Diagnosing pneumonia in primary care

implementation of C-reactive protein point-of-care testing in daily practice

Margaretha Catharina Minnaard

Colophon

Margaretha Catharina Minnaard
Diagnosing pneumonia in primary care,
implementation of C-reactive protein point-of-care testing in daily practice,
164 pages.

Dissertation, Utrecht University, Utrecht, the Netherlands. (2016)

Graphic design cover and inside: Rachel van Esschoten, DivingDuck Design (www.divingduckdesing.nl)
Printed by: Gildeprint Drukkerijen, Enschede (www.gildeprint.nl)

ISBN: 978-90-393-6473-4

Diagnosing Pneumonia In Primary Care

implementation of C-reactive protein point-of-care testing in daily practice

Diagnostiek Van Pneumonie In De Eerste Lijn

implementatie van C-reactive protein sneltesten in de huisartsenpraktijk

(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. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties in het openbaar te verdedigen
op

donderdag 21 januari 2016
des ochtends te 10.30 uur
door

Margaretha Catharina Minnaard

geboren op 18 maart 1979 te Leiden

Promotoren: Prof. dr. N.J. de Wit
Prof. dr. T.J.M. Verheij

Copromotoren: Dr. A.C. van de Pol
Dr. B.D.L. Broekhuizen

The studies described in this thesis were financially supported by Saltro, Diagnostic Center for Primary Care, Utrecht and STAR-MDC, Medical Diagnostic Center, Rotterdam.

The printing of this thesis was financially supported by SBOH, employer of GP trainees in the Netherlands.

Thesis committee: Prof. dr. R.A.M.J. Damoiseaux
Prof. dr. R.J.P.M. Scholten
Prof. dr. F.G. Schellevis
Prof. dr. P. Little
Prof. dr. M.J.M. Bonten

Contents

Chapter 1

Introduction

Lower Respiratory Tract Infections

Cough is a very common symptom in the community, and lower respiratory tract infections (LRTIs) occur very frequently in all seasons. In the United Kingdom the reported incidence is 54/1000 annually among the "previously well" adults (defined as patients without underlying disease, e.g. asthma, chronic obstructive pulmonary disease, heart disease and diabetes).[1] In the Netherlands the reported incidence of patients presenting with acute cough (lasting 3 weeks or less) is 29.4/1000, of community acquired pneumonia (CAP) 13.3/1000 and of acute bronchitis 24.4/1000.[2]

Various definitions for LRTI exist. In most primary care guidelines LRTI is defined as an acute illness, present for a maximum of 3 weeks caused by inflammation in the respiratory tract distally from the larynx, including the trachea-bronchial tree and the alveoli. LRTI typically presents with cough as the main symptom accompanied by at least one other (LRTI) symptom such as sputum production, dyspnoea, wheeze or chest pain, without structural airway disease such as COPD or astma.[3] This pragmatic definition is based on the anatomy of the airways in combination with the clinical presentation. Alternative LRTI definitions are based on causative pathogens.[3] Most commonly identified (respiratory) pathogens in primary care patients with abnormal pulmonary auscultation are *rhinovirus*, *influenza virus A*, *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Mycoplasma pneumoniae*.[4,5] However in a large proportion, ranging from 33 to 60% of coughing patients in primary care no causative pathogen can be identified.[5,6]

The main challenge for clinicians in patients presenting with symptoms of LRTI is to differentiate community acquired pneumonia (CAP) from self-limiting conditions like acute bronchitis. CAP is an inflammation of lung parenchyma with *Streptococcus pneumoniae* as most commonly identified bacterial pathogen,[5,7] which requires antibiotic treatment.[1,8–12] Acute bronchitis is an inflammation of the mucous membranes of the bronchus or bronchioles which is usually self-limiting. It is commonly caused by viral infections, and according to most guidelines prescribing antibiotics is not beneficial for the majority of patients with acute bronchitis. Despite its mostly self-limiting disease course, many general practitioners (GPs) still treat a vast proportion of their patients with LRTI with antibiotics.[13] A recent Dutch study on antibiotic prescribing for LRTI (including bronchitis) concluded that 50% of prescription could be considered as over-prescription, not in accordance with current guidelines.[13] High outpatient use of antibiotics increases antibiotic resistance, which is considered a global health care threat and societal burden.[14] Moreover unnecessary use of antibiotics enhances costs and unwanted side effects.[15,16] The primary goal in the diagnostic work-up of LRTI is therefore to differentiate serious conditions like CAP from self-limiting conditions and to prevent unnecessary prescribing of antibiotics.

Diagnostic work-up of patients symptoms of with LRTI in primary care

Ruling out CAP among the many patients consulting with LRTI is challenging for GPs. Routine use of chest radiography (CXR), which is considered the gold standard for diagnosing CAP,[3,17] in all patients presenting with LRTI is neither legitimate nor feasible in primary care. For adequate management of LRTI GPs primarily rely on history and physical examination, in this thesis further referred to as 'signs and symptoms'. Prediction models for CAP based on signs and symptoms have been proposed to support clinical assessment.[18–23] External validation, i.e. evaluation of these models in populations different from the ones in which they were developed,[24–26] showed however that prediction models using only signs and symptoms cannot adequately predict CAP. To improve diagnostic accuracy additional diagnostic tests are therefore needed.[22,23,27] Routine use of microbiological tests, such as culture analysis and gram stains, which are commonly applied in secondary care to discriminate between viral and bacterial LRTI are not recommended in primary care, as only in a minority of LRTI patients a causative organism is found.[3,5,6,22,28] Additional testing with markers of inflammation in blood is potentially useful, but the diagnostic contribution depends on the protein used. Procalcitonin for instance ,[29,30] did not provide additional diagnostic value for CAP in primary care.[23]

C-reactive protein

So far, C-reactive protein (CRP) is the only inflammation marker that, in addition to signs and symptoms, was demonstrated to support diagnosing CAP in primary care.[23] This acute phase protein was named after its capacity to precipitate the C-polysaccharide of *Streptococcus pneumoniae*. Plasma CRP is synthesized by hepatocytes. Serum concentrations rise above 5 mg/l after about 6 hours after a single stimulus, with a peak in production after 48 hours.[31] CRP was demonstrated to be a sensitive, though nonspecific, systemic marker of inflammation, infection and tissue damage.[32–34] CRP levels in patients with LRTI do not differentiate very clearly between a viral or bacterial origin of disease. Both bacterial[1,4,5,28] and viral infections[35] can cause high levels of CRP.[28, 33,36] On the other hand, low values of CRP can help to rule out relevant disease such as CAP, which is particularly useful in primary care.[23,28] Three systematic reviews reported on the diagnostic value of CRP in diagnosing CAP in primary care.[37-39] Two of these reviews evaluated the diagnostic value of CRP as a single test[38, 39], and not its added value to signs and symptoms. The authors of these reviews considered CRP not sufficiently sensitive and specific to discriminate between patients with and without CAP in primary care. Authors of a more recent systematic review concluded that measurement of CRP has (limited) added value in diagnosing CAP.[37] Recently three new large diagnostic studies in primary care have been published that were not integrated

in these reviews.[23,40,41] The reviews mentioned above were based on published data only and authors had thus to work with different combinations of signs and symptoms to evaluate the added value of CRP. This impeded valid comparison between studies and limited the recommendations about the added value of CRP. To overcome this problem, an individual data analysis of all diagnostic studies in primary care is needed to adequately determine the added value of CRP above a fixed combination of signs and symptoms in diagnostic discrimination of LRTI patients with and without.

Point-of-care testing and antibiotic stewardship

In recent years CRP measurement has been introduced in primary care to guide antibiotic stewardship.[42] A point-of-care (POC) test in (capillary) blood yields CRP results within a few minutes, and can thus be used during consultation. For that reason the POC CRP test is increasingly used in the diagnostic work-up of LRTI.[43,44] Several studies have shown that POC CRP testing is cost-effective,[45–47] and can reduce antibiotic prescribing.[42,48–56] In these randomized controlled trials GPs were explicitly instructed on test indications and interpretation of CRP results. The actual impact of POC CRP implementation in routine GP practice, where the test is used at the discretion of the GP without a strict protocol for monitoring and interpretation, will probably differ from the experimental setting.[57] In Scandinavian countries for example, widespread introduction of POC CRP without proper instructions led to excessive use of the test in mild respiratory tract infections, misinterpretation of test results and improper antibiotic prescribing.[52, 58,59] Currently most national and international guidelines provide guidance on indications for POC CRP testing and interpretation of test results in patients who present with acute cough.[3,60] POC CRP testing was introduced in Dutch primary care after incorporation in the guideline LRTI of the Dutch College of General Practice. So far, little is known about the compliance of GPs with these guidelines and the impact of its introduction on prescription of antibiotics for LRTI.

The guideline did not provide guidance on the best POC CRP test device to be used in primary care. Different POC CRP test devices are available for use in primary care, often provided by primary care laboratories with an accompanying quality assurance program.[61,62] However, studies directly comparing all available devices are scarce [63–71] and the extent to which variability of POC CRP test results translates into differences in diagnostic accuracy for CAP is unknown.

In summary

The following questions regarding implementation of POC CRP testing in patients in LRTI in primary care need to be answered:

1. What is the accurate added value of CRP in the diagnostic work-up of LRTI in primary care?
2. Which POC CRP test device has best test characteristics for use in primary care?
3. How does the use of POC CRP affect GPs antibiotic prescribing in patients presenting with symptoms of LRTI in daily primary care practice?

