82.
24. Hamilton F, Arnold D, Henley W, Payne RA. Aspirin reduces cardiovascular events in patients with pneumonia: a prior event rate ratio analysis in a large primary care database. *Eur Respir J*. 2021;**57**(2):2002795.
25. Eikelboom JW, Jolly SS, Belley-Cote EP, et al. Colchicine and aspirin in community patients with COVID-19 (ACT): an open-label, factorial, randomised, controlled trial. *Lancet Respir Med*. 2022;**10**(12):1160–8.
26. Neumann V, O'Connor RJ, Bhakta B, Tennant A. Correspondence: DVT and pulmonary embolism after acute infection. *Lancet*. 2006 Jul 15;**368**(9531):201.
27. Rafful C, García-Rodríguez O, Wang S, et al. Predictors of quit attempts and successful quit attempts in a nationally representative sample of smokers. *Addict Behav*. 2013;**38**(4):1920–3.
28. Yang HM, Ryu MH, Carey VJ, et al. Chronic obstructive pulmonary disease exacerbations increase the risk of subsequent cardiovascular events: a longitudinal analysis of the COPDGene study. *J Am Heart Assoc*. 2024;**13**(11):e0338.
29. Pride YB, Piccirillo BJ, Gibson CM. Prevalence, consequences, and implications for clinical trials of unrecognized myocardial infarction. *Am J Cardiol*. 2013;**111**(6):914–8.
30. Little P, Vennik J, Rumsby K, et al. Nasal sprays and behavioural interventions compared with usual care for acute respiratory illness in primary care: a randomised, controlled, open-label, parallel-group trial. *Lancet Respir Med*. 2024;**12**(8):619–632.

SUPPLEMENTARY MATERIALS FOR CHAPTER 8

Table S1. Details on study parameters and (code-based) definitions

Parameter	Definition or underlying registration(s)	Codes, if applicable	Coding system, if applicable			Data source
			ICPC	ICD-10	ATC	
Demographics						
Age	Age in years at start follow-up	-				JGPN
Sex	Biological sex	-				JGPN
Other baseline characteristics						
Atrial fibrillation	Atrial fibrillation	K78	x			JGPN
Diabetes	Diabetes mellitus (type 1 or 2)	T90	x			JGPN
Heart failure	Heart failure	K77	x			JGPN
Hypertension	Hypertension	K85, K86, K87	x			JGPN
Myocardial infarction	Acute myocardial infarction	K75	x			JGPN
Rheumatoid arthritis	Rheumatoid arthritis	L88.01	x			JGPN
Stroke	Stroke	K90	x			JGPN
Exposure						
LRTI	Acute bronchitis Pneumonia	R78 R81	x			JGPN
Covariates of self-controlled case-series analysis						
Age	Continuous	-				JGPN
Seasonality	Autumn (from 21 September to 21 December) Winter (from 21 December to 21 March) Spring (from 21 March to 21 June) Summer (from 21 June to 21 September)	-				JGPN

Table S1. Continued

Parameter	Definition or underlying registration(s)	Codes, if applicable	Coding system, if applicable			Data source
			ICPC	*ICD-10*	*ATC*	
(Acquired) history of diabetes	Registration of diabetes mellitus before or during study period	T90	x			JGPN
(Acquired) history of malignancy	Registration of (non-cutaneous) malignancy before or during study period	B72-75, D74-77, R84-85, U75-77, Y77-78, X75-77, A79, F74.01, H75.01, K72.01, L71.01, N74, T71	x			JGPN
Influenza vaccination indication	To evaluate the presence of an indication for influenza vaccination criteria based on guidance issued by the Dutch College of General Practitioners (NHG) were applied.[1] Patients had an indication if meeting one of the following criteria: - Age ≥60 years - Indication due to medical history (not all relevant medical history was covered in the dataset; selection based on available variables only): - Asthma with inhalation corticosteroid use - Chronic obstructive pulmonary disease - Heart failure - History of acute myocardial infarction or angina pectoris - Cardiomyopathy - Valvular disorders - Atrial fibrillation with oral anticoagulant use - Diabetes - Renal disease - HIV - History of stroke - Dementia	R96 + R03BA/R03AK/R03AL R95 K77 K74, K75 K84.03 K71, K83 K78 + B01AA/B01AE07/B01AF T90 U99.01 B90 K75 P70, P20	 x x x x x x x x x x x x	 	 x x 	JGPN
Influenza vaccination	Influenza vaccination prescription; patient considered vaccinated for 365 days following vaccination date	J07BB02			x	JGPN

Table S1. Continued

Parameter	Definition or underlying registration(s)	Codes, if applicable	Coding system, if applicable			Data source
			ICPC	*ICD-10*	*ATC*	
History of MACCE (sensitivity analysis)	Registration of myocardial infarction, stroke, or TIA before study period	K75, K90, G45	x			JGPN
	Prior acute arterial event during study period (see outcomes)	See outcomes	x	x		JGPN, DHD
History of VTE (sensitivity analysis)	Registration of pulmonary embolism or deep venous thrombosis before study period	K93, K94.01 + manual assessment of clinical notes	x			JGPN
	Prior acute venous event during study period (see outcomes)	See outcomes	x	x		JGPN, DHD
Outcomes						
MACCE	Acute coronary syndrome	I20.0, I21, I22		x		DHD
	Stroke	I61, I63, I64	x	x		DHD
	TIA	G45		x		DHD
		K89 + manual assessment of clinical notes				JGPN
VTE	Pulmonary embolism	I26		x		DHD
	Deep venous thrombosis	I80.1, I80.2, I80.3	x	x		DHD
		K94.01 + manual assessment of clinical notes				JGPN
New-onset atrial fibrillation	Atrial fibrillation, registration without prior registration	I48		x		DHD
		K78	x			JGPN
Mortality						
Death	Date of death	-				CBS

[1] NHG-Praktijkhandleiding Griepvaccinaties. Nederlands Huisartsen Genootschap. [Internet]. Accessed on 6 October 2025. Available from: https://www.nhg.org/praktijkvoering/vaccinatieprogrammas/praktijkhandleiding-griepvaccinatie/

ICPC, International Classification of Primary Care; ICD-10, International Classification of Diseases 10th edition; ATC, Anatomical Therapeutic Chemical; JGPN, Julius General Practitioners' Network; LRTI, Lower respiratory tract infection; DHD, Dutch Hospital Data; NHG, *Nederlands Huisartsen Genootschap*; TIA, Transient ischaemic attack; MACCE, major adverse cardiac and cerebrovascular event; VTE, venous thromboembolism; CBS, *Centraal Bureau voor de Statistiek* (Statistics Netherlands).

Supplementary file S2. Methods for calculating excess CVD cases

To determine the excess risk of acute cardiovascular disease (CVD) associated with lower respiratory tract infections (LRTI), we calculated the excess number of major adverse cardiac and cerebrovascular events (MACCE), venous thromboembolism (VTE), and new-onset atrial fibrillation (nAF) diagnoses within 90 days from LRTI diagnosis using the following procedure.

First, we obtained the control period incidence rate of CVD ($IR_{control}$) from the SCCS cohort:

a. $$IR_{control} = \frac{No.\ of\ CVD\ in\ control\ period}{Persontime\ (days)\ in\ control\ period}$$

Next, we calculated the excess incidence rate of CVD (IR_{excess}) during the 90-day risk period in the SCCS cohort. The IR_{excess} was obtained by multiplying the adjusted incidence rate ratio (IRR) following LRTI by the $IR_{control}$. This formula includes the IRR minus one to arrive at the excess rate relative to the control rate. This step was repeated for the values of the IRR reflecting the 95% confidence interval (CI) :

b. $$IR_{excess} = IR_{control} \times ((IRR \pm 95\%\ CI) - 1)$$

Subsequently, we used the excess CVD incidence rate to calculate the total number of excess CVD cases (EC_{total}) in the SCCS cohort that can be attributed to LRTI:

c. $$EC_{total} = IR_{excess} \times (persontime\ (days)\ in\ risk\ period)$$

In the final step we calculated the mean 90-day excess CVD risk per 1,000 LRTI episodes. All CVD cases in the LRTI cohort were included in the SCCS-cohort and by definition, no excess cases among LRTI patients could have occurred outside the SCCS cohort. Therefore, this total number calculated in step c (EC_{total}) was used as the numerator for the total LRTI cohort. The mean 90-day excess CVD risk per 1,000 LRTI episodes was subsequently calculated using the total number of episodes in the LRTI cohort as the denominator:

d. $$Excess\ CVD\ cases\ per\ 1{,}000\ LRTI\ episodes = \frac{EC_{total} \pm 95\%\ CI}{Total\ no.\ of\ LRTI\ episodes} \times 1{,}000$$

For the excess risk of nAF specifically, the total number of LRTI episodes in the denominator was based on patients without a prior history of AF.

Table S3. Baseline characteristics of population cohort and LRTI cohort without history of atrial fibrillation

	Population cohort Without history of AF	LRTI cohort Without history of AF
Total, n	141 449	10 788
Median age at inclusion, years (IQR)	52 (44 – 63)	62 (51 – 72)
Female sex, n (%)	73 107 (51.7)	6 084 (56.4)
Mean follow-up time, years (SD)	3.5 (1.0)	3.8 (0.6)
Medical history, n (%)		
Hypertension	40 702 (28.8)	4 242 (39.3)
Diabetes	3 599 (2.5)	549 (5.1)
Myocardial infarction	3 789 (2.7)	516 (4.8)
Heart failure	1 913 (1.4)	424 (3.9)
Stroke	3 226 (2.3)	411 (3.8)
Rheumatoid arthritis	782 (0.6)	146 (1.4)
Per unique LRTI episode		
Number of LRTI episodes	-	16 092
LRTI diagnosis, (%)		
Acute bronchitis	-	6 916 (43.0)
Pneumonia	-	9 176 (57.0)

LRTI, lower respiratory tract infection; IQR, interquartile range; SD, standard deviation.

Table S4. Acute cardiovascular diseases in the general population during the study period

Diagnosis	Outcomes during study period	
	Separate diagnoses	Total within category
MACCE		3 800
ACS	1 963	
Stroke	1 176	
TIA	661	
VTE		835
Pulmonary embolism	519	
DVT	316	
New-onset AF	2 960	2 960

MACCE, major adverse cardiac and cerebrovascular event; ACS, acute coronary syndrome; TIA, transient ischaemic attack; VTE, venous thromboembolism; DVT, deep venous thrombosis; AF, atrial fibrillation.

Table S5. 90-day incidence rate per acute cardiovascular disease type in the population and within 90 days following pneumonia diagnosis

	90-day incidence rate per 1 000 persons (95% CI)					
	MACCE		**VTE**		**New-onset AF**	
	Population	*Pneumonia*	*Population*	*Pneumonia*	*Population*	*Pneumonia*
Overall	1.8 (1.8 – 1.9)	4.7 (3.4 – 6.3)	0.4 (0.4 – 0.4)	4.4 (3.1 – 5.9)	1.5 (1.4 – 1.5)	8.4 (6.6 – 10.7)
Age						
40 – 49	0.4 (0.4 – 0.5)	2.1 (0.4 – 6.0)	0.2 (0.1 – 0.2)	5.5 (2.4 – 10.8)	0.2 (0.2 – 0.2)	0.7 (0.0 – 4.0)
50 – 59	1.2 (1.2 – 1.3)	2.9 (1.0 – 6.9)	0.3 (0.2 – 0.3)	3.5 (1.3 – 7.7)	0.5 (0.5 – 0.6)	1.2 (0.2 – 4.5)
60 – 69	2.2 (2.1 – 2.4)	6.2 (3.4 – 10.3)	0.5 (0.4 – 0.6)	6.6 (3.7 – 10.9)	1.9 (1.7 – 2.0)	11.3 (7.2 – 17.0)
70 – 79	3.9 (3.7 – 4.1)	5.6 (3.0 – 9.6)	0.9 (0.8 – 1.0)	3.9 (1.8 – 7.4)	4.1 (3.9 – 4.4)	11.8 (7.5 – 17.7)
≥80	6.5 (6.0 – 7.0)	5.6 (2.6 – 10.6)	0.8 (0.6 – 1.0)	1.9 (0.4 – 5.4)	6.9 (6.4 – 7.5)	16.5 (10.1 – 25.6)
Sex						
Female	1.3 (1.3 – 1.4)	3.0 (1.7 – 5.0)	0.4 (0.3 – 0.4)	3.4 (2.0 – 5.5)	1.3 (1.2 – 1.4)	6.6 (4.4 – 9.4)
Male	2.3 (2.3 – 2.4)	6.6 (4.4 – 9.5)	0.5 (0.4 – 0.5)	5.5 (3.5 – 8.1)	1.6 (1.6 – 1.7)	10.6 (7.6 – 14.4)
Medical history						
Diabetes	4.2 (3.7 – 4.7)	5.6 (1.2 – 16.5)	0.6 (0.4 – 0.8)	3.8 (0.5 – 13.6)	2.9 (2.5 – 3.4)	12.0 (3.9 – 27.9)
Hypertension	3.3 (3.2 – 3.5)	8.5 (5.8 – 12.0)	0.6 (0.6 – 0.7)	3.5 (1.8 – 5.9)	2.8 (2.7 – 3.0)	13.8 (10.0 – 18.6)
Rheumatoid arthritis	3.4 (2.4 – 4.7)	5.7 (0.1 – 32.0)	0.9 (0.4 – 1.7)	5.7 (0.1 – 32.0)	2.7 (1.8 – 3.9)	14.2 (1.7 – 51.4)
Myocardial infarction	17.2 (16.1 – 18.3)	17.0 (7.8 – 32.3)	0.7 (0.5 – 0.9)	11.4 (4.2 – 24.7)	5.4 (4.8 – 6.1)	22.0 (10.1 – 41.7)
Stroke	19.9 (18.6 – 21.3)	44.2 (26.6 – 69.0)	1.0 (0.7 – 1.3)	4.7 (0.6 – 16.8)	5.2 (4.5 – 6.0)	37.2 (19.8 – 63.7)

Table S5. Continued

	90-day incidence rate per 1 000 persons (95% CI)					
	MACCE		**VTE**		**New-onset AF**	
	Population	*Pneumonia*	*Population*	*Pneumonia*	*Population*	*Pneumonia*
Heart failure	6.8 (5.8 – 7.8)	22.9 (12.2 – 39.1)	1.4 (1.0 – 1.9)	1.8 (0.0 – 9.8)	12.6 (11.2 – 14.2)	46.4 (27.0 – 74.2)
Atrial fibrillation	5.0 (4.5 – 5.6)	3.7 (0.8 – 10.8)	0.5 (0.3 – 0.6)	2.5 (0.3 – 8.9)	-	-

CI, confidence interval; MACCE, major adverse cardiac and cerebrovascular event; VTE, venous thromboembolism; AF, atrial fibrillation.