Aims/objectives of this thesis

1. To validate and compare the performance of diagnostic models based on signs and symptoms for the diagnosis of CAP in primary care and to provide an accurate and generalizable estimate of the added value of CRP test result above clinically relevant signs and symptoms.
2. To evaluate the analytical performance, agreement with the laboratory reference test, user-friendliness and the diagnostic accuracy of available POC CRP tests in diagnosing CAP in primary care.
3. To determine the effect of POC CRP testing on antibiotic prescribing and the extent of guideline adherence when implemented in routine primary care practice.

Outline

In chapter 2 and 3 we used individual patient data from eight diagnostic studies on adult outpatients with LRTI suspected of CAP for meta-analysis. In chapter 2 six different published prediction models, based on signs and symptoms, for diagnosing CAP were validated by assessing discrimination and calibration of these models in the individual patient data. In chapter 3 we evaluated the added diagnostic value of CRP in diagnosing CAP in individual patient data, based on the improvement in discrimination and risk classification.

In chapter 4 and 5 we determined the suitability of five different POC CRP tests for use in primary care. In chapter 4 we used anonymous blood samples with pre-defined CRP levels to examine the analytical performance and the agreement of five POC CRP tests with a laboratory reference standard analyser. Subsequently we evaluated the user-friendliness of these POC CRP devices. In chapter 5 we assessed (single test) accuracy measures and added diagnostic value, reporting discrimination of the POC CRP tests and the laboratory analyser in predicting CAP in outpatients presenting with acute cough or suspected LRTI symptoms.

In chapter 6 we determined the overall effect of POC CRP measurement on antibiotic prescribing on group level in primary care practices. Moreover we assessed how often the POC CRP test results altered the GPs' decision on antibiotic prescribing for individual patients and we evaluated guideline adherence for both testing and antibiotic prescribing. In chapter 7 we elaborated on selection of patients in the study described in chapter 6 and whether this selection potentially biased the association between POC CRP tests and antibiotic prescription.

In chapter 8 the results are summarized, methodological issues are considered and implications of these results for management of LRTI in primary care are discussed.

References

1. Macfarlane J, Holmes W, Gard P, et al. Prospective study of the incidence, aetiology and outcome of adult lower respiratory tract illness in the community. *Thorax*. 2001;56(2):109–114.
2. Nielen MMJ, Spronk I, Davids R,et al. Incidentie en prevalentie van gezondheidsproblemen in de Nederlandse huisartsenpraktijk in 2013. Uit: NIVEL Zorgregistraties eerste lijn URL: www.nivel.nl/node/4096.
3. Woodhead M, Blasi F, Ewig S, et al. Guidelines for the management of adult lower respiratory tract infections--full version. *Clin Microbiol Infect*. 2011;17 Suppl 6:E1–59.
4. Graffelman a W, Knuistingh Neven A, le Cessie S, et al. Pathogens involved in lower respiratory tract infections in general practice. *Br J Gen Pract*. 2004;54(498):15–19.
5. Holm A, Nexoe J, Bistrup L a, et al. Aetiology and prediction of pneumonia in lower respiratory tract infection in primary care. *Br J Gen Pract*. 2007;57(540):547–554.
6. Graffelman . W, Knuistingh Neven A, Le Cessie S, et al. A diagnostic rule for the aetiology of lower respiratory tract infections as guidance for antimicrobial treatment. *Br J Gen Pract*. 2004;54(498):20–24.
7. Torres A, Blasi F, Peetermans WE, et al. The aetiology and antibiotic management of community-acquired pneumonia in adults in Europe: a literature review. *Eur J Clin Microbiol Infect Dis*. 2014;33(7):1065–79.
8. Butler CC, Hood K, Verheij T, et al. Variation in antibiotic prescribing and its impact on recovery in patients with acute cough in primary care: prospective study in 13 countries. *BMJ*. 2009;338:b2242.
9. Mortensen EM, Coley CM, Singer DE, et al. Causes of death for patients with community-acquired pneumonia: results from the Pneumonia Patient Outcomes Research Team cohort study. *Arch Intern Med*. 2002;162(9):1059–64.
10. Niederman MS, Mandell LA, Anzueto A, et al. Guidelines for the management of adults with community-acquired pneumonia. Diagnosis, assessment of severity, antimicrobial therapy, and prevention. *Am J Respir Crit Care Med*. 2001;163(7):1730–54.
11. Meehan TP, Fine MJ, Krumholz HM, et al. Quality of care, process, and outcomes in elderly patients with pneumonia. *JAMA*. 1997;278(23):2080–4.
12. Smucny J, Fahey T, Becker L, Glazier R. Antibiotics for acute bronchitis. *Cochrane database Syst Rev*. 2004;(4):CD000245.
13. Dekker AR, Verheij TJ, van der Velden AW. Inappropriate antibiotic prescription for respiratory tract indications: most prominent in adult patients. *Fam Pract*. 2015;32(4):401–407.
14. Goossens H, Ferech M, Vander Stichele R, et al. Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet*. 2005;365(9459):579–87.
15. Bradley JS, Guidos R, Baragona S, et al. Anti-infective research and development--problems, challenges, and solutions. *Lancet Infect Dis*. 2007;7(1):68–78.
16. Tacconelli E, De Angelis G. Fighting antibiotic resistance all over Europe. *Expert Rev Anti Infect Ther*. 2010;8(7):761–3.

17. Metlay JP, Fine MJ. Testing strategies in the initial management of patients with community-acquired pneumonia. *Ann Intern Med*. 2003;138(2):109–18.

18. Diehr P, Wood RW, Bushyhead J, et al. Prediction of pneumonia in outpatients with acute cough--a statistical approach. *J Chronic Dis*. 1984;37(3):215–25.

19. Singal BM, Hedges JR, Radack KL. Decision rules and clinical prediction of pneumonia: evaluation of low-yield criteria. *Ann Emerg Med*. 1989;18(1):13–20.

20. Heckerling PS, Tape TG, Wigton RS, et al. Clinical prediction rule for pulmonary infiltrates. *Ann Intern Med*. 1990;113(9):664–70.

21. Melbye H, Straume B, Aasebø U, et al. Diagnosis of pneumonia in adults in general practice. Relative importance of typical symptoms and abnormal chest signs evaluated against a radiographic reference standard. *Scand J Prim Health Care*. 1992;10(3):226–33.

22. Hopstaken RM, Muris JW, Knottnerus JA, et al. Contributions of symptoms, signs, erythrocyte sedimentation rate, and C-reactive protein to a diagnosis of pneumonia in acute lower respiratory tract infection. *Br J Gen Pract*. 2003;53(490):358–364.

23. 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:f2450.

24. Steyerberg EW. Clinical Prediction Models - A Practical Approach to | Ewout Steyerberg | Springer. 2008. Available at: www.springer.com/us/book/9780387772431.

25. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128–38.

26. Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14:40.

27. Graffelman AW, le Cessie S, Knuistingh Neven A, et al. Can history and exam alone reliably predict pneumonia? *J Fam Pract*. 2007;56(6):465–70.

28. Hjortdahl P, Landaas S, Urdal P, et al. C-reactive protein: a new rapid assay for managing infectious disease in primary health care. *Scand J Prim Health Care*. 1991;9(1):3–10.

29. Briel M, Schuetz P, Mueller B, et al. Procalcitonin-guided antibiotic use vs a standard approach for acute respiratory tract infections in primary care. *Arch Intern Med*. 2008;168(18):2000–2007; discussion 2007–2008.

30. Christ-Crain M, Jaccard-Stolz D, Bingisser R, et al. Effect of procalcitonin-guided treatment on antibiotic use and outcome in lower respiratory tract infections: cluster randomised, single-blinded intervention trial. *Lancet*. 2004;363(9409):600–605.

31. Pepys MB, Hirschfield GM. C-reactive protein: a critical update. *J Clin Invest*. 2003;111(12):1805–12.

32. Pepys MB, Baltz ML. Acute phase proteins with special reference to C-reactive protein and related proteins (pentaxins) and serum amyloid A protein. *Adv Immunol*. 1983;34:141–212.

33. Clyne B, Olshaker JS. The C-reactive protein. *J Emerg Med*. 1999;17(6):1019–1025.

34. Volanakis J. Human C-reactive protein: expression, structure, and function. *Mol Immunol*. 2001;38:189–197.

35. Hopstaken RM, Stobberingh EE, Knottnerus J a., et al. Clinical items not helpful in differentiating viral from bacterial lower respiratory tract infections in general practice. *J Clin Epidemiol*. 2005;58(2):175–183.

36. Holmberg H, Bodin L, Jönsson I, et al. Rapid aetiological diagnosis of pneumonia based on routine laboratory features. *Scand J Infect Dis*. 1990;22(5):537–45.

37. Engel MF, Paling FP, Hoepelman , AI, et al. Evaluating the evidence for the implementation of C-reactive protein measurement in adult patients with suspected lower respiratory tract infection in primary care: a systematic review. *Fam Pract*. 2012;29(4):383–393.

38. van der Meer V, Knuisting Neven A, van den Broek PJ, et al. Diagnostic value of C reactive protein in infections of the lower respiratory tract: systematic review. *BMJ*. 2005;331(7507):26.

39. Falk G, Fahey T. C-reactive protein and community-acquired pneumonia in ambulatory care: Systematic review of diagnostic accuracy studies. *Fam Pract*. 2009;26(1):10–21.

40. Steurer J, Held U, Spaar A, et al. A decision aid to rule out pneumonia and reduce unnecessary prescriptions of antibiotics in primary care patients with cough and fever. *BMC Med*. 2011;9:56.

41. Rainer TH, Chan CPY, Leung MF, et al. Diagnostic utility of CRP to neopterin ratio in patients with acute respiratory tract infections. *J Infect*. 2009;58(2):123–30.

42. Cooke J, Butler C, Hopstaken R, et al. Narrative review of primary care point-of-care testing (POCT) and antibacterial use in respiratory tract infection (RTI). *BMJ open Respir Res*. 2015;2(1):e000086.

43. Price CP. Point of care testing. *BMJ*. 2001;322(7297):1285–8.

44. Kazmierczak SC. Point-of-care testing quality: some positives but also some negatives. *Clin Chem*. 2011;57(9):1219–20.

45. Cals JW, Ament AJ, Hood K, et al. C-reactive protein point of care testing and physician communication skills training for lower respiratory tract infections in general practice: Economic evaluation of a cluster randomized trial. *J Eval Clin Pract*. 2011;17(6):1059–1069.

46. Dahler-Eriksen BS, Lauritzen T, Lassen JF, et al. Near-patient test for C-Reactive protein in general practice: Assessment of clinical, organizational, and economic outcomes. *Clin Chem*. 1999;45(4):478–485.

47. Oppong R, Jit M, Smith RD, et al. Cost-effectiveness of point-of-care C-reactive protein testing to inform antibiotic prescribing decisions. *Br J Gen Pract*. 2013;63(612):465–471.

48. Cals JW, Butler CC, Hopstaken RM, et al. Effect of point of care testing for C reactive protein and training in communication skills on antibiotic use in lower respiratory tract infections: cluster randomised trial. *BMJ*. 2009;338:b1374.

49. Cals JW, Schot MJ, de Jong SA, et al. Point-of-care C-reactive protein testing and antibiotic prescribing for respiratory tract infections: a randomized controlled trial. *Ann Fam Med*. 2010;8(2):124–33.

50. Andreeva E, Melbye H. Usefulness of C-reactive protein testing in acute cough/respiratory tract infection: an open cluster-randomized clinical trial with C-reactive protein testing in the intervention group. *BMC Fam Pract*. 2014;15(1):80.
51. Kavanagh KE, O'Shea E, Halloran R, et al. A pilot study of the use of near-patient C-Reactive Protein testing in the treatment of adult respiratory tract infections in one Irish general practice. *BMC Fam Pract*. 2011;12(1):93.
52. Diederichsen HZ, Skamling M, Diederichsen A, et al. A randomized controlled trial of the use of CRP rapid test as a guide to treatment of respiratory infections in general practice. *Ugeskr Laeger*. 2001;163(27):3784–3787.
53. Aabenhus R, Jensen-Jens US, Jørgensen KJ, et al. Biomarkers as point-of-care tests to guide prescription of antibiotics in patients with acute respiratory infections in primary care. Cochrane Database Syst Rev. 2014 Nov 6;11:CD010130.
54. Llor C, Cots JM, López-Valcárcel BG, et al. Interventions to reduce antibiotic prescription for lower respiratory tract infections: Happy Audit study. *Eur Respir J*. 2012;40(2):436–441.
55. Little P, Stuart B, Francis N, et al. Effects of internet-based training on antibiotic prescribing rates for acute respiratory-tract infections: A multinational, cluster, randomised, factorial, controlled trial. *Lancet*. 2013;382(9899):1175–1182.
56. Huang Y, Chen R, Wu T, et al. Association between point-of-care CRP testing and antibiotic prescribing in respiratory tract infections: A systematic review and meta-analysis of primary care studies. *Br J Gen Pract*. 2013;63(616):787–794.
57. Williams BA. Perils of evidence-based medicine. *Perspect Biol Med*. 2010;53(1):106–20.
58. André M, Schwan A, Odenholt I. The use of CRP tests in patients with respiratory tract infections in primary care in Sweden can be questioned. *Scand J Infect Dis*. 2004;36(3):192–7.
59. Engström S, Mölstad S, Lindström K, et al. Excessive use of rapid tests in respiratory tract infections in Swedish primary health care. *Scand J Infect Dis*. 2004;36(3):213–218.
60. Verlee L, Verheij TJM, Hopstaken RM, et al. [Summary of NHG practice guideline "Acute cough"]. *Ned Tijdschr Geneeskd*. 2012;156:A4188.
61. Hopstaken RM , Kleinveld HA, Balen van JAM, et al. Richtlijn point of care testing (POCT) in de huisartsenzorg. www.nhg.org/actueel/nieuws/nieuwe-richtlijn-point-care-testing-de-huisartsenzorg.
62. ISO 22870:2006 Point-of-care testing (POCT) - Requirements for quality and competence. International Organization of Standards. 2006.
63. Dahler-Eriksen BS, Lassen JF, Petersen PH, et al. Evaluation of a near-patient test for C-reactive protein used in daily routine in primary healthcare by use of difference plots. *Clin Chem*. 1997;43(11):2064–2075.
64. Dinant GJ, Costongs R, Leclercq RM, et al. Reliability of C-reactive protein measurement in general practice in The Netherlands. *Scand J Clin Lab Invest*. 1994;54(2):113–117.
65. Esposito S, Tremolati E, Begliatti E, et al. Evaluation of a rapid bedside test for the quantitative determination of C-reactive protein. *Clin Chem Lab Med*. 2005;43(4):438–440.

66. Monteny M, Ten Brinke MH, Van Brakel J, et al. Point-of-care C-reactive protein testing in febrile children in general practice. *Clin Chem Lab Med*. 2006;44(12):1428–1432.

67. Seamark DA, Backhouse SN, Powell R. Field-testing and validation in a primary care setting of a point-of-care test for C-reactive protein. *Ann Clin Biochem*. 2003;40(Pt 2):178–80.

68. Urdal P, Borch SM, Landaas S, et al. Rapid immunometric measurement of C-reactive protein in whole blood. *Clin Chem*. 1992;38(4):580–584.

69. Zecca E, Barone G, Corsello M, et al. Reliability of two different bedside assays for C-reactive protein in newborn infants. *Clin Chem Lab Med*. 2009;47(9):1081–1084..

70. http://madox.org/horizon-scanning-reports/20110017/point-of-care-test-poct-for-c-reactive-protein-crp

71. http://www.skup.dk/flx/afproevninger/?mode=liste&category=12.


2

Chapter 2

External validation of prediction models for pneumonia in primary care patients with lower respiratory tract infection: an individual patient data meta-analysis

Alwin Schierenberg
Margaretha C. Minnaard
Rogier M. Hopstaken
Alma C. van de Pol
Berna D.L. Broekhuizen
Niek J. de Wit
Johannes B. Reitsma
Saskia F. van Vugt
Aleida W. Graffelman
Hasse Melbye
Timothy H. Rainer
Johann Steurer
Anette Holm
Ralph Gonzalez
Geert-Jan Dinant
Joris de A.H. Groot
Theo J.M. Verheij

In revision for PLOS ONE

2

Abstract

Background

Community acquired pneumonia (CAP) remains difficult to diagnose in primary care. Prediction models based on signs and symptoms (S&S) serve to minimize the diagnostic uncertainty. External validation of these models is essential before implementation into routine practice. In this study all published S&S models for prediction of CAP in primary care were externally validated in the individual patient data (IPD) of previously performed diagnostic studies.

Methods

S&S models for diagnosing CAP in adults presenting to primary care with lower respiratory tract infection and IPD for validation were identified through a systematical search. Six prediction models and IPD of eight diagnostic studies (N total=5308, prevalence CAP 12%) were included. Models were assessed on discrimination and calibration. Discrimination was measured using the pooled area under the curve (AUC) and delta AUC, representing the performance of an individual model relative to the average performance of all models. Calibration was measured using calibration plots.

Results

Prediction models by van Vugt *et al*. and Heckerling *et al*. demonstrated the highest pooled AUC of 0.79 (95% CI 0.74-0.85) and 0.72 (0.68-0.76), respectively. Other models by Diehr *et al*., Singal *et al*., Melbye *et al*., and Hopstaken *et al*. demonstrated pooled AUCs of 0.65 (0.61-0.68), 0.64 (0.61-0.67), 0.56 (0.49-0.63) and 0.53 (0.5-0.56), respectively. A similar ranking was present based on the delta AUCs of the models. Calibration demonstrated close agreement of observed and predicted probabilities in the models by van Vugt *et al*. and Singal *et al*., other models lacked such correspondence. The absence of predictors in the IPD on dataset level hampered a systematical comparison of model performance and could be a limitation to the study.

Conclusions

The model by van Vugt *et al*. demonstrated the highest discriminative accuracy coupled with reasonable to good calibration across the IPD of different study populations. This model is therefore the main candidate for primary care use.