Table S6. Acute cardiovascular diseases in LRTI cohort during study period

Diagnosis	**Study Period**	**Following LRTI**			
		0 – 90 days	*1 – 90 days*	*2 – 90 days*	*3 – 90 days*
MACCE					
ACS	273	30	28	28	26
Stroke	158	27	25	23	21
TIA	111	12	11	10	10
VTE					
Pulmonary embolism	83	37	32	30	27
DVT	41	7	7	7	7
New-onset AF	477	96	77	69	65

LRTI, lower respiratory tract infection; MACCE, major adverse cardiac and cerebrovascular event; ACS, acute coronary syndrome; TIA, transient ischaemic attack; VTE, venous thromboembolism; DVT, deep venous thrombosis; AF, atrial fibrillation.

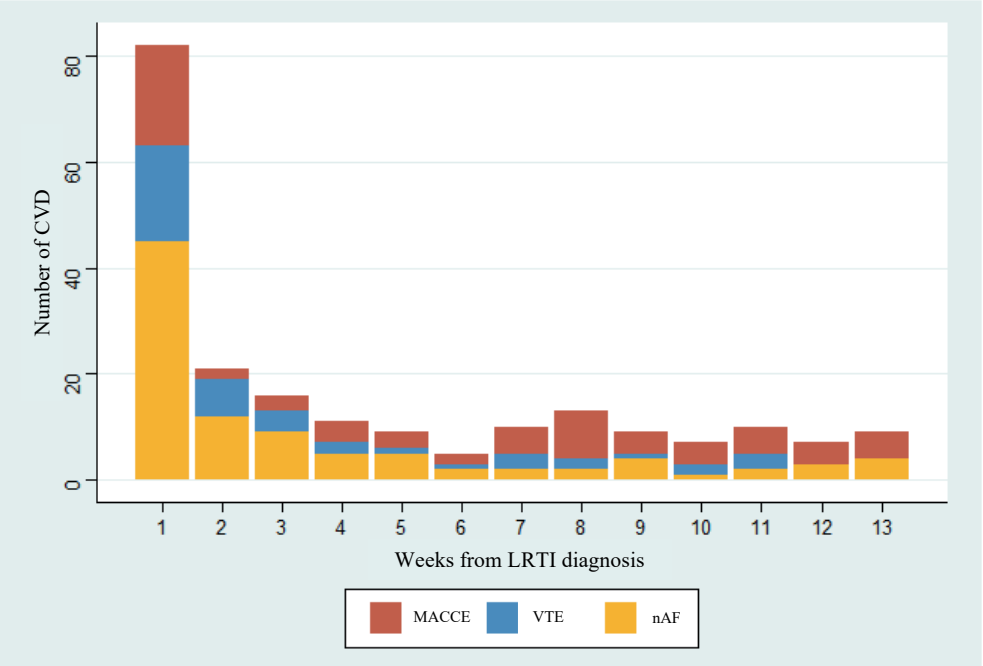


Figure S7. Time to acute cardiovascular disease in weeks from LRTI diagnosis

LRTI, lower respiratory tract infection; CVD, cardiovascular disease; MACCE, major adverse cardiac and cerebrovascular event; VTE, venous thromboembolism; nAF, new-onset atrial fibrillation.

Table S8. Baseline characteristics of SCCS cohorts

	MACCE	VTE	New-onset AF
Total, n	406	99	477
Median age at inclusion, years (IQR)	69.5 (62 – 78)	67 (58.5 – 74)	72 (65 – 79)
Female sex, n (%)	163 (40.1)	56 (56.6)	237 (49.7)
Mean follow-up time, years (SD)	3.9 (0.4)	3.8 (0.5)	3.8 (0.5)
Influenza vaccinations between 2015–2019, n (%)			
No indication	21 (5.2)	12 (12.1)	18 (3.8)
0	145 (35.7)	38 (38.4)	145 (30.4)
1	33 (8.1)	10 (10.1)	47 (9.9)
2	33 (8.1)	<10 (*)	43 (9.0)
3	42 (10.3)	<10 (*)	54 (11.3)
4	51 (1.3)	<10 (*)	61 (12.8)
5	81 (20.0)	18 (18.2)	109 (22.9)
(Acquired) history of, n (%)			
Diabetes	31 (7.6)	<10 (*)	40 (8.4)
Malignancy	52 (12.8)	18 (18.2)	79 (16.6)
Number of respective CVD during study period, n (%)			
1	317 (78.1)	78 (78.8)	477 (100.0)
≥2	89 (21.9)	21 (21.2)	-
Death during study period	18 (4.4)	5 (5.0)	24 (5.0)

* As per output restrictions of Statistics Netherlands (CBS), it is not allowed to present values based on fewer than 10 observations. These values are therefore indicated with '<10'. To be able to report the number of deaths during the study period we received an exemption for these specific values.

SCCS, self-controlled case-series; MACCE, major adverse cardiac and cerebrovascular event; VTE, venous thromboembolism; AF, atrial fibrillation; CVD, cardiovascular disease; SD, standard deviation; MACCE, major adverse cardiac and cerebrovascular event.

Table S9. Estimated incidence rate ratios per cardiovascular disease type within 90 days following LRTI diagnosis from additional sensitivity analyses

Risk period	**Incidence rate ratio (95% CI)**		
Days following LRTI diagnosis	**MACCE**	**VTE**	**New-onset AF**
Primary analysis			
0–90 days	1.3 (1.0–1.8)	6.8 (4.4–10.4)	2.9 (2.3–3.7)
Sensitivity analyses (0–90 days)			
Excluding patients who died during study period	1.3 (1.0–1.8)	6.4 (4.1–9.9)	2.7 (2.1–3.5)
Extended follow-up after death	1.3 (1.0–1.8)	6.8 (4.4–10.4)	2.9 (2.3–3.7)
First events only	1.2 (0.8–1.6)	5.3 (3.0–9.2)	2.9 (2.3–3.7)
Adjusted for MACCE and VTE recurrence	1.3 (1.0–1.8)	6.8 (4.4–10.4)	2.9 (2.3–3.7)
Extended pre-exposure period (28 days)	1.3 (1.0–1.8)	6.6 (4.3–10.2)	3.0 (2.3–3.8)

LRTI, lower respiratory tract infection; CI, confidence interval; MACCE, major adverse cardiac and cerebrovascular event; VTE, venous thromboembolism; AF, atrial fibrillation.

Table S10. Calculations of for excess CVD in risk period based on incidence rate ratios obtained from SCCS analyses

Risk period	No. outcomes in control period	Follow-up time in control period (person-days)	Incidence rate control period (per person-day)	IRR (95% CI)	Excess CVD incidence rate in risk period per person-day (95% CI)	Follow-up time in risk period (person-days)	Number of LRTI episodes in LRTI cohort
Days following LRTI				**MACCE**			
0 – 90 days	395	436 668	0.000905	1.3 (1.0 – 1.8)	0.00031 (0.00002 – 0.00069)	57 237	16 497
1 – 90 days				1.3 (0.9 – 1.6)	0.00023 (-0.00006 – 0.00059)	56 630	
2 – 90 days				1.2 (0.9 – 1.6)	0.00020 (-0.00008 – 0.00056)	56 023	
3 – 90 days				1.1 (0.8 – 1.5)	0.00014 (-0.00014 – 0.00048)	55 417	
				VTE			
0 – 90 days	67	106 426	0.000630	6.8 (4.4 – 10.4)	0.0037 (0.0021 – 0.0059)	13 587	16 497
1 – 90 days				5.9 (3.8 – 9.3)	0.0031 (0.0018 – 0.0060)	13 442	
2 – 90 days				5.6 (3.6 – 8.8)	0.0029 (0.0016 – 0.0049)	13 298	
3 – 90 days				5.1 (3.2 – 8.1)	0.0026 (0.0014 – 0.0045)	13 155	
				New-onset AF			
0 – 90 days	295	514 472	0.000573	2.9 (2.3 – 3.7)	0.0011 (0.00072 – 0.0015)	64 780	16 092
1 – 90 days				2.3 (1.8 – 3.0)	0.00077 (0.00045 – 0.0012)	64 082	
2 – 90 days				2.1 (1.6 – 2.8)	0.00065 (0.00035 – 0.0010)	63 383	
3 – 90 days				2.0 (1.5 – 2.7)	0.00059 (0.00029 – 0.00096)	62 683	

SCCS, self-controlled case-series; LRTI, lower respiratory tract infection; IRR, incidence rate ratio; CI, confidence interval; CVD, cardiovascular disease; MACCE, major adverse cardiac and cerebrovascular event; VTE, venous thromboembolism; AF, atrial fibrillation.

9 General discussion

In this thesis, I aimed to improve the prognostication of adults diagnosed with a lower respiratory tract infection (LRTI) by their general practitioner. While doing so, I considered a broad range of relevant outcomes, with a particular focus on cardiovascular disease (CVD). To this end, I combined state-of-the-art prognostic and etiologic research methods, including natural language processing (NLP), to evaluate the impact of CVD on the prognosis of LRTI as both a predictor and an outcome.

This thesis is the result of a trajectory combining a GP residency training with a position as clinician-researcher and I will therefore primarily discuss the implications of my thesis' findings in the context of the general practitioner's (GP) consultation room.

I will first briefly summarise the main findings of this thesis. Next, I will discuss how the individual risk prediction model for 30-day all-cause hospitalisation or mortality based on structured electronic health records (EHR) data may improve patient outcomes of patients with LRTI in general practice (Chapter 5 and 7), while reflecting upon common challenges and pitfalls of implementing prediction models in clinical practice. Then, I will focus on the additional burden of LRTIs due to the temporary increased risk of acute cardiovascular events as outlined in Chapter 8 and discuss potential strategies to reduce the risk of these events. Finally, using an illustrative example of a targeted case-finding strategy for new onset atrial fibrillation (nAF) in LRTI patients, I will elaborate on balancing the risks and benefits of such a strategy and emphasize the need for tailored interventions to avoid overdiagnosis and (subsequent) overtreatment as much as possible – in line with the "first do no harm" principle.

Main findings of this thesis

The majority of my thesis centred around improving the prognostication of patients presenting to general practice with LRTI in terms of a complicated disease trajectory, that is, all-cause hospitalisation or mortality within 30 days from LRTI diagnosis. I synthesised available prognostic literature for these outcomes (**Chapter 2**) and found that previously developed prediction models are currently not suitable for implementation in everyday clinical practice due to high risk of bias and incomplete assessment of model performance. However, promising prognostic factors to be considered when developing a new model were identified. As preparatory work, I elaborated on the design and rationale of the prediction modelling study (**Chapter 3**) and evaluated recording patterns of signs and symptoms in EHR clinical notes (**Chapter 4**) – in which a substantial amount of non-recorded, and, when recorded, potentially selective values were observed. In **Chapter 5**, I developed and externally validated a new prediction model based on readily available structured EHR data on demographics, medical history, medication use, and type of LRTI. The model stratifies LRTI patients based on 30-day risk of all-cause hospitalisation or mortality and showed promising model performance upon external validation. During further preparatory work, I fine-tuned and compared Dutch large language models for

extraction of data on signs and symptoms from EHR clinical notes (**Chapter 6**), demonstrating good performance. In **Chapter 7**, it was reported that, although these signs and symptoms improved model performance beyond age and sex, their incremental value beyond the structured EHR-based predictors from Chapter 5 was limited for prediction of 30-day all-cause hospitalisation or mortality.