Introduction

Community acquired pneumonia (CAP) is a major cause of death in developed countries[1,2] and requires clinical treatment, whereas other lower respiratory tract infections (LRTIs) such as acute bronchitis are self-limiting.[3] The accurate diagnosis of CAP by a general practitioner (GP) is therefore important, but challenging as the routine use of chest x-radiography (CXR) for all patients presenting with LRTI is not feasible. Consequently, GPs mainly rely on signs and symptoms (S&S) in the diagnosis of CAP. Prediction models based on S&S have been proposed to decrease diagnostic uncertainty and prevent improper prescription of antibiotics and accompanying bacterial resistance.[4–7] Before considering the use of a prediction model in daily clinical practice, it is essential that its performance is empirically evaluated in datasets that were not used in the model development.[8–10] Such a study, in which the discrimination and calibration[11] of a prediction model are evaluated in new patients, is referred to as external validation.[10,12] Discrimination is the ability of the model to differentiate between diseased and non-diseased patients, whilst calibration signifies the agreement between predicted and observed probability of disease.[12] Evaluation of clinical usefulness with regard to improving patients outcomes or changing GP behaviour are not part of external validation.[13] External validation is required to quantify optimism caused by model overfitting[14] or deficiencies in the statistical modelling during model development, such as incorrect handling of missing data or a small sample size. Validation is also important to assess the model's transportability to other sites with arguably similar patients.[9,12,15] External validation of newly developed prediction models is rarely performed and generally of poor quality[13], but a necessary step before use in clinical care. Therefore, this type of prediction model studies is receiving increasingly more attention and has a central role in the recently published reporting guideline for prediction research (TRIPOD statement[16]). A limited number of external validation studies on diagnostic models for CAP have been performed[17–19], but none included patient data of multiple study sites and recently developed models.[19] Therefore, a meta-analysis using individual patient data (IPD) from multiple studies was performed in order to extensively validate and compare the performance of all published S&S prediction models for the diagnosis of CAP in primary care.

Methods

Selection of published models

Models eligible for inclusion were logistic regression models including S&S for predicting the probability of CAP in primary care patients with acute cough or suspected LRTI. Prediction models were identified through the following strategy: (a) screening references of the ERS management guidelines for adults with LRTI[20]; (b) eligibility assessment of models included in previously published validation studies[17–19]; (c) system-

atically searching PubMed, EMBASE and the Cochrane Library (Appendix 2.1, reference date: August 2012, 21st). After the identification of all eligible models, experts in the field were asked to identify missing models.

Selection of IPD for validation of published models

IPD for model validation was identified through an electronic systematical search in PubMed, EMBASE and the Cochrane Library (Appendix 2.1). C-reactive protein (CRP), an inflammation marker, was incorporated in the search because of a subsequent study investigating its added value.[21] Prospective studies diagnosing CAP were included when containing patients who: (a) were at least 18 years old; (b) presented through self-referral in primary care, ambulatory care or at an emergency department with an acute or worsened cough (≤28 days of duration) or any other clinical presentation potentially caused by LRTI; (c) consulted for the first time for this disease episode; (d) were immunocompetent. Eligible studies should have recorded clinical S&S and verified the disease status by CXR.[22]

Methodological quality assessment of IPD

Two reviewers (AS, JG) independently assessed the characteristics and methodological quality of the included IPD using the QUADAS-2[23] in order to identify potential sources of bias and improve the interpretation of results (Appendix 2.2). IPD were compared to the original study report on the total number of patients and the frequencies of single variables for error checking. If necessary, authors were contacted for information on quality assessment criteria or when datasets showed unexpected missing or invalid values.

Missing data

Missing values in IPD were regarded as missing at random (MAR). Single imputation was performed on individual dataset level[24] when missing values per IPD dataset did not exceed 33%. Predictors were considered absent when missing values exceeded 33% or when a predictor was not recorded entirely. Models could not be validated in IPD datasets containing absent predictors. This implies that the number of analysed patients might differ between the models validated.

Statistical analysis

The performance of included prediction models was assessed by discrimination and calibration. Discrimination was quantified using the pooled Area Under the (ROC) Curve (AUC) and the delta AUC (dAUC). Pooled AUC was quantified by first calculating the AUC and 95% confidence interval (CI) for each model individually per IPD dataset, followed by combining the individual AUCs in a pooled AUC using inverse variance weighing.[25,26] This two-step approach ensures accurate estimation of the pooled AUC in account for potential heterogeneity in AUC estimates.[27] As the absolute value of discrimination may differ considerably between IPD datasets, model performance was

subsequently evaluated on a relative scale, using the dAUC. The dAUC represents the difference in discriminative performance between an individual model (AUC) and the average performance of all models (mean AUC) within an IPD dataset. Calibration of included prediction models was assessed across different risk groups in each individual dataset. Risk groups with a low (0-10%) predicted risk of CAP, an intermediate risk (10-30%) and a high risk (30-100%) were defined. Per risk group the average predicted probability was calculated and compared to the proportion of CAP (i.e. the observed prevalence of CAP) in this group of patients. To obtain reliable estimates, the average probabilities were only calculated when at least 5 subjects per risk group could be included. In the case both a model and its development dataset were included in this study; the IPD of such a study was excluded from the external validation process. Data were analysed with SPSS (v20 for Windows), R (v2.15) including the "RMS" and "ROCR" packages for R[28] and Microsoft Excel (2010 for Windows). A prospective study protocol was formulated, indicating the main study objectives of the IPD study and the general methods for the current external validation study.

The Institutional Review Board of the University Medical Center Utrecht was not consulted for this meta-analysis as the study used only anonymous data from previously performed studies for which both informed consent and ethical approval had already been obtained.

Results

Selection of models

After assessment of published studies validating S&S models[17–19] and the European Respiratory Society guideline[20], six CAP prediction models for primary care use were included.[18,19,29-32] No suitable additional models were identified, neither through our systematic search nor after inquiry with experts in the field. The prediction models included between three to six predictors, the most frequent being fever (in 5 models), crackles (in 4 models), coryza (in 3 models), cough, dyspnoea, diminished breath sounds and tachycardia (in 2 models). The predictors asthma, duration of illness, chest pain, diarrhoea, fever (symptom), myalgia, phlegm, sore throat, sweating and tachypnoea were all included in one model (Table 2.1 and Appendix 2.3).

Selection of IPD for validation of published models

Eighteen of the 3676 identified studies appeared eligible for inclusion. Authors of these eighteen studies were requested to provide additional information and original data. Six studies did not fit the inclusion criteria, one author did not respond to our request and three authors were unable to provide the original study data (Figure 2.1). Eventually, the IPD of eight studies (N=5308) were included.[17,19,30,31,33–36]

Table 2.1: included prediction models.
Overview of included prediction models to diagnose community acquired pneumonia in a primary care setting and their incorporated predictors.

	Total	Diehr *et al.*	Singal *et al.*	Heckerling *et al.*	Melbye *et al.*	Hopstaken *et al.*	van Vugt *et al.*
Total predictors in model		**6**	**3**	**5**	**6**	**3**	**6**
History							
Absence of asthma	1			•			
Duration of illness	1				•*		
Symptoms							
Chest pain	1				•		
Coryza (absence)	3	•			•		•
Cough (dry)	2		•			•	
Diarrhoea	1					•	
Dyspnoea	2				•		•
Fever	1				•*		
Myalgia	1	•					
Phlegm	1	•					
Sore throat	1				•		
Sweats (night)	1	•					
Signs							
Crackles	4		•	•	•		•
Diminished breath sounds	2			•			•
Fever	5	•	•	•		•	•
Tachycardia	2			•			•
Tachypnoea	1	•					

*• = predictor present, *combined predictor.*

Characteristics of IPD

Of the eight included studies, five included patients visiting a GP,[17,19,30,34,36] one included patients visiting a primary care out-of-hours service[31] and two studied self-referred patients to an emergency department[33,35] (Table 2.2). The study by van Vugt *et al*. contained 55% (N=2820) of all IPD. The mean age of all individual patients combined was 49 years (SD=18). The mean age of separate studies was lower in patients from Melbye *et al*. and Flanders *et al*., with a mean age of 33 (SD=14) and 40 (SD=16) years, respectively. In individual datasets the proportion of males varied between 40 and 50%.

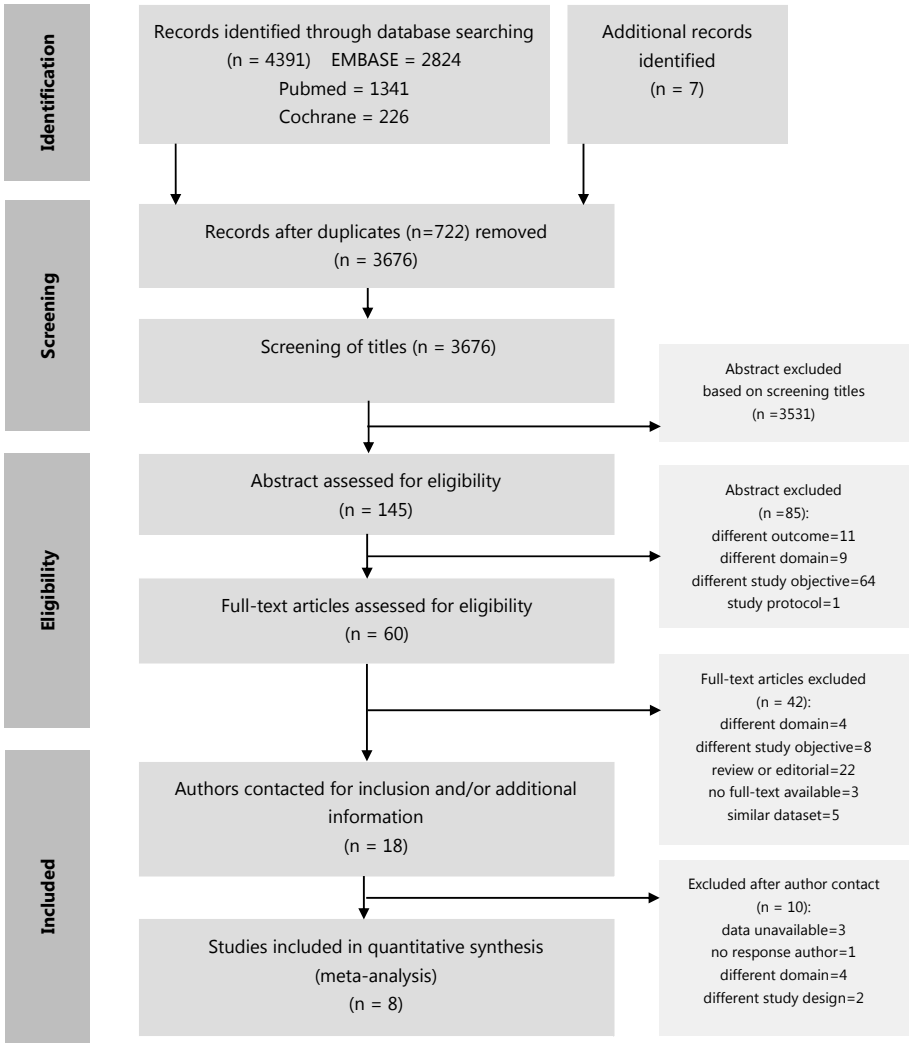


Figure 2.1: PRISMA flow diagram.
PRISMA flow diagram of the selection process of IPD used for external validation of prediction models.[37]