In **Chapter 8**, I shift the focus to the temporary increased risk of acute cardiovascular events attributable to LRTIs. Although this association has been well-established in previous studies, absolute risks were not reported, thus leaving the population impact unknown. I estimated that, among adult patients with LRTI, the risk of major adverse cardiac and cerebrovascular events (defined as acute MI, stroke, or transient ischaemic attack (TIA)) increased by 1.3-fold, the risk of venous thromboembolism (VTE, defined as PE or DVT) by 6.8-fold, and the risk of nAF by 2.9-fold within 90 days following LRTI diagnosis, culminating in approximately five to nine excess cardiovascular events per 1,000 LRTI patients.

PREDICTION MODELS: FROM PAPER TO PRACTICE

The ultimate goal of a newly derived prediction model is to improve individual patient outcomes in every clinical practice. For diagnostic models, this may entail accurate prediction of a patient's diagnosis to effectively guide management strategies. Prognostic models, on the other hand, can guide preventative or therapeutic (mitigative) management decisions when accurately estimating individuals' risk of future health outcomes.

Although thousands of predictions models have been developed, only few make it to clinical practice guidelines and are applied in everyday practice.[1] Guidelines of the Dutch College of General Practitioners currently include prognostic and diagnostic models, mainly in the field of CVD. For example, to support GPs in the diagnostic process of VTE, two models can estimate risk of PE (the YEARS algorithm[2]) and DVT.[3] In addition, the SCORE2 model predicts the 10-year risk of fatal and non-fatal cardiovascular events in community individuals without prior CVD or diabetes and can be used to guide preventive CVD strategies, including lifestyle.[4]

The model reported in Chapter 5 of my thesis was developed to predict individual risk of all-cause hospitalisation or mortality within 30 days following LRTI diagnosis. It could therefore aid GPs in identifying LRTI patients at highest risk of a complicated disease trajectory. Given the high incidence of LRTIs in general practice – roughly three percent of the adults present to their GP with a community-acquired LRTI annually[5] – and the diagnostic challenges to identify those with more severe infections, e.g., community-acquired pneumonia (CAP)[6,7], the prediction model addresses a relevant clinical problem.

Hurdles on the way to implementation

Notwithstanding the potential clinical relevance of a prediction model, there are many reasons why a prediction model does not make it to clinical practice, as has previously been depicted as the 'leaky prognostic model adoption pipeline' by *Van Royen et al*.[8] Applying this 'framework' to the prediction models reported in Chapter 5 and 7, with the 'final' model as presented in Chapter 5 considered as the preferred model, we could argue that the first two out of four hurdles of bringing a prediction model to clinical practice have successfully been overcome (Table 1). First, I have developed a prediction model for a clinically relevant problem based on predictors that are readily available from EHR. Integration within EHR systems should therefore be relatively straightforward, facilitating easy use by GPs in the consultation room. Second, the model has been developed using state-of-the-art techniques and has been externally validated in recent (post-pandemic) data, showing promising model performance. Whether the model will have a positive impact on individual patient outcomes (hurdle three) and, will ultimately be adopted (hurdle four) is, however, yet to be determined.

Table 1. Hurdles on the way to model adoption applied to our model

Hurdle	**How this was overcome**
Fit for purpose	Relevant clinical problem in well-defined population, readily available predictors, easy-to-use model, reliable outcome measurement
Validation	State-of-the-art model development (minimising overfitting), temporal internal-external cross-validation, external validation in (recent) post-pandemic data
	How to overcome
Implementation	Define position within current clinical practice, define actionable risk thresholds, apply user- and patient centred implementation strategy and assess effects on relevant patient outcomes
Adoption	Integration within general practice electronic health records to provide point-of-care predictions

As proposed by *Van Royen et al.* Eur Resp J (2022).

Indeed, a well-developed and accurate prediction model does not guarantee that it will have clinical value when used for a patient in the GP's consultation room. A common issue for models developed using EHR data is that these may incorporate predictors that are selectively reported, as is the case for signs, symptoms, and vital sign measurements as described in Chapter 4. When clinicians – urged by using such a prediction model – start to routinely assess these predictors, their predictive value, and thus the predictive performance of the model as a whole will change. This problem is also referred to as 'the curse of knowing'.[9] Prediction models that do not include predictors with informative missingness, as is the case for our 'final' model reported in Chapter 5 which incorporates solely structured EHR data, may circumvent this problem and provide accurate

predictions when used in clinical practice. Nevertheless, even when a model generates accurate predictions it does not necessarily lead to improvement of patient outcomes, as previous studies have shown. For example, the impact of using a model that predicts 31-day mortality risk among patients presenting to the emergency department ($RISK^{INDEX}$) – developed to support rapid risk stratification and optimise allocation of healthcare resources – was evaluated in a randomised controlled trial. Although the model yielded accurate predictions that outperformed both clinical intuition and prevailing risk prediction tools (i.e., NEWS[10], APACHE II[11], and SOFA[12]), it did not alter treatment decisions or clinical outcomes.[13] Moreover, in the absence of benefits for individual patients, use of prediction models can even result in adverse effects in terms of increasing the burden on healthcare resources. This was the conclusion of the PRISMATIC randomised stepped-wedge trial, in which the introduction of a predictive risk stratification model (PRISM) – estimating risk of emergency hospital admission within a year for patients registered in general practice – led to an increase in emergency department visits, hospitalisations, and healthcare costs.[14] Furthermore, recent insights on the clinical impact of prediction models emphasize that, although accurate, predictions may even lead to harmful self-fulfilling prophecies.[15] For example, clinicians may be more inclined to refrain from intensive treatment in patients with a predicted poor prognosis, whereas these patients could actually benefit most. Post-deployment model performance would remain favourable in these situations, which underlines the importance of a focus on relevant, and preferably mitigatable, individual patient outcomes rather than prognostic accuracy in the implementation phase of a model.

Bringing the newly derived model to the consultation room

What can we learn from these examples, in which accurate predictions have no, or even counterintuitive, effects? And how could this help to bring the model developed in Chapter 5 this thesis closer to successful implementation in clinical practice?

These studies show us that well-developed prediction models do not necessarily lead to successful implementation. They underline the importance of acknowledging that a prediction model will always be used alongside clinical judgement, and not as a stand-alone decision tool – it is a decision support tool. This implies that, while developing and implementing prognostic models, its role in the clinical decision-making process should be given careful consideration before implementation, and its use by clinicians should be actively studied during implementation to ensure it aligns with clinical practice to maximise impact.[16] Such a 'user-centred approach' also warrants predicted outcomes that are clinically relevant and actionable, with easily interpretable guidance on the down-stream interventions that could follow from explicitly defined thresholds. In other words, it should be acknowledged that introducing a prediction model in clinical practice, even in the case of accurate, user-centred and actionable models, is a complex intervention demanding understanding, trust, and adherence of clinicians and clinical practice.[17]

Returning to the prediction model developed in Chapter 5 of this thesis, I aimed to aid GPs with stratification of LRTI patients based on short-term risk of hospitalisation or mortality from any cause. I would argue that the outcome predicted by this model – hospitalisation or mortality irrespective of the underlying cause – is particularly relevant from the patients' perspective and actionable from a GP's point of view: whether a patient diagnosed with LRTI will be hospitalised due to severe LRTI or, for example, deterioration of concurrent heart failure may not be the relevant question for a patient. Timely recognition of deterioration and appropriate management aimed at preventing hospitalisation is in the interest of both the patients and GP. This implies that the model should aid GPs to identify those patients that need closer monitoring. Although the model is unlikely to yield risks that justify hospital admission decisions and cannot be used to guide antibiotic treatment as this was not modelled during development, the range of risks predicted by the model – between 1.5% and 51.3% in the validation cohort – does include thresholds that can be used to guide close monitoring strategies. Such strategies have no clear adverse effects and thus a larger proportion of unnecessary monitoring in retrospect ('false positives') is acceptable.

In sum, our prediction model may be used to aid GPs in deciding which LRTI patients deserve closer monitoring to promptly recognize deterioration. Moreover, it can easily be integrated in existing EHR systems since it is based on structured EHR data only. However, before widespread implementation, GPs, patients, and other relevant stakeholders should inform clinically relevant risk thresholds and downstream intervention strategies to ensure successful implementation, GPs' adherence and trust, and alignment with the patient's perspective.

LRTI-ASSOCIATED CARDIOVASCULAR RISK: FROM KNOWLEDGE TO PRACTICE

In Chapter 8 of this thesis, I have provided unique data on the absolute risk of a broad range of cardiovascular events associated with community-acquired LRTI and estimated excess cardiovascular event rates to provide insight into the population impact. Albeit these absolute risks may seem modest at first glance, the studied acute cardiovascular events are associated with high morbidity and mortality.[18] Moreover, LRTIs are among the most common infections encountered in general practice – with an incidence of approximately three percent annually[5] – thus culminating in a substantial number of excess cardiovascular events.

Prediction of cardiovascular risk following LRTI could allow for early detection or even prevention of cardiovascular events. However, the low (absolute) outcome event rates as reported in Chapter 8 impede the development of clinically relevant prediction models.[19]

In such cases, a model's discriminative ability may be high, but is largely driven by predicting absence of the outcome. An illustrative example of this problem is the DASHI score, predicting individual risk of a composite outcome including MI, stroke, TIA, or death from these diagnoses within 28 days following RTI diagnosis in general practice.[20] With an overall event rate of 0.3%, model performance in terms of discrimination (c-statistic 0.85 (95% CI 0.85–0.85)) and calibration (observed to expected ratio 0.85 (95% CI 0.85–0.85)) was promising. Individual predictions ranged from 0.04% to 35.6%, but the distribution of predicted risks was not reported leaving the clinical relevance of the risks, theoretically, unknown. Decision curve analysis, however, indicated negligible clinical benefit of the DASHI score with a net benefit below 0.001 at a risk threshold of 1%.

Given the low absolute cardiovascular event rates following LRTI, I refrained from my intention, as outlined in Chapter 3, to develop a prediction model for these outcomes. Alternatively, I performed an etiologic study to estimate the excess cardiovascular risk associated with LRTI to unravel the population impact. The gained insights could guide potential interventions aimed at mitigating this risk. I found that, for every 1,000 patients with GP-diagnosed LRTI, 0.5 to 1.1 cases of MI, stroke, or TIA, 2.1 to 3.0 cases of VTE, and 2.3 to 4.4 new diagnoses of AF within 90 days are attributable to LRTI. These figures were based on estimated incidence rate ratios, indicating that LRTIs increased the risk of 90-day cardiovascular events by 1.3-fold (95% CI 1.0–1.8) for a composite of MI, stroke, or TIA, 6.8-fold (95% CI 4.4–10.4) for VTE, and 2.9-fold (95% CI 2.3–3.7) for nAF.

LRTI-associated cardiovascular risk in context of well-established CVD risk factors

How should we move from these risks to clinical implications? Do such risks justify efforts to develop and test interventions aimed at early detection or prevention of cardiovascular events?

To provide some context to my findings in Chapter 8, let me compare the increased short-term cardiovascular risk associated with LRTI to other, well-established, risk factors and triggers for cardiovascular events (Figure 1). Smoking, for example, is well known to increase the risk of a variety of cardiovascular diseases, including major adverse cardiac and cerebrovascular events (MACCE). A large prospective cohort study with up to 20 years follow-up found that current smokers have a significantly increased risk of MACCE compared to never-smokers (hazard ratio 1.7 (95% CI 1.7–1.8)).[21] In addition, evidence on the association between hypertension and increased risk of MACCE is ample. A systematic review and meta-analysis of randomised trials generally spanning between 1 and 5 years found that, among individuals with hypertension, every 10 mmHg increase in systolic blood pressure resulted in a relative risk (RR) of 1.3 (95% CI 1.2–1.3) for MACCE.[22] Importantly, these effects occurred following long-term exposure to either smoking or hypertension and over a long follow up time. This is significantly different from a transient

exposure such as LRTI, lasting days to some weeks, with only a temporarily increased cardiovascular risk. Impact of long-term and transient exposures should therefore be compared with caution. As a transient exposure, use of non-steroidal anti-inflammatory drugs (NSAIDs) has also been linked to increased risk of MI. Use of any NSAID was related to an increased odds of MI, with a 1.6-fold higher odds of MI (95% CI 1.2–2.0).[23] This finding resulted in diclofenac being contraindicated in, among others, patients with a history of ischaemic heart disease.[24]

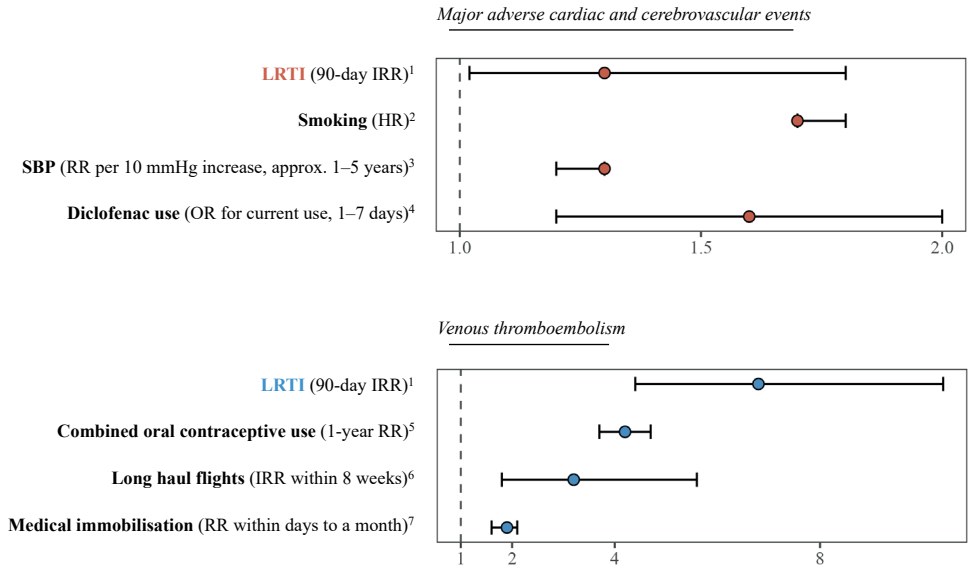


Figure 1. Comparison of cardiovascular risk following LRTI compared to risk effect estimates of well-established CVD risk factors

1. Chapter 8 of this thesis. 2. Tasdighi et al. PLoS Medicine (2025). 3. Ettehad et al. Lancet (2016). 4. Bally et al. BMJ (2017). 5. Lidegaard et al. BMJ (2017). 6. Kuipers et al. PLoS Medicine (2007). 7. Pottier et al. Thromb Res (2009).