The prevalence of CAP ranged from 5% to 43%. In only one study providing IPD all predictors were present,[33] in all other studies proving IPD one or more predictors were not recorded. If predictor were recorded the highest percentage of missing values per predictor never exceeded 33% (max. 28%). One of the included IPD datasets had previously been imputed using hot-deck imputation.[33] Table 2.2 gives a detailed presentation of the baseline characteristics in all included IPD datasets.

Table 2.2: Baseline characteristics

Baseline characteristics of included individual patient datasets used in the external validation of prediction models for community acquired pneumonia in primary care setting.

	Validation dataset								
	Melbye *et al.*	Hopstaken *et al.*	Flanders *et al.*	Graffelman *et al.*	Holm *et al.*	Rainer *et al.*	Steurer *et al.*	van Vugt *et al.*	All datasets
Patient characteristics									
Setting	OHD	GP	ED/AC	GP	GP	ED	GP, ED	GP	-
Number of patients	402	243	168	129	364	561	621	2820	5308
CAP	5%	13%	12%	20%	13%	43%	21%	5%	12%
Age, mean (SD)	33 (14)	52 (16)	40 (16)	50 (14)	50 (16)	53 (22)	47 (16)	50 (17)	49 (18)
Gender, Male	41%	47%	41%	47%	49%	53%	50%1	40%	44%
Duration illness in days,mean (SD)	10 (14)	Categorized[2]	7 (5)	9 (6)	--	17 (9)	7 (10)	10 (10)	8,4 (10)
Smoker	56%	33%	11%	36%	45%	17%	29%	28%	30%
Asthma	10%	19%	11%	6%	8%	--	--	10%	10%
Symptoms									
Cough	91%	92%	100%	98%	98%	88%	97%	100%	97%
Chest pain (lateral)	53%	60%	40%	23%	64%	40%	29%	46%	45%
Coryza	80%	38%	69%	59%	--	50%	--	71%	67%
Diarrhoea	--	8%	14%	24%	--	9%	--	7%	8%
(Daily) Fever, subjective	31%	35%	59%	85%	42%	83%	56%	35%	47%
Dyspnoea	69%	77%	51%	76%	72%	56%	36%	57%	57%
Myalgia	54%	62%	55%	59%	--	50%	--	50%	52%
Sore throat	73%	39%	65%	39%	--	50%	--	--	55%
Phlegm	88%	55%	55%	79%	81%	77%	49%	79%	75%
(Night) Sweats	84%	61%	58%	--	--	42%	--	--	60%
Signs									
Crackles	11%	21%	9%	60%	--	--	20%	9%	573
Diminished breath sounds	5%	--	17%	12%	--	--	12%	13%	13%
Heart rate, p.m. (SD)	79 (13)	--	85 (19)	82 (11)	81 (15)	98 (18)	--	77 (12)	81 (15)
Respiratory rate, p.m. (SD)	--	Categorized[3]	18 (4)	21 (4)	19 (4)	19 (3)	17 (6)	17 (4)	18 (4)
Temperature, C° (SD)	37.3 (0.7)	37.5 (0.8)	37.3 (0.8)	37.9 (0.7)	37.4 (0.6)	37.8 (1.1)	37.4 (1)	36.7 (0.6)	37.1 (1)

OHD=Out of Hours Department, GP=General Practitioner, ED=Emergency Department, AC=Ambulatory Clinic, [1] Data from original publication, [2] Categorized as ≤2, 3-7, 8-28 days , [3] Categorized as >20 p.m., "--"= Variable missing

Table 2.3: Discriminative performance.
Discriminative performance of community acquired pneumonia prediction models per dataset, measured as Area Under the ROC Curve (AUC) and as pooled AUC in all suited individual patient data (IPD).

	Validation dataset										
Model	**Melbye *et al.***	**Hopstaken *et al.***	**Flanders *et al.***	**Graffelman *et al.***	**Holm *et al.***	**Rainer *et al.***	**Steurer *et al.***	**Van Vugt *et al.***	**Development AUC (95% CI)**	**Pooled AUC (95% CI)†**	**Patients in IPD/ development (N/N)**
van Vugt *et al.*	0.78	X	0.89	0.60	X	X	X	D	0.70 (0.65-0.75)	0.79 (0.74-0.85)	699/2820
Heckerling *et al.*	0.69	X	0.89	0.62	X	X	X	0.66	0.82 (0.78-0.86)	0.72 (0.68-0.76)	3519/1134
Diehr *et al.*	X	0.57	0.76	X	X	0.64	X	X	NA	0.65 (0.61-0.68)	972/474
Singal *et al.*	0.68	0.62	0.81	0.63	0.62	X	0.61	0.64	0.73 (0.69-0.77)	0.64 (0.61-0.67)	4747/255
Melbye *et al.*	D	0.57	0.62	0.49	X	X	X	X	0.75 (0.66-0.84)	0.56 (0.49-0.63)	540/402
Hopstaken *et al.*	X	D	0.58	0.61	X	0.52	X	0.56	0.70 (0.59-0.80*)	0.53 (0.50-0.56)	3678/243

*X = Model not validated in dataset due to missing predictors, D = Development dataset (AUCs shown under "Development"), NA= Not available (none reported in development study), * 95% CI not available in original study report (recalculated in original dataset), †AUC of Development dataset ("D") not included.*

Methodological quality assessment of IPD

In general the assessment of study quality of the included datasets raised little concern of bias (Appendix 2.2). Nonetheless, four studies showed a risk of bias and/or applicability concerns in the patient selection.[17,31,33,35] Two studies presented potential bias concerning flow and timing,[31,33] as the acquisition of the reference test was left up to the physician's judgment (partial verification), which may have induced misclassification of CAP. To adjust for potential misclassification one of these two study performed the reference standard in a 25% random sample (showing no additional cases of CAP).[31] Furthermore, in one IPD dataset the CXR results were missing; therefore the discharge diagnosis (primarily based on CXR results) was used in the meta-analysis to define CAP.[35] Moreover, this study reported a high prevalence of CAP (43%),[35] indicating a potential applicability concern in the patient selection for the purposes of this validation study. Two studies demonstrated negligible concerns in the study's flow and timing or patient selection.[17,19]

Performance of models in individual patient datasets

Each of the six included models could be externally validated in the IPD of at least three and up to seven datasets (Table 2.3). The model by Diehr *et al.* in three datasets

(N=972), Singal *et al.* in seven datasets (N=4747), Heckerling *et al.* in four datasets (N=3519), Melbye *et al.* in three datasets (N=540), Hopstaken *et al.* in four datasets (N=3678) and the model by van Vugt *et al*. in three datasets (N=699).

The model by van Vugt *et al.* demonstrated the highest pooled AUC of 0.79 (95% CI 0.74-0.85), compared to an AUC of 0.7 in the development study. The model by Heckerling *et al.* demonstrated a pooled AUC of 0.72 (95% CI 0.68-0.76, development AUC of 0.82), Diehr *et al.* of 0.65 (0.61-0.68, development not available), Singal *et al.* of 0.64 (0.61-0.67, development 0.73), Melbye *et al*. of 0.56 (0.49-0.63, development 0.75) and Hopstaken *et al*. of 0.53 (0.5-0.56, development 0.7). When evaluating the individual model performance relative to the average dataset performance, using the dAUC as measure, a similar ranking was demonstrated compared to the ranking based on the pooled AUC. The model by van Vugt *et al*. demonstrated a higher than average AUC in all datasets, with dAUCs ranging from 0.14 to 0.01 (Figure 2.2 and Appendix 2.4), and

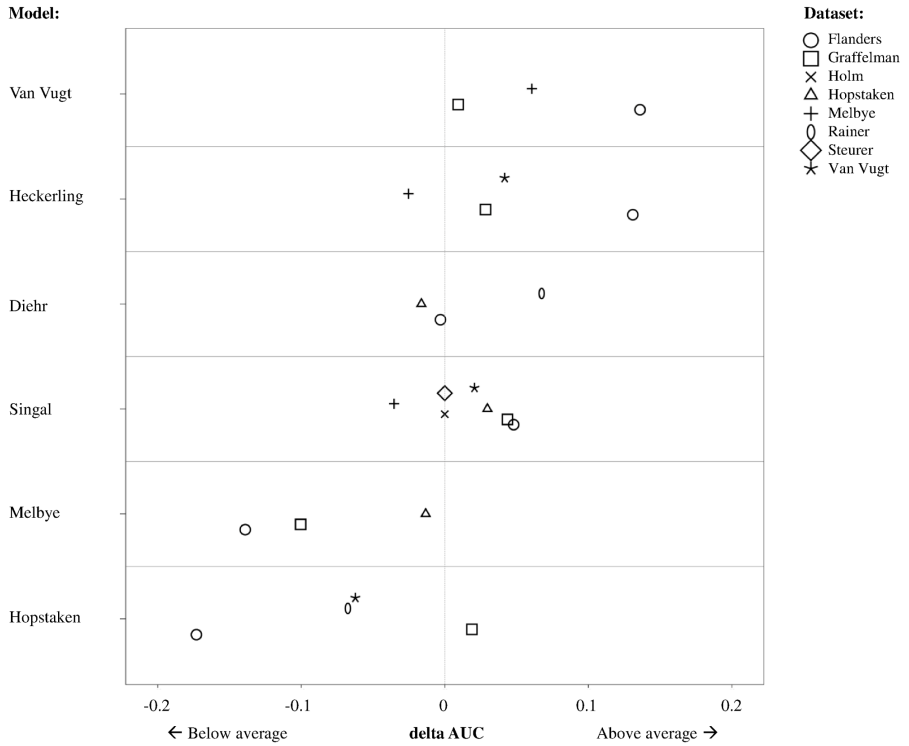
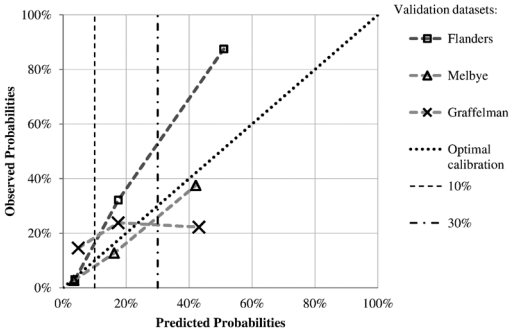
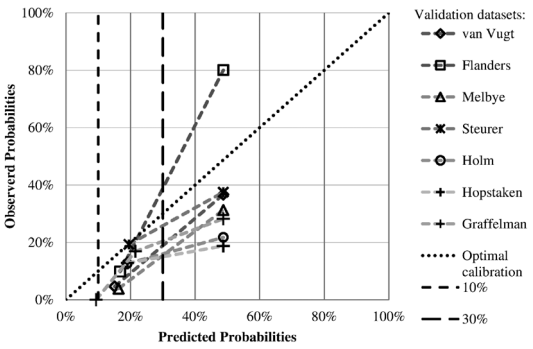


Figure 2.2: Graphic representation of model performance relative to dataset average AUC, measured as delta AUC.

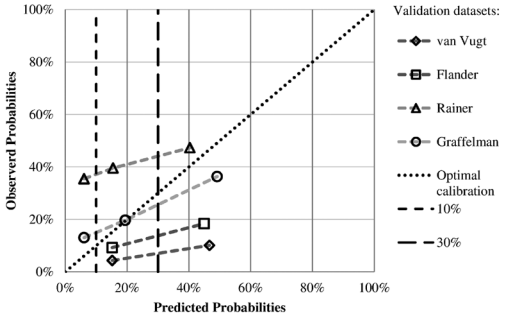
Each point represents the performance of an individual model relative to the average performance of all models per dataset (dAUC) and can be either above the average (better performance) of the dataset or below the average (worse performance). The plot is clustered per model. The dAUC is calculated as: individual model AUC - μAUC of dataset.



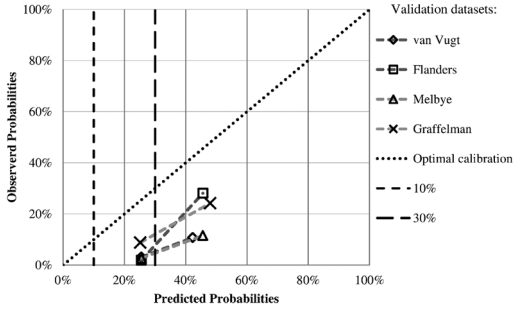
2.3A: Calibration plot of the model by van Vugt et al.



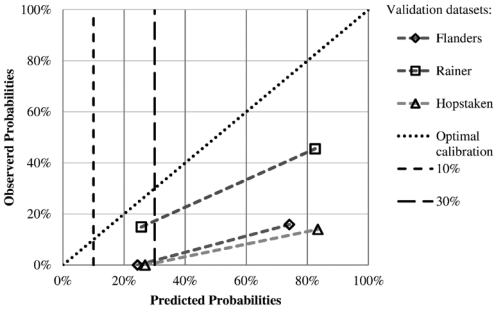
2.3B: Calibration plot of the model by Singal et al.



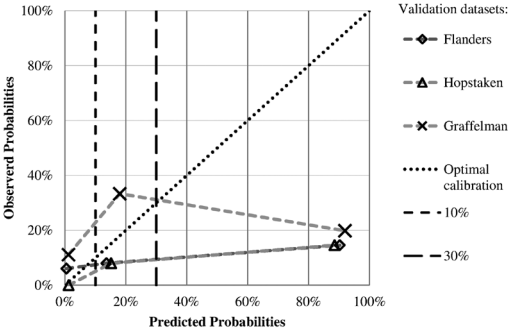
2.3C: Calibration plot of the model by Hopstaken et al.



2.3D: Calibration plot of the model by Heckerling et al.



2.3E: Calibration plot of the model by Diehr et al.



2.3F: Calibration plot of the model by Melbye et al.

Figure 2.3: Calibration plots.
Calibration plots of prediction models clustered per risk group with low (0-10%), intermediate (10-30%) and high (30-100%) predicted probabilities. Calibration results are presented for each validation dataset where the model could be validated. Plots show how well the predicted probabilities (x-axis) agree with observed probabilities (y-axis). For perfect agreement, the calibration curve falls on the ideal diagonal line (optimal calibration). Two vertical cut-off lines for 10% and 30% risk of CAP are depicted.

was followed by Heckerling *et al.* (0.13 to -0.03), Diehr *et al.* (0.07 to -0.02), Singal *et al.* (0.05 to -0.04), Melbye *et al.* (-0.01 to -0.14) and Hopstaken *et al.* (0.02 to -0.17).

The calibration plot of the model by van Vugt *et al.* demonstrated the closest agreement between the model's predictions and the observed prevalence of CAP (Figure 2.3a). The model by Singal *et al.* lacked the potential to assign patients to a low risk of CAP, but showed a rather uniform prediction pattern in the other risk groups, where in general the model slightly overestimated the predicted probabilities (Figure 2.3b). The model by Hopstaken *et al.* showed a linear relation between the predicted probabilities and prevalence of CAP in all datasets. However, this relation varied considerably, from consistent overestimation in one dataset and an underestimation in another (Figure 2.3c). The models by Heckerling *et al.* and Diehr *et al.* demonstrated consistent overestimation of the predicted probabilities and lacked the potential to assign patients to a low risk of CAP (Figures 2.3d and 2.3e, respectively). The model by Melbye *et al.* lacked a clear linear relation between the observed probabilities and prevalence of CAP (Figure 2.3f). Pooling of calibration results was not possible due to heterogeneity of results and, therefore, not further pursued.

Discussion

Main findings

This study assessed discrimination (pooled AUC and dAUC) and calibration of six previously published S&S models for diagnosing CAP in primary care in the IPD of eight diagnostic studies (N=5308).

The model by van Vugt *et al.*, demonstrated the highest pooled and relative discriminative performance, with a pooled AUC of 0.79 and dAUCs between 0.14 and 0.01. The model by Heckerling *et al.* followed with a pooled AUC of 0.73 and dAUCs between 0.13 and -0.03. The models by Diehr *et al.*, Singal *et al.* Melbye *et al.* and Hopstaken *et al.*, demonstrated lower average and relative discriminative performance, with respective pooled AUCs of 0.65, 0.64, 0.56 and 0.53, and marginally positive or negative dAUCs values. Calibration of the models by van Vugt *et al.* and Singal *et al.* was acceptable, demonstrating reasonable overall agreement in the predicted probabilities and presence of CAP, and allows for optimization with simple recalibration methods. Calibration of the remaining models showed signs of overfitting or varying degrees of systematic over- or underestimation of predicted probabilities between different datasets, impeding simple recalibration.