Abbreviations: LRTI, lower respiratory tract infections; IRR, incidence rate ratio; HR, hazard ratio; SBP, systolic blood pressure; RR, relative risk; OR, odds ratio.

Well-known transient exposures that increase the risk of VTE include combined oral contraceptive use, long haul flights, and immobilisation. Compared to non-users, users of combined oral contraceptives have a 4.2 times higher risk (95% CI 3.7–4.7) of VTE during the first year.[25] Long haul flights, defined as a flight duration of at least four hours, are associated with a 3.2-fold (95% CI 1.8–5.6) increased incidence in VTE within 8 weeks, a risk that further increases with duration of travel and number of flights.[26] Evidence suggests that individuals at highest VTE risk – based on factors such as recent trauma or surgery, previous VTE or varicose veins, active malignancies, hormone therapy, and obesity – could benefit from wearing compression stockings on such flights.[27] Lastly, recent

medical immobilisation, regardless of the underlying disease, was found to increase the risk of VTE by 1.9-fold (95% CI 1.6–2.1) based on a meta-analysis of studies that applied time-windows varying from several days to a month.[28] For hospitalised patients, surgical and non-surgical, prophylactic antithrombotic therapy can therefore be indicated based on individual VTE risk factors or by using the Padua prediction rule.[29,30]

Potential cardiovascular risk mitigating strategies for patients with an LRTI

Summarising these traditional risk factors, including the magnitude of their effects and the currently adopted preventive strategies for some of these factors and subsequently comparing these to the increased risk of cardiovascular events associated with LRTI, warrants, at minimum, increased awareness among clinicians that LRTIs are a non-traditional and currently undervalued risk factor for acute cardiovascular events. In addition, acknowledging the temporarily increased cardiovascular risk following LRTI diagnosis may provide patients and GPs with a novel and unique opportunity for early detection or prevention of acute cardiovascular events. Especially since the period at highest risk is short (i.e., several weeks following LRTI diagnosis) and interventions may therefore be only of temporary nature, potentially increasing patient's adherence and minimising adverse effects. This may entail strategies aimed at close monitoring of unfolding cardiovascular symptoms, short-course antiplatelet therapy to prevent MI or stroke, short-course anticoagulant treatment to prevent thromboembolic events (DVT or PE), and/or targeted case finding of nAF, as briefly outlined in Table 2. In addition to these strategies, effective and low-cost treatments that reduce viral spread in the community and reduce symptomology, such as saline nasal spray, can prevent patients from progressing from a 'simple' upper RTI to LRTI and may therefore also contribute to mitigation of complication risks, including cardiovascular.[31]

Given the low absolute (and temporary) risk of cardiovascular events, the question is, however, what risk threshold justifies an intervention, and what intervention is most appropriate. This implies that benefits and risks should be carefully weighted and warrants tailoring interventions to those LRTI patients at highest cardiovascular risk. Input on such risk thresholds and acceptability of preventive interventions is needed from both patients and clinicians. Both perspectives are pivotal to guide potentially effective and acceptable risk mitigation strategies. In future projects on this topic, researchers should therefore actively engage with patients and clinicians to co-create, fine-tune and pilot test appropriate preventive interventions in patients with LRTI.

Table 2. Potential strategies to mitigate cardiovascular risk in LRTI patients

Aim	Strategy	Potential benefits	Potential adverse effects
Early detection	Risk-based close monitoring of unfolding cardiovascular symptoms ('watchful waiting')	Prompt recognition and treatment (if indicated)	Increased workload for GP
Early detection	Targeted case-finding of nAF among LRTI patients at highest risk of clinically relevant nAF	Prompt recognition and treatment of nAF and mitigation of complication risk (e.g., stroke)	Providing patients with a diagnosis of unknown clinical relevance (i.e., for subclinical AF) and uncertainties regarding anticoagulant treatment and the duration of such treatment
Prevention	Short-course anticoagulant treatment among LRTI patients at highest risk of VTE	Prevention of VTE	(Major) bleedings, other side-effects
Prevention	Short-course antiplatelet treatment among LRTI patients at highest risk of MACCE	Prevention of MACCE	(Major) bleedings, other side-effects

LRTI, lower respiratory tract infection; GP, general practitioner; nAF, new-onset atrial fibrillation; VTE, venous thromboembolism; MACCE, major adverse cardiac and cerebrovascular event.

Such project may be informed by the MRC framework for developing and evaluating complex interventions, with parallel process evaluation and pilot-testing of interventions using Plan-Do-Study-Act cycles.[32] In addition, while determining the need for these interventions, societal issues such as limited health care resources and the impact of increasing use of diagnostic tests and potential overtreatment should be considered. For example, low-cost strategies aimed at close monitoring of unfolding cardiovascular symptoms will likely not result in unnecessary treatment whereas preventative antithrombotic treatment could. As such, potential benefits of interventions should be carefully balanced against risks of adverse effects and societal impact.

BALANCING EARLY DETECTION STRATEGIES AND OVERDIAGNOSIS: AN ILLUSTRATIVE EXAMPLE

Patients with increased risk of major adverse cardiac and cerebrovascular events or VTE following LRTI may benefit from interventions aimed at early detection or preventative treatment. Given the short-term increase in cardiovascular risk, such interventions will usually be of temporary nature during the at-risk period only. Risks or disadvantages of these interventions may include a temporary risk of adverse effects, such as risk of bleeding associated with anticoagulant use, that should not outweigh the potential benefits.

An early detection strategy aimed at nAF in LRTI patients has, however, more implications to consider.

AF is a leading cause of ischaemic stroke – increasing its risk up to five-fold – and timely oral anticoagulant treatment lowers this risk substantially.[33–35] Furthermore, AF is known to increase the risk of heart failure and dementia.[36,37] This implies a need for effective and early AF detection and, based on my findings in Chapter 8, LRTIs could provide a unique and novel opportunity for such targeted case-finding strategies. Targeted screening for nAF among LRTI patients will, however, also result in an increasing number of patients diagnosed with AF for whom the question will be whether anticoagulant treatment is necessary, and – when considered necessary – whether this should be short-course or permanent anticoagulant treatment. The underlying question is whether the risk of complications, such as ischaemic stroke, is similarly increased for (subclinical) AF triggered by LRTIs and how we should define clinically relevant AF in LRTI patients in terms of AF duration.

Uncertainties regarding infection-related atrial fibrillation

On this topic, a Danish population-based study found that infection-related AF was associated with AF recurrence – which occurred in over a third of the patients with infection-related AF – and a 2-fold increased risk of thrombo-embolic events within one year (absolute risk 7.6%).[38] However, evidence on the added value of population-based screening for AF is contradicting. Although health economic evaluation of nation-wide AF screening among patients aged ≥65 years in general practice was considered cost-effective and other studies on population-based screening or implantable loop recorders yielded increased AF detection, these strategies showed limited to no benefit on stroke or systemic embolism occurrence.[39–41] This implies that not all AF may be clinically relevant to detect. Although subclinical (or device-detected) AF is associated with increased stroke risk, this is approximately only half the risk compared to clinical AF, which may no longer outweigh bleeding risk due to anticoagulant treatment.[42]

Currently, there is no consensus on whether long-term anticoagulant treatment is indicated in patients with subclinical or infection-related AF, and definitions of relevant subclinical AF vary in terms of duration. Main and post-hoc analyses of the ASSERT trial on device-detected AF among 2,580 patients with a pacemaker found that stroke risk was not increased in patients with subclinical AF lasting less than 24 hours compared to patients without subclinical AF.[43] On the contrary, subclinical AF of at least 24 hours carried an absolute risk of ischaemic stroke of 3.1% per year – similar to that of clinical AF. Given the uncertainty regarding the definition of clinically relevant infection-related AF and its appropriate management, the development of targeted case-finding strategies for AF in patients with LRTI should be carefully weighed against the burden of providing

patients with a permanent diagnosis of unknown clinical relevance and without a well-defined treatment strategy.

Diagnosis-expansion

So, should we explore targeted case-finding strategies for nAF in patients presenting to general practice with LRTI? How confident are we that this would be of any benefit for individual patients and from a societal perspective? And what risks come with lowering thresholds for AF diagnoses?

In 2025, the Council for Public Health & Society of the Netherlands published a report on so-called diagnosis-expansion as an emerging societal issue in healthcare that may provide guidance on these questions.[44] Diagnosis-expansion is defined as a phenomenon of (1) increasing use of diagnostic tests and screening and (2) lowering thresholds for diagnoses ('diagnosis inflation'). Whereas increasing screening and diagnostic testing may seem positive at first glance, downsides are numerous, both from the patient's perspective as for society as a whole. Abnormal test findings of unknown clinical relevance may, for example, result in anxiety for future health issues, a worsening of experienced health (i.e., the 'nocebo' effect), or additional follow-up diagnostic testing methods that carry risks of adverse effects. Moreover, additional testing, re-consultations, and referrals to secondary care pose an increasing burden on limited healthcare resources, both economic and in terms of healthcare professionals. Apart from the individual responsibility of healthcare professionals to weigh the need, costs, and benefits of diagnostic testing, the issue of diagnosis-expansion warrants accountability of other parties involved in healthcare management, politics and the academic community. Especially in times of continuously increasing national healthcare expenditures – increasing with almost 9% in 2024[45] –, an aging population, and a limited number of healthcare professionals, this obligation should be felt in the consultation room and among clinician-researchers. Although general practice accounts for only a small proportion of the total health care costs[46], diagnosis-expansion in general practice will likely result in additional downstream health care costs.

In their report, the Council for Public Health & Society of the Netherlands recommends quantifying adverse effects of lowering thresholds for disease. In addition, the Counsel states that the risk of overdiagnosis and societal effects – such as psychological effects, the burden on limited healthcare resources and diagnosis inflation –, in particular for patients without health complaints as may be the case for LRTI patients with subclinical AF, should be carefully considered before including new management strategies in clinical guidelines. Following this advice, a targeted case-finding strategy for nAF following LRTI should be designed in collaboration with patients to balance costs and benefits that align with their perspective and avoid 'top down' development of interventions.

Considering the beforementioned evidence on the potential of screening strategies for (subclinical) nAF among LRTI patients in light of these societal issues, targeted case-finding strategies should, as a start, be tailored to a selected subgroup LRTI patients at highest risk of clinically relevant AF. Other risk mitigation strategies for cardiovascular events following LRTI should be similarly tailored to those patients at highest risk. This may increase overall benefit while mitigating risk of overdiagnosis and the downstream adverse effects. To safeguard input from a societal perspective, patients and clinicians should be actively involved to inform reasonable risk thresholds and shape appropriate interventions that may ultimately improve individual patient outcomes in the consultation room.

CONCLUSIONS AND FUTURE DIRECTIONS

In this thesis, I aimed to contribute to improving the prognostication of patients presenting to general practice with LRTI. The synthesis of available prognostic literature on a complicated disease trajectory in LRTI patients, covering both individual prognostic factors and prediction models, may inform future clinical guidelines. In addition, the prediction model based on readily available data from EHR – and validated in recent post-pandemic data – can easily be integrated in general practice EHR systems to generate point-of-care predictions. However, whether the use of this model actually improves patient outcomes and how it could complement current clinical practice should first be studied in a future project focussed on the various aspects of model implementation.

The unique insights on the absolute population impact of cardiovascular risk attributable to LRTI provided in this thesis further established LRTIs as a 'non-traditional' risk factor for cardiovascular events. At a minimum, awareness among GPs should be increased that LRTIs are a non-traditional and currently undervalued risk factor for acute cardiovascular events. These findings could also provide a unique opportunity for early detection or prevention of acute cardiovascular events. However, given the low absolute short-term cardiovascular risk following LRTI, future studies aimed at developing risk mitigation strategies among LRTI patients should carefully weigh benefits against adverse effects and societal impact in co-creation with patients, GPs, medical specialists, and other relevant stakeholders. Before doing so, further refinement of cardiovascular risk across various patient subgroups should inform which LRTI patients are at highest risk of cardiovascular events, and in whom cardiovascular risk mitigating strategies may thus yield most benefit.

REFERENCES

1. Arshi B, Wynants L, Rijnhart E, et al. Number of publications on new clinical prediction models: a bibliometric review. *JMIR Med Inform*. 2025;**13**:e62710.
2. Geersing GJ, Takada T, Klok FA, et al. Ruling out pulmonary embolism across different healthcare settings: a systematic review and individual patient data meta-analysis. *PLoS Med*. 2022;**19**(1):e1003905.
3. Oudega R, Moons KGM, Hoes AW. Ruling out deep venous thrombosis in primary care. A simple diagnostic algorithm including D-dimer testing. *Thromb Haemost*. 2005;**94**(1):200–205.
4. SCORE2 working group and ESC Cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J*. 2021;**42**(25):2468–2471.
5. Hak E, Rovers MM, Kuyvenhoven MM, Schellevis FG, Verheij TJM. Incidence of GP-diagnosed respiratory tract infections according to age, gender and high-risk co-morbidity: the second dutch national survey of general practice. *Fam Pract*. 2006;**23**(3):291–294.
6. Teepe J, Broekhuizen BDL, Loens K, et al. Predicting the presence of bacterial pathogens in the airways of primary care patients with acute cough. *CMAJ*. 2017;**189**(2):E50–E55.
7. Minnaard MC, De Groot JAH, Hopstaken RM, 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–E63.
8. van Royen FS, Moons KGM, Geersing GJ, van Smeden M. Developing, validating, updating and judging the impact of prognostic models for respiratory diseases. *Eur Respir J*. 2022;**60**(3):2200250.
9. Groenwold RHH. Informative missingness in electronic health record systems: the curse of knowing. *Diagn Progn Res*. 2020;**4**:8.
10. Royal College of Physicians. *National Early Warning Score (NEWS): Standardising the Assessment of Acute-Illness Severity in the NHS*. Report of a Working Party. London: Royal College of Physicians; 2012.
11. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med*. 1985;**13**(10):818–829.
12. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med*. 1996;**22**(7):707–710.
13. van Dam PMEL, van Doorn WPTM, Sevenich L, et al. Machine learning for risk stratification in the emergency department (MARS-ED): a randomized controlled trial. *Nat Commun*. 2025;**17**(1):242.
14. Snooks H, Bailey-Jones K, Burge-Jones D, et al. Predictive risk stratification model: a randomised stepped-wedge trial in primary care (PRISMATIC). *Health Serv Delivery Res*. 2018;**6**(1):1–164.
15. van Amsterdam WAC, van Geloven N, Krijthe JH, Ranganath R, Cinà G. When accurate prediction models yield harmful self-fulfilling prophecies. *Patterns*. 2025;**6**(4):101229.
16. Seneviratne MG, Li RC, Schreier M, et al. User-centred design for machine learning in health care: a case study from care management. *BMJ Health Care Inform*. 2022;**29**(1):e100656.
17. Kappen TH, van Klei WA, van Wolfswinkel L, et al. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. *Diagn Progn Res*. 2018;**2**(1):11.
18. Global Burden of Cardiovascular Diseases and Risks 2023 Collaborators. Global, regional, and national burden of cardiovascular diseases and risk factors in 204 countries and territories, 1990-2023. *JACC*. 2025;**86**(22):2167–2243.
19. Cartus AR, Samuels EA, Cerdá M, Marshall BDL. Outcome class imbalance and rare events: an underappreciated complication for overdose risk prediction modelling. *Addiction*. 2023;**118**(6):1167–1176.

20. Lee JJ, Koshiaris C, Wright-Drakesmith C, et al. Development and external validation of a risk prediction score (DASHI) for cardiovascular events following acute respiratory infections: derivation and validation retrospective cohort study. *eClinicalMedicine*. 2025;**84**:103273.
21. Tasdighi E, Yao Z, Dardari ZA, et al. Association between cigarette smoking status, intensity, and cessation duration with long-term incidence of nine cardiovascular and mortality outcomes: the Cross-Cohort Collaboration (CCC). *PLoS Med*. 2025;**22**(11):e1004561.
22. Ettehad D, Emdin CA, Kiran A, et al. Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. *Lancet*. 2016;**387**(10022):957–967.
23. Bally M, Dendukuri N, Rich B, et al. Risk of acute myocardial infarction with NSAIDs in real world use: bayesian meta-analysis of individual patient data. *BMJ*. 2017;**357**:j1909.
24. Keizer D, Luiten W, Schouten F, et al. *NHG-richtlijn Pijn* [internet]. Utrecht: Nederlands Huisartsen Genootschap; 2024. Accessed January 22, 2026. Available at: https://richtlijnen.nhg.org.
25. Lidegaard Ø, Løkkegaard E, Svendsen AL, Agger C. Hormonal contraception and risk of venous thromboembolism: national follow-up study. *BMJ*. 2009;**339**(7720):557–560.
26. Kuipers S, Cannegieter S, Middeldorp S, et al. The absolute risk of venous thrombosis after air travel: a cohort study of 8,755 employees of international organisations. *PLoS Med*. 2007;**4**(9):e290.
27. Watson HG, Baglin TP. Guidelines on travel-related venous thrombosis. *Br J Haematol*. 2011;**152**(1):31–34.
28. Pottier P, Hardouin JB, Lejeune S, et al. Immobilization and the risk of venous thromboembolism. A meta-analysis on epidemiological studies. *Thromb Res*. 2009;**124**(4):468–476.
29. Nederlandse Internisten Vereniging, Federatie Medisch Specialisten. *Richtlijn antitrombotisch beleid* [internet]. Richtlijnendatabase; 2025. Accessed January 14, 2026. Available from: https://richtlijnendatabase.nl/richtlijn/antitrombotisch_beleid/.
30. Barbar S, Noventa F, Rossetto V, et al. A risk assessment model for the identification of hospitalized medical patients at risk for venous thromboembolism: the Padua prediction score. *J Thromb Haemost*. 2010;**8**(11):2450–2457.
31. Little P, Vennik J, Rumsby K, et al. Nasal sprays and behavioural interventions compared with usual care for acute respiratory illness in primary care: a randomised, controlled, open-label, parallel-group trial. *Lancet Respir Med*. 2024;**12**(8):619–632.
32. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of medical research council guidance. *BMJ*. 2021;**374**:n2061.
33. Hart RG, Benavente O, Mcbride R, Pearce LA. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. *Ann Intern Med*. 1999;**131**(7):492–501.
34. Giugliano RP, Ruff CT, Braunwald E, et al. Edoxaban versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2013;**369**(22):2093–2104.
35. Granger CB, Alexander JH, McMurray JJV, et al. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2011;**365**(11):981–992.
36. Nicoli CD, O'Neal WT, Levitan EB, et al. Atrial fibrillation and risk of incident heart failure with reduced versus preserved ejection fraction. *Heart*. 2022;**108**(5):353–359.
37. Rivard L, Friberg L, Conen D, et al. Atrial fibrillation and dementia: a report from the AF-SCREEN international collaboration. *Circulation*. 2022;**145**(5):392–409.
38. Gundlund A, Olesen JB, Butt JH, et al. One-year outcomes in atrial fibrillation presenting during infections: a nationwide registry-based study. *Eur Heart J*. 2020;**41**(10):1120–1122.
39. Svendsen JH, Diederichsen SZ, Højberg S, et al. Implantable loop recorder detection of atrial fibrillation to prevent stroke (The LOOP Study): a randomised controlled trial. *Lancet*. 2021;**398**(10310):1507–1516.

40. Svennberg E, Friberg L, Frykman V, et al. Clinical outcomes in systematic screening for atrial fibrillation (STROKESTOP): a multicentre, parallel group, unmasked, randomised controlled trial. *Lancet*. 2021;**398**(10310):1498–1506.
41. Van Hulst M, Tieleman RG, Zwart LAR, et al. Health economic evaluation of nation-wide screening programmes for atrial fibrillation in the Netherlands. *Eur Heart J Qual Care Clin Outcomes*. 2023;**9**(4):408–416.
42. Healey JS, Connolly SJ, Gold MR, et al. Subclinical atrial fibrillation and the risk of stroke. *N Engl J Med*. 2012;**366**(2):120–9.
43. Van Gelder IC, Healey JS, Crijns HJ, et al. Duration of device-detected subclinical atrial fibrillation and occurrence of stroke in ASSERT. *Eur Heart J*. 2017;**38**(17):1345–1347.
44. Raad voor Volksgezondheid en Samenleving. *Iedereen bijna ziek: over de keerzijden van diagnose-expansie* [internet]. Den Haag: Raad voor Volksgezondheid en Samenleving; 2025. Accessed January 20, 2026. Available from: https://www.raadrvs.nl/documenten/2025/04/15/advies-iedereen-bijna-ziek.
45. Statistics Netherlands (CBS). *Spending on health and social care up by 8.9 percent in 2024* [internet]. Den Haag: Statistics Netherlands (CBS); 2025. Accessed January 20, 2026. Available from: https://www.cbs.nl/en-gb/news/2025/41/spending-on-health-and-social-care-up-by-8-9-percent-in-2024.
46. Zorginstituut Nederland. *Zorgcijfersdatabank* [internet]. Den Haag: Zorginstituut Nederland; 2025. Accessed January 30, 2026. Available from: https://www.zorgcijfersdatabank.nl.

Appendices

SUMMARY

The rationale behind this thesis

Lower respiratory tract infections (LRTI) are among the most common infections encountered in general practice with around three percent of adults consulting their general practitioner (GP) with an LRTI annually. Uncomplicated LRTIs generally run a favourable course in the absence of 'risk factors' for a complicated trajectory. Nevertheless, adverse outcomes, such as hospitalisation or mortality, do occur – especially in patients with more severe LRTI such as community-acquired pneumonia (CAP). A prediction model to aid GPs in identifying LRTI patients at highest risk of adverse outcomes could improve clinical decision-making but is currently lacking in everyday practice.

Complications from LRTIs reach beyond a complicated disease trajectory in terms of hospitalisation or mortality. Evidence consistently indicates that LRTIs can trigger acute cardiovascular events, such as myocardial infarction (MI), stroke, venous thromboembolism (VTE), and atrial fibrillation (AF). Although previous studies found that LRTIs increase the risk of cardiovascular events up to five-fold, none reported estimates of absolute risk. This leaves the population impact of cardiovascular risk associated with LRTIs unknown.

Throughout this thesis, we focussed on improving the prognostication of patients presenting to general practice with LRTI. While doing so, we considered a broad range of relevant outcomes, ranging from adverse outcomes such as hospitalisation and mortality to acute cardiovascular events. In the various chapters of this thesis, we applied state-of-the-art prognostic and etiologic research methods to electronic health records (EHR) data. In addition, we evaluated whether data derived from EHR unstructured clinical notes by natural language processing (NLP) improved prognostication of LRTI patients in general practice.

Findings of this thesis

Chapter 2 reports the findings of a systematic literature review to identify and synthesise the available evidence base on relevant existing prognostic factors and prediction models for all-cause hospitalisation and mortality within 90 days for adults with a GP-diagnosed LRTI. Increasing age, sex, current smoking, a history of diabetes, stroke, cancer, or heart failure, previous hospitalisation, influenza vaccination status, current use of systemic corticosteroids, recent antibiotic use, respiratory rate ≥25/minute, and a clinical diagnosis of pneumonia were identified as promising prognostic factors to be considered when developing or updating a prediction model. Currently available prediction models were considered not suitable for implementation in everyday clinical practice due to high risk of bias and incomplete assessment of model performance.

Chapter 3 describes the rationale for, and design of the development and external validation of two EHR-based prediction models that predict individual risk of (1) 30-day all-cause hospitalisation or mortality and (2) 90-day cardiovascular events in patients presenting to the GP with LRTI. For the first model, the prespecified design was followed with the findings reported in Chapter 5. The observed low event rate of cardiovascular events following LRTI (as reported in Chapter 8) precluded development of a clinically meaningful model. As an alternative to the intended prediction model development study, an etiologic study design was used to quantify the excess cardiovascular risk attributable to LRTI with the findings reported in Chapter 8.

In **Chapter 4**, recording patterns of signs, symptoms, and vital sign measurements in EHR clinical notes of patients presenting to the GP with an LRTI are evaluated to inform future prediction modelling studies. For these potentially relevant predictors, a substantial amount of nonrecorded values were noted, and when recorded, this was likely selective with abnormal findings more likely to be recorded. This underlines the challenges of using unstructured EHR data in prediction research and implies that, prior to including signs, symptoms, and vital sign measurements from clinical notes, recording patterns of such unstructured data should be carefully assessed to tailor missing data handling techniques.