Interpretation of findings

It is common that performance of a prediction model decreases when validated in new patients. Such a decrease is typically caused by the difference in case-mix of, arguably similar, patients. However, when the decrease in performance is larger than expected other mechanisms could have caused overfitting of the model in the development

study, such as a (too) small development dataset or a too elaborate selection of candidate predictors.[14] Furthermore, in some cases the replacement of absent predictors in the external validation data may have led to lower discriminative performance of the model, e.g. 'dry cough' in the model of Hopstaken *et al*. was only measured in a single dataset[17] and therefore the predictor 'cough' was used. The model by van Vugt *et al*. showed a better discrimination in external validation compared to the development study. This somewhat unusual finding might be caused by the partial verification of the disease status in two of the included datasets.[31,33] In both datasets CXR acquirement was dependent on physician judgment, whereas patients not receiving a CXR were considered healthy. Consequently, clinical information (e.g. signs and symptoms) could have influenced the disease status and lead to an overestimation of the discriminative performance of a prediction model.[38] However, it is likely that all models would equally benefit from potential overestimation of the discriminative performance in these two datasets and also be of little impact, as most models could be validated in these datasets. Concern in performance differences would not have existed if all models would have been validated in all IPD. In our study such a comparison was not conceivable as in five of the included IPD dataset one or more required predictors were absent. To approach an equal comparison between models and minimize the performance differences we used the dAUC. Here both methods (pooled and dAUC) demonstrated similar results. Performance between models could also be affected by the inclusion criteria used in a study contributing IPD. For example, when patient are selected on the basis of specific clinical characteristics (e.g. fever) one might expect that the performance of models including such variables (predictors) will be negatively influenced in a validation study.[39] However, the good performance of some of the included models, when evaluated in a mixed IPD population including patients with various likelihoods of CAP, indicates that they can be used beyond the first step of the diagnostic process. In this study we performed a visual assessment of calibration in various clinically relevant risk groups. Per group it was assessed how the predicted risk of CAP compared to the true prevalence of CAP. In general, included models failed to assign extreme predictions (closer to 0 or 1), meaning it is challenging to completely rule out or prove the presence of CAP. Either such extreme predictions were not made at all by the model (e.g. for low risks <10%) or did not correspond well with the true prevalence of CAP (e.g. for higher predicted risks >30%). This phenomenon is can be expected when presenting patients are in general reasonably healthy and when studying a clinically heterogeneous disease, like CAP. Future research should focus on the recalibration of original models to ensure the accurate predictions in all types of patient populations, while preserving discrimination.[40] However, in models lacking consistency in calibration (e.g. by overfitting), simple recalibration methods may not suffice. Two of the included models cases (Diehr *et al*. and Melbye *et al*.) included no intercept. This might be an explanation for the poor calibration of these models. In subsequent investigations it is recommended to add an intercept to improve performance of these models. However, such amendments were beyond the scope of this review.

Finally, the diagnostic properties found in the present analysis may be lower, or arguably higher, when applied to settings where alternative reference standards for CAP than CXR are applied. Although, no complete consensus on a gold standard for CAP exists, CXR is the most commonly used reference standard. In this study all of the included studies providing a dataset utilized CXR as a reference standard, but non raised concern about the reference standard and in most studies the CXR was review by a second blinded radiologist to minimize inter-observer variability.[17,19,30,31,35,36]

Strengths and limitations

To our knowledge this IPD meta-analysis validated all primary care S&S prediction models for CAP in a large composite dataset of IPD of high quality diagnostic studies. Included models could be validated in at least three external data sources, providing reliable estimates of the pooled AUC. Nonetheless, it is important when comparing models to focus on results obtained within the same validation dataset, in a paired comparison using dAUCs, as the absolute value of discrimination differed between validation datasets. Calibration of models in multiple validation datasets is notoriously hard to quantify. Therefore, we created clinically relevant risk groups to detect potential weaknesses in calibration that can be translated to the clinical setting.

A potential limitation of this study was the use of alternative (definitions for) predictors when specific predictors from published models were missing (Appendix 2.3). However, we only used these alternative predictors when sufficiently appropriate or when they could be calculated with the help of other predictors. Moreover, we presume that these types of predictors (e.g. “sweats” for “night sweats”) are often used in a similar and interchangeable fashion in daily practice and are therefore comparable. Even when alternative predictors were considered, the performance evaluation of several models was hampered due to absence of predictors. This complicates straightforward comparison of these models and could have theoretically induced bias in model performance. However, by assessment of the discriminative performance according to two different methods, which incorporated a within model comparison (i.e. dAUC) this evaluation was arguably justified. Lastly, in our study the prevalence of CAP ranged between 5-43% in the included IPD datasets, which generally higher than the prevalence of 6%, previously found in a primary care setting.[41] The large variation in prevalence reflects both a variation in setting of included studies and a difference in the inclusion criteria applied in included studies. This may have led to the inclusion of IPD with a broad case-mix, ranging from patients with acute cough to suspected CAP. However, as the key purpose of an external validations study is to evaluate the performance of prediction models in other – but arguably comparable – patients, the heterogeneity in the IPD patient population due to differences in inclusion criteria does not interfere with the primary aim of our study.

Conclusions

Prediction models can be of value for GPs by discriminating between patients with and without CAP but they fail to assign very high or low risks. Of all published primary care S&S prediction models for CAP, the model by van Vugt *et al*. demonstrated the highest discriminative accuracy coupled with reasonable to good calibration in IPD of different study populations. This model is therefore the main candidate for use in primary care.

2

References

1. Mortensen EM, Coley CM, Singer DE, et al. Causes of death for patients with community-acquired pneumonia: results from the Pneumonia Patient Outcomes Research Team cohort study. *Arch Intern Med*. 2002;162(9):1059–64.
2. Niederman MS, Mandell LA, Anzueto A, et al. Guidelines for the management of adults with community-acquired pneumonia. Diagnosis, assessment of severity, antimicrobial therapy, and prevention. *Am J Respir Crit Care Med*. 2001;163(7):1730–54.
3. Smucny J, Fahey T, Becker L, et al. Antibiotics for acute bronchitis. *Cochrane database Syst Rev*. 2004;(4):CD000245.
4. Bradley JS, Guidos R, Baragona S, et al. Anti-infective research and development--problems, challenges, and solutions. *Lancet Infect Dis*. 2007;7:68–78.
5. Tacconelli E, De Angelis G. Fighting antibiotic resistance all over Europe. *Expert Rev Anti Infect Ther*. 2010;8(7):761–3.
6. Cals JWL, de Bock L, Beckers P-JHW, et al. Enhanced communication skills and C-reactive protein point-of-care testing for respiratory tract infection: 3.5-year follow-up of a cluster randomized trial. *Ann Fam Med*. 11(2):157–64.
7. Cals JWL, Butler CC, Hopstaken RM, et al. Effect of point of care testing for C reactive protein and training in communication skills on antibiotic use in lower respiratory tract infections: cluster randomised trial. *BMJ*. 2009;338:b1374.
8. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS Med*. 2013;10(2):e1001381.
9. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. *BMJ*. 2009;338(may28 1):b605–b605.
10. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med*. 2000;19(4):453–73.
11. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128–38.
12. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer Science & Business Media; 2008.
13. Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*. 2014;14:40.
14. Harrell FE, Lee KL, Mark DB. Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. *Stat Med*. 1996;15(4):361–387.
15. Bleeker SE, Moll HA, Steyerberg EW, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol*. 2003;56(9):826–32.

16. Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). *Ann Intern Med*. 2015;162(10):735–6.

17. Graffelman AW, le Cessie S, Knuistingh Neven A, et al. Can history and exam alone reliably predict pneumonia? *J Fam Pract*. 2007;56(6):465–70.

18. Heckerling PS, Tape TG, Wigton RS, et al. Clinical prediction rule for pulmonary infiltrates. *Ann Intern Med*. 1990;113(9):664–70.

19. 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(apr30 1):f2450–f2450.

20. Woodhead M, Blasi F, Ewig S, et al. Guidelines for the management of adult lower respiratory tract infections - Full version. *Clin Microbiol Infect*. 2011;17:E1–E59.

21. Minnaard MC, et al. The added value of C-reactive protein in diagnosing pneumonia in primary care: an individual patient data meta-analysis of 5,308 adult patients. *this thesis*. 2015; chapter 3.

22. Mandell LA, Wunderink RG. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis*. 2007 Mar 1;44 Suppl 2:S27-72.

23. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–36.

24. Rubin DB. Multiple imputation for nonresponse in surveys. *Wiley*. 2009.

25. Janes H, Pepe MS. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. *Am J Epidemiol*. 2008;168:89–97.

26. Hedges L V. A random effects model for effect sizes. *Psychol Bull*. 1983;93(2):388–395.

27. Mathew T, Nordström K. Comparison of one-step and two-step meta-analysis models using individual patient data. *Biom J*. 2010;52(2):271–87.

28. Harrell FE. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. *Springer*. 2001.

29. Diehr P, Wood RW, Bushyhead J, et al. Prediction of pneumonia in outpatients with acute cough--a statistical approach. *J Chronic Dis*. 1984;37(3):215–25.

30. Hopstaken RM, Muris JW, Knottnerus JA, et al. Contributions of symptoms, signs, erythrocyte sedimentation rate, and C-reactive protein to a diagnosis of pneumonia in acute lower respiratory tract infection. *Br J Gen Pract*. 2003;53(490):358–64.

31. Melbye H, Straume B, Aasebø U, et al. Diagnosis of pneumonia in adults in general practice. Relative importance of typical symptoms and abnormal chest signs evaluated against a radiographic reference standard. *Scand J Prim Health Care*. 1992;10(3):226–33.

32. Singal BM, Hedges JR, Radack KL. Decision rules and clinical prediction of pneumonia: evaluation of low-yield criteria. *Ann Emerg Med*. 1989;18(1):13–20.
33. Flanders SA, Stein J, Shochat G, et al. Performance of a bedside C-reactive protein test in the diagnosis of community-acquired pneumonia in adults with acute cough. *Am J Med*. 2004;116:529–535.
34. Holm A, Nexoe J, Bistrup LA, et al. Aetiology and prediction of pneumonia in lower respiratory tract infection in primary care. *Br J Gen Pract*. 2007;57:547–554.
35. Rainer TH, Chan CP, Leung MF, et al. Diagnostic utility of CRP to neopterin ratio in patients with acute respiratory tract infections. *J Infect*. 2009;58:123–30.
36. Steurer J, Held U, Spaar A, et al. A decision aid to rule out pneumonia and reduce unnecessary prescriptions of antibiotics in primary care patients with cough and fever. *BMC Med*. 2011;9:56.
37. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339:b2700.
38. De Groot JAH, Dendukuri N, Janssen KJM, et al. Adjusting for differential-verification bias in diagnostic-accuracy studies: a Bayesian approach. *Epidemiology*. 2011;22(2):234–41.
39. Melbye H, Straume B. The spectrum of patients strongly influences the usefulness of diagnostic tests for pneumonia. *Scand J Prim Health Care*. 1993;11(4):241–6.
40. Steyerberg EW, Borsboom GJJM, van Houwelingen HC, et al. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23(16):2567–86.
41. Macfarlane J, Holmes W, Gard P, et al. Prospective study of the incidence, aetiology and outcome of adult lower respiratory tract illness in the community. *Thorax*. 2001;56(2):109–114.
42. Haynes RB, Wilczynski NL. Optimal search strategies for retrieving scientifically strong studies of diagnosis from Medline: analytical survey. *BMJ*. 2004;328:1040.
43. Nancy L Wilczynski, R Brian Haynes, et al. EMBASE search strategies for identifying methodologically sound diagnostic studies for use by clinicians and researchers. *BMC Med*. 2005;3:7.