Chapter 5 reports the stepwise development and external validation of a model predicting individual risk of 30-day all-cause hospitalisation or mortality in adults aged ≥40 years presenting to the GP with an LRTI. Predictors of the 'final' model, based on structured EHR data, include demographics (age and sex), cardiometabolic diseases (history of diabetes, heart failure, ischaemic heart disease, stroke, transient ischaemic attack (TIA) pulmonary embolism (PE), deep venous thrombosis (DVT), atrial fibrillation (AF), and peripheral artery disease), other medical history (history of pneumonia, non-dermatological malignancies, chronic obstructive pulmonary disease, asthma, dementia, hospitalisation in previous year, and influenza vaccination in previous year year), current medication use (antibiotic prescription in previous month, immunosuppressants, inhalation medication, and antidepressants), and a clinical diagnosis of pneumonia. The model was developed in a pre-COVID-19 pandemic cohort (2016–2019, outcome rate: 7.8%) from the region of Utrecht, the Netherlands, and was externally validated in a post-COVID-19 pandemic cohort (2022–2023, outcome rate: 11.4%) from the region of Amsterdam, the Netherlands. External validation yielded a c-statistic of 0.71 (95% confidence interval (CI) 0.69 – 0.73), a calibration intercept of 0.28 (95% CI 0.20 – 0.36), and a calibration slope of 0.95 (95% CI 0.85 – 1.05), with predicted risks ranging from 1.5% to 51.3% (median 6.9%). In addition, the incremental predictive value of cardiometabolic diseases beyond age and sex was found to be low (Dc-statistic +0.02). The 'final' structured EHR data-based model holds promise to aid GPs in identifying LRTI patients at highest risk of 30-day hospitalisation or

mortality. However, before adoption in everyday general practice, future research should assess the model's impact on individual patient outcomes.

Chapter 6 reports the fine-tuning and evaluation of the performance of two Dutch large language models (LLM) (MedRoBERTa.nl and RobBERT) for extraction of candidate predictor information on signs and symptoms – identified as promising in Chapter 2 – from EHR clinical notes of patients presenting to the GP with LRTI. These models demonstrated good performance when implemented as direct classifier – labelling signs and symptoms as either present, absent, or not reported – with an average F1-score of 0.74 (range 0.56 – 0.87) and 0.69 (range 0.46 – 0.86) using 1,600 manually annotated training samples, respectively. Model performance varied substantially across signs and symptoms, with performance decreasing with increasing missingness and class-imbalance. Nevertheless, these findings support the feasibility of using LLMs for automated extraction of signs and symptoms captured in Dutch EHR clinical notes.

In **Chapter 7**, the incremental predictive value of signs and symptoms beyond structured EHR data is evaluated by adding these as candidate predictors to the models presented in Chapter 5. These signs and symptoms were derived from unstructured EHR clinical notes by NLP using the LLM (MedRoBERTa.nl) that was fine-tuned as reported in Chapter 6. Although signs and symptoms derived from unstructured EHR clinical notes improved model performance beyond age and sex, their incremental value beyond structured EHR-based predictors for prediction of 30-day all-cause hospitalisation or mortality was limited (Dc-statistic +0.02, without substantial improvement of model calibration). In addition, decision curve analysis revealed limited net benefit of adding the NLP-derived predictors when compared to the model including structured EHR data only. This implies that signs and symptoms derived from unstructured EHR clinical notes do not add substantial predictive information in LRTI patients, provided that GPs adequately incorporate all relevant structured information available from patients' EHR.

Chapter 8 focusses on the temporary increased risk of acute cardiovascular events following an LRTI. In a population-based nested self-controlled case-series, the absolute excess cardiovascular risk associated with LRTI in adults presenting to the GP with LRTI was estimated for a broad range of cardiovascular events to quantify the population impact of LRTIs on cardiovascular events. The risk of major adverse cardiac and cerebrovascular events (a composite of MI, stroke, and TIA), venous thromboembolism (VTE, including PE and DVT) and new-onset AF (nAF) was particularly increased during the first week following LRTI diagnosis. Incidence rate ratios of these cardiovascular events within 90 days following LRTI diagnosis were 1.3 (95% CI 1.0 – 1.8), 6.8 (95% CI 4.4 – 10.4), and 2.9 (95% CI 2.3 – 3.7) per 1,000 LRTI patients, respectively. This culminates in approximately five to nine excess cardiovascular events attributable to LRTI per 1,000 LRTI patients. Incidence of major adverse cardiac and cerebrovascular events and nAF within 90 days following

LRTI diagnosis increased with age and was particularly high in patients with a history of heart failure, diabetes or hypertension, whereas incidence of VTE was most pronounced among younger patients. Given the large number of LRTI patients during seasonal outbreaks, this implies a need for targeted preventative interventions and patient-focussed campaigns to increase seasonal vaccine uptake and awareness of LRTI as trigger for cardiovascular events.

Discussion of this thesis

In the general discussion of this thesis (**Chapter 9**), I reflect upon the thesis' implications in the context of the GP's consultation room. I discuss how the 'final' model reported in Chapter 5 based on structured EHR data may be used to guide close monitoring strategies in LRTI patients in general practice while reflecting upon common challenges and pitfalls of implementing prediction models in clinical practice. I argue that well-developed prediction models do not necessarily lead to successful implementation. Therefore, apart from assessment of the model's impact on individual patient outcomes, GPs and other relevant stakeholders should inform clinically relevant risk thresholds and downstream intervention strategies to ensure successful adoption. Subsequently, I focus on the increased cardiovascular risk attributable to LRTIs and place these risk estimates in the context of traditional cardiovascular risk factors. This illustrates that the risks reported in Chapter 8 warrant, at a minimum, increased awareness among clinicians that LRTIs are an undervalued risk factor for acute cardiovascular events. In addition, I propose potential risk mitigating strategies aimed at early detection and prevention of cardiovascular events, including close monitoring strategies and short-course preventative treatment regimens. I emphasize that input from both patients and GPs on appropriate risk thresholds and acceptability and usability of potentially effective interventions are pivotal for the development and refinement of future risk mitigation strategies. Finally, using an illustrative example of a targeted case-finding strategy for nAF in LRTI patients, I discuss the clinical uncertainties and potential individual and societal adverse effects that should be explored and weighed against potential benefits while developing new cardiovascular risk mitigating interventions for LRTI patients in general practice. I argue that, especially in the case of nAF, targeted case-finding strategies should be tailored to a subgroup of LRTI patients at highest risk of clinically relevant AF to reduce overdiagnosis and (subsequent) overtreatment. In general, when developing interventions to mitigate cardiovascular risk in patients with LRTI, I highlight the importance of carefully balancing benefits with potential risks and adverse effects to adhere to the "first do no harm" principle, both from the individual patient's perspective and for society as a whole.

SAMENVATTING

De aanleiding voor dit proefschrift

Naar schatting bezoekt drie procent van de volwassenen jaarlijks de huisarts met klachten van een acute lage luchtweginfectie. Ongecompliceerde lage luchtweginfecties hebben over het algemeen een gunstig beloop, zeker bij afwezigheid van zogenoemde 'risicofactoren' voor een ernstig beloop. Complicaties, zoals ziekenhuisopname of zelfs overlijden, komen echter wel degelijk voor, met name bij patiënten met ernstigere lage luchtweginfecties zoals een longontsteking. Een voorspelmodel dat patiënten kan herkennen die, in het geval van een lage luchtweginfectie, een verhoogd risico hebben op zulke complicaties kan huisartsen ondersteunen bij hun klinische besluitvorming. Een dergelijk model is momenteel echter niet beschikbaar.

Complicaties ten gevolge van acute lage luchtweginfecties zijn breder dan enkel ziekenhuisopname of overlijden. Er is toenemend wetenschappelijk bewijs dat lage luchtweginfecties acute hart- en vaatziekten zoals een hartaanval, beroerte, trombose en boezemfibrilleren kunnen uitlokken. Hoewel uit eerder onderzoek blijkt dat het risico op dergelijke acute hart- en vaatziekten tijdens of kort na een lage luchtweginfectie tot vijfmaal hoger is, zijn er tot op heden geen absolute kansen op deze complicaties beschreven. Daarmee blijft de exacte omvang van het aantal acute hart- en vaatziekten gerelateerd aan lage luchtweginfecties in de populatie, en daarmee de extra ziektelast, onbekend.

De focus van dit proefschrift ligt op het verbeteren van de inschatting van het ziektebeloop – ofwel de prognose – van volwassenen die hun huisarts bezoeken met klachten van een acute lage luchtweginfectie. We richtten ons hierbij op een breed scala aan relevante uitkomsten, variërend van complicaties zoals ziekenhuisopname en overlijden tot acute hart- en vaatziekten. In de verschillende hoofdstukken van dit proefschrift werd gebruik gemaakt van onderzoeksgegevens verzameld uit elektronisch patiëntendossiers én moderne onderzoeksmethoden op het gebied van zowel prognostisch als etiologisch (oorzaak-gevolg) onderzoek. Daarnaast werd met behulp van technieken op het gebied van 'natuurlijke taalverwerking', onderzocht of ongestructureerde informatie zoals door de huisarts in het elektronische patiëntendossiers genoteerd, bijdraagt aan het beter voorspellen van de prognose van patiënten met een lage luchtweginfectie.

Bevindingen op basis van dit proefschrift

In **Hoofdstuk 2** worden de bevindingen van een systematisch literatuuronderzoek naar bestaande prognostische factoren en voorspelmodellen voor ziekenhuisopname en overlijden binnen 90 dagen na diagnose van een lage luchtweginfectie bij volwassenen in de huisartspraktijk beschreven. De volgende prognostische factoren bleken relevant: hogere leeftijd, geslacht, roken, een voorgeschiedenis met suikerziekte, beroerte, kanker of hartfalen, recente ziekenhuisopname, griepvaccinatiestatus, gebruik van systemische

glucocorticosteroïden, recent antibioticagebruik, een ademhalingsfrequentie van ≥25 per minuut en een klinische diagnose van een longontsteking. Deze factoren kunnen overwogen worden bij het ontwikkelen van een nieuw voorspelmodel of het verbeteren van een bestaand voorspelmodel. Bestaande voorspelmodellen zijn in hun huidige vorm niet geschikt voor gebruik in de dagelijkse huisartspraktijk vanwege de beperkte onderzoekskwaliteit en onvolledige beoordeling van modelprestatie.

Hoofdstuk 3 beschrijft de rationale voor en onderzoeksopzet van de ontwikkeling en externe validatie van twee voorspelmodellen op basis van gegevens verzameld uit het elektronisch patiëntendossier. Deze modellen richten zich op het voorspellen het individuele risico op (1) ziekenhuisopname of overlijden, ongeacht oorzaak, binnen 30 dagen en (2) acute hart- en vaatziekten binnen 90 dagen na diagnose van een lage luchtweginfectie in de huisartspraktijk. Voor het eerste model is deze vooraf uitgewerkte onderzoeksopzet ook gevolgd. De bevindingen hiervan worden beschreven in Hoofdstuk 5. Het lage percentage patiënten met een acute hart- en vaatziekte binnen 90 dagen na een lage luchtweginfectie belemmerde echter de ontwikkeling van een tweede klinisch relevant voorspelmodel. In plaats van het ontwikkelen van een tweede voorspelmodel, werd een etiologisch onderzoek uitgevoerd waarin het extra aantal acute hart- en vaatziekten ten gevolge van lage luchtweginfecties binnen onze onderzoekspopulatie werd geschat. De resultaten hiervan worden beschreven in Hoofdstuk 8.

In **Hoofstuk 4** wordt de volledigheid van rapportage van klachten, bevindingen van het lichamelijk onderzoek en metingen zoals genoteerd door huisartsen in het elektronisch patiëntendossier van volwassenen met een lage luchtweginfectie onderzocht, om zo inzicht te krijgen in de bruikbaarheid van deze gegevens bij het ontwikkelen van een voorspelmodel. In een aanzienlijk deel van de patiënten bleken deze potentieel relevante voorspellers niet gerapporteerd te worden. Daarnaast is de rapportage van deze gegevens waarschijnlijk selectief: abnormale waarden worden relatief vaak gerapporteerd. Dit onderschrijft de uitdagingen die gepaard gaan met het gebruik van gegevens over klachten, lichamelijk onderzoek en metingen zoals genoteerd door huisartsen in het elektronisch patiëntendossier bij het ontwikkelen van een voorspelmodel. Het is raadzaam om, voorafgaand aan het gebruik van dergelijke gegevens, de rapportage ervan goed te onderzoeken en technieken om met ontbrekende gegevens om te gaan hierop af te stemmen.

Hoofdstuk 5 beschrijft de stapsgewijze ontwikkeling en externe validatie van een voorspelmodel om het individuele risico op ziekenhuisopname of overlijden, ongeacht de onderliggende oorzaak, binnen 30 dagen na diagnose van een lage luchtweginfectie bij volwassenen ≥40 jaar in de huisartspraktijk te voorspellen. Voorspellers in dit model zijn gebaseerd op gecodeerde gegevens uit het elektronisch patiëntendossier en bestaan uit demografische gegevens (leeftijd en geslacht), cardiometabole ziekten (voorgeschiedenis

van suikerziekte, hartfalen, hartaanval, beroerte, TIA, longembolie, trombosebeen, boezemfibrilleren of perifeer vaatlijden), overige medische voorgeschiedenis (eerdere longontsteking, kanker die niet is ontstaan in de huid, chronische obstructieve longziekte, astma, dementie, recente ziekenhuisopname en griepvaccinatiestatus), medicatiegebruik (recent antibioticagebruik, afweeronderdrukkende medicatie, inhalatiemedicatie en antidepressiva) en een klinische diagnose van een longontsteking (versus bronchitis) door de huisarts. Het voorspelmodel werd ontwikkeld bij patiënten die zich met een lage luchtweginfectie presenteerden bij de huisarts uit de regio Utrecht vóór de COVID-19 pandemie (2016 – 2019, uitkomstpercentage: 7,8%) en werd vervolgens extern gevalideerd onder vergelijkbare patiënten uit de regio Amsterdam ná de COVID-19 pandemie (2022 – 2023, uitkomstpercentage: 11,4%). Deze externe validatie resulteerde in een onderscheidend vermogen, ofwel c-statistiek, van 0,71 (95% betrouwbaarheidsinterval (BI) 0,69 – 0,73) met overwegend goed gekalibreerde en dus nauwkeurige voorspellingen ('*intercept*' 0.28 (95% CI 0,20 – 0,36) en '*slope*' 0,95 (95% BI 0,85 – 1,05)). De door het model voorspelde kansen liggen tussen 1,5% en 51,3% (mediaan 6,9%). De toegevoegde voorspellende waarde van cardiometabole ziekte naast leeftijd en geslacht bleek laag (Dc-statistiek +0,02). Het ontwikkelde voorspelmodel, gebaseerd op gecodeerde gegevens uit het elektronisch patiëntendossier, lijkt veelbelovend ter ondersteuning van huisartsen bij het inschatten van de prognose van patiënten met een lage luchtweginfectie. Voordat het model gebruikt kan worden in de dagelijkse huisartspraktijk zal verder onderzoek moeten uitwijzen of – en hoe – het model uitkomsten van patiënten ook daadwerkelijk verbetert.

In **Hoofdstuk 6** wordt beschreven hoe twee Nederlandse taalmodellen (MedRoBERTa.nl en RobBERT) zijn aangepast en vervolgens getest om informatie over klachten en bevindingen van het lichamelijk onderzoek – geïdentificeerd als potentieel belangrijke voorspellers in Hoofdstuk 2 – met behulp van natuurlijke taalverwerking automatisch uit het elektronisch patiëntendossier van patiënten met een lage luchtweginfectie in de huisartspraktijk te verkrijgen. De modellen presteerden goed wanneer ze gebruikt werden voor directe classificatie – waarbij klachten en bevindingen werden aangemerkt als aanwezig, afwezig of niet gerapporteerd – met een respectievelijke gemiddelde F1-score van 0,74 (reikwijdte 0,56 – 0,87) en 0,69 (reikwijdte 0,46 – 0,86). Als trainingsmateriaal werden 1600 handmatig gescoorde consulten gebruikt. Modelprestatie bleek te variëren tussen de verschillende klachten en bevindingen en was afhankelijk van de mate waarin deze gerapporteerd werden en de verdeling over de verschillende classificatiegroepen. Desondanks laten deze resultaten zien dat taalmodellen bruikbaar zijn om automatisch informatie over klachten en bevindingen van het lichamelijk onderzoek uit Nederlandse elektronische patiëntdossiers te verkrijgen.

In **Hoofdstuk 7** wordt onderzocht in hoeverre door de huisarts genoteerde klachten en bevindingen van het lichamelijk onderzoek toegevoegde voorspellende waarde hebben

bovenop gecodeerde gegevens uit het elektronisch patiëntendossier door deze toe te voegen aan het voorspelmodel zoals beschreven in Hoofdstuk 5. De door de huisarts genoteerde klachten en bevindingen werden met behulp van natuurlijke taalverwerking door één van de taalmodellen uit Hoofstuk 6 (MedRoBERTa.nl) automatisch uit het elektronisch patiëntendossier verkregen. Hoewel deze klachten en bevindingen, wanneer toegevoegd aan een basismodel met enkel leeftijd en geslacht, voorspellende waarde lieten zien, bleken deze, wanneer toegevoegd aan het model met gecodeerde gegevens uit het elektronisch patiëntendossier, niet te leiden tot betere voorspellingen van ziekenhuisopname of overlijden binnen 30 dagen na diagnose van een lage luchtweginfectie in de huisartspraktijk (Dc-statistiek +0,02 met vergelijkbaar nauwkeurige voorspellingen). De winst in termen van terecht en onterecht geïdentificeerde hoog-risicopatiënten door het toevoegen van klachten en bevindingen bleek daarnaast, op basis van analyse van de besliscurve, beperkt ten opzichte van het model met enkel gecodeerde gegevens. Deze resultaten suggereren dat het meewegen van door de huisarts genoteerde klachten en bevindingen van het lichamelijk onderzoek, die door middel van natuurlijke taalverwerking automatisch werden verkregen uit het elektronisch patiëntendossier, niet leidt tot betere voorspellingen bij patiënten met een lage luchtweginfectie, onder de voorwaarde dat huisartsen alle relevante gecodeerde gegevens uit het elektronisch patiëntendossier goed in weten te zetten.

Hoofstuk 8 richt zich op het verhoogde risico op acute hart- en vaatziekten tijdens en kort na diagnose van een lage luchtweginfectie. Met een zogenoemde '*self-controlled case-series'* analyse onder volwassenen in de huisartspraktijk werd het extra aantal acute hart- en vaatziekten gerelateerd aan lage luchtweginfecties geschat, om zo de hieruit voortvloeiende extra ziektelast in de populatie in kaart te brengen. Een breed scala aan acute hart- en vaatziekten werd meegenomen in de analysen: ernstige cardiovasculaire en cerebrovasculaire aandoeningen (hartaanval, beroerte en TIA), veneuze trombose (longembolie en trombosebeen) en nieuw ontstaan boezemfibrilleren. Het risico op deze acute hart- en vaatziekten bleek met name verhoogd gedurende de eerste week na diagnose van een lage luchtweginfectie. Het aantal nieuwe acute hart- en vaatziekten binnen 90 dagen per 1000 volwassenen met een lage luchtweginfectie bleek 1.3 (95% BI 1.0 – 1.8) voor ernstige cardiovasculaire en cerebrovasculaire aandoeningen, 6.8 (95% BI 4.4 – 10.4) voor veneuze trombose en 2.9 (95% BI 2.3 – 3.7) voor nieuw ontstaan boezemfibrilleren. Dit resulteert in, naar schatting, vijf tot negen extra acute hart- en vaatziekten per 1000 patiënten met een lage luchtweginfectie die toe te schrijven zijn aan de lage luchtweginfectie. Het risico op ernstige cardiovasculaire en cerebrovasculaire aandoeningen en nieuw ontstaan boezemfibrilleren steeg met de leeftijd en was met name hoog bij patiënten met een voorgeschiedenis van hartfalen, suikerziekte of een hoge bloeddruk. Het risico op veneuze trombose was daarentegen groter onder jongere patiënten. Omdat lage luchtweginfecties, zeker in de winterperiode, vaak voorkomen, onderschrijven onze resultaten het belang van het ontwikkelen van gerichte preventieve (be)handelingen en

publiekelijke voorlichtingscampagnes om vaccinatiegraad van vaccinaties tegen belangrijke verwekkers van luchtweginfecties te bevorderen. Daarnaast geven onze bevindingen aanleiding tot extra bewustzijn onder huisartsen van het verhoogde risico op acute hart- en vaatziekten tijdens en kort na een lage luchtweginfectie.

Discussie en implicaties van dit proefschrift

In de algemene discussie (**Hoofdstuk 9**) schets ik de implicaties van dit proefschrift voor de zorg in de huisartspraktijk. Zo bespreek ik hoe het voorspelmodel uit Hoofdstuk 5, gebaseerd op gecodeerde gegevens uit het elektronisch patiëntendossier, gebruikt kan worden om patiënten met een lage luchtweginfectie en een verhoogd risico op ziekenhuisopname of overlijden te herkennen en vervolgens nauwkeuriger te monitoren. Daarnaast bespreek ik veelvoorkomende uitdagingen en valkuilen bij het implementeren van voorspelmodellen in de dagelijkse praktijk aan de hand van voorbeelden uit de literatuur, en concludeer hierbij dat goed ontwikkelde voorspelmodellen geen garantie bieden voor succesvolle implementatie. Om de kans hierop te vergroten is het daarom belangrijk om, naast te evalueren of – en hoe – gebruik van het door ons ontwikkelde voorspelmodel leidt tot verbetering van patiëntuitkomsten, samen met huisartsen en patiënten na te denken over relevante risicocategorieën en bijbehorende (behandel) adviezen. Vervolgens richt ik me op het verhoogde risico op verschillende acute hart- en vaatziekten tijdens en kort na een lage luchtweginfectie en vergelijk deze met traditionele risicofactoren voor hart- en vaatziekten. Hieruit concludeer ik dat de risico's zoals beschreven in Hoofdstuk 8 op zijn minst aanleiding geven tot extra bewustzijn onder huisartsen en andere zorgverleners van het – momenteel nog onderbelichte – verhoogde risico op acute hart- en vaatziekten tijdens en kort na een lage luchtweginfectie. Daarbij bespreek ik mogelijke strategieën gericht op het tijdig herkennen of voorkómen van acute hart- en vaatziekten bij patiënten met een lage luchtweginfectie, waaronder intensieve monitoring en kortdurende preventieve behandeling met bijvoorbeeld bloedverdunners. Wederom benadruk ik hierbij dat het belangrijk is om tijdens het uitwerken van deze strategieën actief met huisartsen en patiënten samen te werken. Dit vergroot de kansen op haalbare en praktisch goed toepasbare behandelstrategieën die aansluiten bij het perspectief en de behoeften van zowel huisarts als patiënt, te meer omdat de voor- en nadelen van nieuwe behandelingen zoals bloedverdunning goed afgewogen moeten worden. Tenslotte bespreek ik, aan de hand van een te ontwikkelen gerichte screeningsmethode om nieuw ontstaan boezemfibrilleren na een lage luchtweginfectie te herkennen, de klinische vraagstukken en mogelijke persoonlijke en maatschappelijke nadelige effecten van zulke strategieën. Deze dienen, samen met de mogelijke winst die deze strategieën kunnen opleveren, zorgvuldig afgewogen te worden. Ik benadruk hierbij, zeker in het geval van screening op nieuw ontstaan boezemfibrilleren, dat deze strategieën zich moeten richten op patiënten met een lage luchtweginfectie die het hoogste risico op klinisch relevant boezemfibrilleren hebben om zo de kansen op ongewenste overdiagnose en overbehandeling te minimaliseren, zeker in de context van het fenomeen

'diagnose-expansie' zoals beschreven in een rapport van de Raad voor Volksgezondheid en Samenleving in 2025. Samengevat dienen voor- en nadelen van nieuwe behandelstrategieën, gericht op het tijdig herkennen of voorkómen van acute hart- en vaatziekten tijdens of kort na een lage luchtweginfectie, zorgvuldig afgewogen te worden om ons, als medici en onderzoekers, zo aan het principe 'ten eerste, niet schaden' te houden, zowel vanuit individueel patiëntperspectief als voor de maatschappij als geheel.

DANKWOORD

De afgelopen jaren heb ik veel geleerd over predictie, over het zo goed mogelijk proberen te voorspellen van 'de toekomst'. Als ik terugblik op mijn promotietraject ligt het belangrijkste inzicht op het gebied van predictie voor mij echter buiten het onderzoeksveld. De belangrijkste voorspeller voor, bijvoorbeeld, een succesvol promotietraject, een opleiding tot huisarts en zoveel andere 'prestaties' is volgens mij de plek waar je wieg staat. Ik ben er zeker van dat er kinderen opgroeien in gebieden waar het woord predictiemodel geen betekenis heeft die in potentie betere wetenschappers en huisartsen zouden zijn dan ik ooit zou kunnen zijn, en ben me ervan bewust hoe ongelijk die kansen worden verdeeld. Mij past daarom bescheidenheid. Temeer omdat ik naast deze bepalende uitgangspositie vervolgens op weg geholpen ben door de tweede belangrijke voorspeller: een warme en stabiele basis, hechte familie, hele fijne vrienden, een promotieteam waar ik altijd op kon bouwen en de grootste levensgenieter en positieveling aan mijn zijde.

Zo hebben velen, ieder op zijn of haar eigen manier, aan dit proefschrift bijgedragen en ervoor gezorgd dat de afgelopen jaren, waarin de dit proefschrift tot stand is gekomen, als één groot feest hebben gevoeld. Daar wil ik jullie allemaal voor bedanken.

Allereerst wil ik uiteraard mijn promotieteam, prof. dr. Roderick Venekamp, prof. dr. Frans Rutten en dr. Tamara Platteel, bedanken voor hun steun en vertrouwen in mij als persoon en onderzoeker.

Beste **Roderick**, in 2019 deed ik onder jouw vleugels mijn wetenschappelijke stage in het laatste jaar van mijn Geneeskunde opleiding. Ik heb vanaf het begin jouw vertrouwen gevoeld om mijn weg in de academische wereld te vinden. Ik bewonder je vermogen om snelle analyses te maken en oplossingen te bedenken. Je ziet overal mogelijkheden, durft groot te denken en weet daarbij altijd anderen enthousiast te maken voor je projecten. Geen verrassing dus dat je inmiddels benoemd bent tot hoogleraar. Je gaf me daarnaast ook altijd een gevoel van gelijkwaardigheid tijdens inhoudelijke discussies. Dank voor je vertrouwen en steun in mijn ontwikkeling als onderzoeker.

Beste **Frans**, als enige promotor gedurende mijn onderzoekstijd gaf je de ruimte aan anderen om de lijnen uit te zetten, maar was je er altijd op de momenten dat het nodig was. Ik waardeer je betrokkenheid, eerlijke mening en je gave om buiten de gebaande paden te denken. Ik heb warme herinneringen aan onze gesprekken over reizen en bergvakanties in de gangen (of halverwege de trap) van het Julius Centrum. Ook waardeer ik je nimmer aflatende pogingen om het onderwerp hartfalen in de marges van mijn papers te schrijven tijdens revisierondes. En terecht, gezien de complexiteit van het ziektebeeld. Omdat je mijn dankwoord niet kon reviseren doe ik het nu dus voor je.

Beste **Tamara**, wat was het fijn om met jou samen te werken. Ik waardeer je vriendelijke directheid, je humor en je enthousiasme. Je frisse blik gaf steeds weer inzichten in hoe mijn projecten te vertalen naar de praktijk van de huisarts, en ik wist altijd precies wat ik aan je had. Daarnaast had je ook aandacht voor de dingen die mij buiten het werk bezighielden. Zo konden we overlegmomenten moeiteloos volpraten als er inhoudelijk weinig te bespreken was.

Ook wil ik jullie, Roderick, Frans en Tamara, nogmaals bedanken voor de vrijheid die jullie mij gegeven hebben om in de zomer van 2023 samen met Maïté onze droom na te jagen en in twee maanden de Pyreneeën te doorkruisen op de *Haute Randonnée Pyrénéenne*. Een ervaring voor het leven.

De leden van de **beoordelingscommissie**, prof. dr. J. Bont, prof. dr. R.A.M.J. Damoiseaux, prof. dr. M.H. Emmelot-Vonk, prof. dr. H.M. den Ruijter en prof. dr. E.W. Steyerberg, wil ik bedanken voor het lezen en beoordelen van mijn proefschrift. Ook wil ik prof. dr. N.D. Scherpbier en dr. M.E.W. Hemels bedanken voor hun deelname aan de oppositie tijdens mijn verdediging.

Naast mijn promotieteam hebben verschillende collega's een belangrijke rol gespeeld bij de verschillende projecten beschreven in dit proefschrift. Jullie hebben er samen met mijn promotieteam voor gezorgd dat ik ontzettend trots ben op de brede inhoud en kwaliteit ervan.

Beste prof. dr. Geersing, beste **Geert-Jan**, ondanks dat je formeel geen onderdeel was van mijn promotieteam ben je vanaf het begin altijd actief betrokken geweest. Met je snelle analyses kon je geregeld mijn projecten aanscherpen met het oog op de praktijk, en je daagde mij tijdens onze maandelijkse teamoverleggen uit om mijn projecten alvast te 'verdedigen'. Dank voor je betrokkenheid, positiviteit en stimulerende woorden.

Beste dr. van Smeden, Beste **Maarten**, dank voor je geduld om mij, niet gehinderd door enige kennis van predictieonderzoek, vanaf de start van dit traject bij te brengen hoe je goed predictieonderzoek doet en wat je vooral niét moet doen. Dit alles altijd met een stellige mening, duidelijke argumenten en veel humor. Een betere 'predictie opvoeding' kun je niet krijgen. Ik ben trots op wat we aan resultaten op het gebied van predictie in dit proefschrift hebben neergezet. Dit was nooit mogelijk geweest zonder jou.

Beste dr. Leeuwenberg, beste **Tuur**, ik had nooit gedacht dat ik tijdens dit promotietraject in de wereld van de taalmodellen zou belanden. Gelukkig hebben we ruimte genomen om nieuwe (ongeplande) wegen in te slaan, waaronder het verkennen van de waarde van notities in het elektronisch patiëntendossier voor predictieonderzoek. Onze samenwerking is hét bewijs dat het Julius Centrum een plek is waar klinisch en

methodologisch onderzoek elkaar direct kunnen versterken. Dank voor je betrokkenheid en enthousiasme, en ook veel dank aan **Isa** voor de fijne samenwerking.

Beste prof. dr. Bruijning-Verhagen, beste **Patricia**, toen we in Hoofdstuk 8 van dit proefschrift een voor mij geheel nieuwe onderzoeksopzet bedachten was jij enthousiast om ons hierbij van het nodige advies voorzien. Je legt de lat terecht hoog en mede dankzij jouw betrokkenheid ben ik ontzettend trots op de studie die we hebben neergezet.

Je kunt nog zo'n sterk team aan onderzoekers hebben, maar zonder data is er geen onderzoek. Allereerst wil ik daarom de deelnemende patiënten, huisartspraktijken en collega's van het **Julius Huisartsen Netwerk (JHN)** uit de regio Utrecht bedanken voor hun bijdrage aan het verzamelen van de benodigde gegevens voor mijn onderzoeken. Zonder jullie zou onderzoek op deze schaal niet mogelijk zijn, en jullie dragen hiermee bij aan onderzoek dat zich richt op directe verbetering van de zorg in de huisartspraktijk. Mijn speciale dank gaat uit naar de onvermoeibare **Marloes van Beurden**, datamanager bij JHN, voor haar harde werk om de benodigde gegevens voor mijn onderzoeken uit de JHN-database te verkrijgen. Ook wil ik de collega's, deelnemende praktijken en patiënten van het **Academisch Netwerk Huisartsgeneeskunde Amsterdam UMC (ANHA)** bedanken voor hun bijdrage.

Beste prof. dr. Schoonhoven en dr. Elias, beste **Lisette en Sjoerd**, jullie waren als onafhankelijk mentoren betrokken gedurende mijn gehele promotietraject. Hoewel de jaarlijkse voortgangsbesprekingen vooral een fijn bijpraat moment waren, voelde ik het vertrouwen dat ik ook bij jullie terecht kon als ik ergens tegenaan zou lopen. Dank voor jullie betrokkenheid.

Van alle collega's heb ik de afgelopen jaren de meeste uren doorgebracht met een fantastische groep **collega promovendi** bij de Huisartsgeneeskunde. Hiske, Joline, Anique, Mathé, Florien, Carlijn, Saskia, Anne, Lisa, Sophie, Fenja, Thomas, Eva, Josi, Lucien, Rianne, Hannah, Lion, Veerle, Gerdine, Pauline, Esmée, Michelle en Karin. Dankzij jullie was mijn promotietijd één groot feest. Ik ben trots op de mooie club HAG-promovendi die wij door de jaren heen gevormd hebben. Een 'writing & kiting retreat' (waar niet is gekite en eerlijk gezegd ook nauwelijks iets geschreven, toch Joline en Anique?), lange tennis en squash pauzes bij Olympos, congres in wijnstad Würzburg, borrels in Santorini en buiten werktijd, kantoorhumor en een daadwerkelijke schrijfretraite in een hutje op de hei. Dat ga ik allemaal missen. Ik hoop jullie, als collega huisarts of onderzoeker, of gewoon buiten werk om, nog veel tegen te komen.

Ook mijn collega's binnen de **HAG infectieziekten onderzoeksgroep** en alle andere collega's binnen de **afdeling Huisartsgeneeskunde & Verplegingswetenschappen** wil ik bedanken voor de fijne samenwerking.

Freek, Geertje, Aisha en andere collega's van **Medisch Centrum de Fonteijn** in Culemborg: bedankt voor een fantastisch eerste opleidingsjaar in jullie praktijk. Een bevestiging dat het huisartsenvak in al zijn veelzijdigheid bij me past. In het bijzonder wil ik jou, **Freek**, bedanken voor je vertrouwen, enthousiasme en vele taxi-ritjes naar de praktijk en de HAP in Tiel. Je bent een fantastische opleider.

De afgelopen jaren heb ik zoveel meer meegemaakt dan het schrijven van dit proefschrift, en er zijn velen die ik wil bedanken voor de manier waarop zij kleur aan mijn leven geven. Een hoop van jullie zullen misschien nauwelijks kunnen vertellen waar mijn promotietraject over ging – en dat is misschien maar goed ook, want er is zoveel meer in dit leven.

Thomas, **Jaap** en **Jumaa**, maatjes uit Breda al voelen we ons misschien het meest thuis op Sicilië: altijd een feestje om elkaar weer te zien. **Ceder** en **Tijn**, wie had gedacht dat die broekies in het Diakonessenhuis alle drie zouden promoveren? Het is genieten om met jullie, Jels en Jasper te zijn; dat voelt altijd als een thuiswedstrijd. **Daan** en **Julian**, als clubgenootjes aan het studentenleven begonnen en samen inmiddels volwassen. Met jullie kan ik áltijd lachen. De mannen van de 'Bergkriebels', **Kas**, **Mark** en **Rick**, de afgelopen jaren heb ik met jullie en Maïté waarschijnlijk de meeste grenzen verlegd. Van beginnende alpinisten tot onze eerste 4000'er. Er ligt nog een hoop avontuur op ons te wachten. Dear **Matteo** – my Sicilian brother – and wonderful family, I am beyond grateful that our friendship has survived the years and distance, and to know that I have a home in Sicily. Ook de vrouwen (en mannen) van de **KDS** en **Viatores** wil ik bedanken. Ik ben de afgelopen 10 jaar bijzonder aan jullie en de mannen gehecht geraakt en kijk terug op heel veel mooie momenten samen.

Ik ben een familiemens, en dat is natuurlijk niet zonder reden. Lieve **familie**, opa Piet, oma Gemma, oma Ans, ooms en tantes van beide kanten, neefjes en nichtjes: jullie zijn het warme bad waarin ik ben opgegroeid. Opa Huub, jouw naam draag ik met trots.

Lieve **Pierre**, **Annelies**, **Margaux**, **Mark** en **Poé**, wat ben ik blij om onderdeel van jullie familie te zijn. Een studie Geneeskunde verbleekt bij wat ik allemaal bij heb moeten leren over de Franse cultuur in de afgelopen 10 jaar, maar daar hebben jullie mij ontelbare mooie herinneringen voor teruggegeven. Jullie voelen als thuis. En wat betreft dit proefschrift: die schrijfretraites onder de Zuid-Franse zon hebben dit proefschrift altijd weer verder geholpen. Lieve **Juul**, dank voor alle fijne momenten samen op de Loeff, waar we allebei aan ons promotietraject begonnen.

Lieve **Maud** en **Gilles**, mijn paranimfen, wat ben ik trots dat wij hier vandaag met zijn drieën staan. Jullie zijn mijn beste maatjes, ik kan op jullie bouwen en vertrouwen – en eindeloos met jullie lachen. En ik besef me maar al te goed hoe uniek dat is. Lieve **Lotje**,

wat is het fijn dat jij onderdeel bent van onze familie. Ik bewonder je warmte en aandacht, en hoe je ook altijd je mannetje staat.

Lieve **papa** en **mama**, ik weet niet of woorden op papier recht kunnen doen aan wat ik jullie zou willen zeggen. Jullie waren, en zijn nog steeds, de veilige, warme basis waarin ik me heb kunnen ontwikkelen tot wie ik ben. En jullie zijn op zoveel manieren een voorbeeld voor me. Bedankt voor jullie steun en vertrouwen, en voor de inmiddels ontelbare fijne momenten samen. En pap, wat is het bijzonder dat jij de prachtige voorkant van dit proefschrift hebt gemaakt. Hierin combineerde je het onderwerp van mijn proefschrift met onze gedeelde passie voor de bergen.

Lieve **Maïté**, op de toppen van de Alpen of een ochtendduik in de Lek: wij maken overal geluk. Ik geniet van elke dag met jou, je bent mijn thuis. Je daagt me uit te blijven ontdekken en laat me meeliften op de lichte manier waarmee je je door het leven beweegt. Je inspireert me en houdt me met beide benen op de grond. Ik ben trots op jou en op wie wij samen zijn, en ik kan niet wachten op de rest van ons leven samen.

ABOUT THE AUTHOR

Merijn Huub Rijk was born on September 25th, 1993 in Breda. After completing his secondary education at 'De Nassau', he started his health care career by working in a nursing home. Subsequently, he volunteered as an English teacher in a remote village near Kep, Cambodia and travelled across Asia.

In 2013, Merijn started his medical training at the University Medical Center Utrecht. His interest in combining patient care with research grew during two internships within the Primary Care Infectious Diseases research group of the Julius Center for Health Sciences and Primary Care. After graduation, he worked as a resident in paediatrics at Ziekenhuis Gelderse Vallei in Ede. In 2022, he started a trajectory combining his vocational training as general practitioner with a PhD position under supervision of prof. dr. R.P. Venekamp, prof. dr. F.H. Rutten, and dr. T.N. Platteel. During this trajectory, in which he focussed on the prognosis of lower respiratory tract infections in general practice and the role of cardiovascular disease, he completed a postgraduate master in Epidemiology at Utrecht University.

Merijn lives in idyllic Wijk bij Duurstede together with his wife Maïté and their orange cat Tuca. In his spare time, he is most happy when surrounded by mountains, either climbing alpine summits or wild camping in remote valleys.