Appendix 2.1:
Search strategies for PubMed, EMBASE and the Cochrane Library

General:
Reference date: 21-08-2012. No language or pre-set filters used
Filters in Pubmed and EMBASE: Haynes *et al.* and Wilcyinski et al.[42,43]
Mesh term were first mapped and subsequently added to syntax.

PubMed:
(sensitiv*[Title/Abstract] OR sensitivity and specificity[MeSH Terms] OR diagnos*[Title/Abstract] OR diagnosis[MeSH:noexp] OR diagnostic *[MeSH:noexp] OR diagnosis, differential[MeSH:noexp] OR diagnosis[Subheading:noexp]) AND (c-reactive[tiab] OR reactive protein[tiab] OR CRP[tiab] OR C-Reactive Protein[Mesh]) AND (Pneumonie[tiab] OR Pneumonia[tiab] OR Pneumoniae[tiab] OR Pneumonitis[tiab] OR Pneumonias[tiab] OR Airway infection[tiab] OR Airway infections[tiab] OR Airway inflammation[tiab] OR Airway inflammations[tiab] OR Lower respiratory tract infection[tiab] OR Lower respiratory tract infections[tiab] OR Lower respiratory infection[tiab] OR Lower respiratory infections[tiab] OR Lower airway infection[tiab] OR Lower airway infections[tiab] OR Lower airway inflammation[tiab] OR Respiratory infection[tiab] OR Respiratory infections[tiab] OR Respiratory inflammation[tiab] OR Respiratory inflammations[tiab] OR Respiratory tract infection[tiab] OR Respiratory tract infections[tiab] OR Respiratory tract inflammation[tiab] OR Respiratory tract inflammations[tiab] OR LRTI[tiab] OR LRTIS[tiab] OR RTI[tiab] OR RTIS[tiab] OR respiratory tract illness[tiab] OR Pneumonia[Mesh] OR Respiratory Tract Infections[Mesh])

EMBASE:
('pneumonie':ab,ti OR 'pneumonia':ab,ti OR 'pneumoniae':ab,ti OR 'pneumonitis':ab,ti OR 'pneumonias':ab,ti OR 'airway infection':ab,ti OR 'airway infections':ab,ti OR 'airway inflammation':ab,ti OR 'airway inflammations':ab,ti OR 'lower respiratory tract infection':ab,ti OR 'lower respiratory tract infections':ab,ti OR 'lower respiratory infection':ab,ti OR 'lower respiratory infections':ab,ti OR 'lower airway infection':ab,ti OR 'lower airway infections':ab,ti OR 'lower airway inflammation':ab,ti OR 'respiratory infection':ab,ti OR 'respiratory infections':ab,ti OR 'respiratory inflammation':ab,ti OR 'respiratory inflammations':ab,ti OR 'respiratory tract infection':ab,ti OR 'respiratory tract infections':ab,ti OR 'respiratory tract inflammation':ab,ti OR 'respiratory tract inflammations':ab,ti OR 'lrti':ab,ti OR 'lrtis':ab,ti OR 'rti':ab,ti OR 'rtis':ab,ti OR 'respiratory tract illness':ab,ti OR 'pneumonia'/exp OR 'respiratory tract infections'/exp) AND ('specificity' OR predict* OR 'diagnosis':lnk) AND ('c-reactive':ab,ti OR 'reactive protein':ab,ti OR 'crp':ab,ti OR 'c-reactive protein'/exp)

2

Cochrane:

#1:(pneumonie:ab,ti) OR (pneumonia:ab,ti) OR (pneumoniae:ab,ti) OR (pneumonitis:ab,ti) OR (pneumonias:ab,ti) OR (airway infection:ab,ti) OR (airway infections:ab,ti) OR (airway inflammation:ab,ti) OR (airway inflammations:ab,ti) OR (lower respiratory tract infection:ab,ti) OR (lower respiratory tract infections:ab,ti) OR (lower respiratory infection:ab,ti) OR (lower respiratory infections:ab,ti) OR (lower airway infection:ab,ti) OR (lower airway infections:ab,ti) OR (lower airway inflammation:ab,ti) OR (respiratory infection:ab,ti) OR (respiratory infections:ab,ti) OR (respiratory inflammation:ab,ti) OR (respiratory inflammations:ab,ti) OR (respiratory tract infection:ab,ti) OR (respiratory tract infections:ab,ti) OR (respiratory tract inflammation:ab,ti) OR (respiratory tract inflammations:ab,ti) OR (lrti:ab,ti) OR (lrtis:ab,ti) OR (rti:ab,ti) OR (rtis:ab,ti) OR (respiratory tract illness:ab,ti)
#2: (pneumonia[Mesh])
#3: (respiratory tract infections[Mesh]))
#4: (C-Reactive Protein[Mesh])
#5: (c-reactive:ab,ti) OR (reactive protein:ab,ti) OR (CRP:ab,ti) OR
Final syntax cochrane: (#1 OR #2 OR #3) AND (#4 OR #5)

Appendix 2.2:
Overview of methodological quality

Methodological quality of included validation datasets according to QUADAS-2 assessment.[23]

	Risk of bias				Applicability concerns		
Dataset	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Melbye *et al.*	😐±	☺+	☺+	☹–	😐±	☺+	☺+
Hopstaken *et al.*	☺+	☺+	☺+	☺+	☺+	☺+	☺+
Flanders *et al.*	😐±	☺+	☺+	☹–	😐±	☺+	☺+
Graffelman *et al.*	☺+	☺+	☺+	☺+	😐±	☺+	☺+
Holm *et al.*	☺+	☺+	☺+	☺+	☺+	☺+	☺+
Rainer *et al.*	☺+	☺+	☺+	☺+	☹–	☺+	☺+
Steurer *et al.*	☺+	☺+	☺+	☺+	☺+	☺+	☺+
Van Vugt *et al.*	☺+	☺+	☺+	😐±	☺+	☺+	☺+

"☺+" = Low risk, "😐±" = Intermediate risk, "☹-"= High risk

2

Appendix 2.3: Prediction rules

Prediction rules of included diagnostic models for pneumonia in primary care.

Model	Predictors, n	Predicted probability of pneumonia given by: $1/(1+e^{-y})$
Diehr *et al.*'84 *	6	y= -2×Coryza + 1×Night sweats[2] + 1×Myalgia + 1×Phlegm + 2×Respiratory rate >25[3] + 2×Temperature>37.78°
Singal *et al.*'89	3	y= -3.095 + 1.214×Cough[1] + 1.007×Temperature>37.78° + 0.823×Crackles
Heckerling *et al.*'90	5	y= -1.705 + 0.494×Temperature >37.8°+ 0.428×Pulse>100 + 0.658×Crackles + 0.638×Diminished breath sound + 0.691×Absence of asthma
Melbye *et al.*'92 *	6	y= 4.7×Fever (symptom) & Dur. illness ≥1 week - 4.5×Coryza - 2.1×Sore throat +5.0×Dyspnea, very annoying[1] + 8.2×Chest pain, strong, lateral[1] + 0.9×Crackles
Hopstaken *et al.*'03	3	y= -2.74 + 1.02×Dry cough[1] + 1.78×Diarrhea + 1.13×Temperature ≥38°
van Vugt *et al.*'13	6	y= -3.984 + 0.446×Dyspnea + 0.698×Absence of coryza + 0.596×Diminished breath sounds + 1.404×Crackles + 0.961×Pulse>100 + 0.980×Temperature >37.8°

** model does not incorporate an intercept, [1] If available (otherwise unspecified type of predictor used), [2]If available (otherwise "unspecified sweating", used), [3]In the data of Hopstaken et al. a respiratory rate >20 is used.*

Appendix 2.4:
Relative Discriminative Performance

Relative discriminative performance of pneumonia prediction models within datasets, measured as delta AUC. Numbers depict the difference of individual model's AUC to the average AUC of the dataset.

	Validation dataset							
Model	**Melbye *et al.***	**Hopstaken *et al.***	**Flanders *et al.***	**Graffelman *et al.***	**Holm *et al.***	**Rainer *et al.***	**Steurer *et al.***	**van Vugt *et al.***
Van Vugt ***et al.***	0,06	X	0,14	0,01	X	X	X	D
Heckerling ***et al.***	-0,03	X	0,13	0,03	X	X	X	0,04
Diehr ***et al.***	X	-0,02	0,00	X	X	0,07	X	X
Singal ***et al.***	-0,04	0,03	0,05	0,04	NA	X	NA	0,02
Melbye ***et al.***	D	-0,01	-0,14	-0,10	X	X	X	X
Hopstaken ***et al.***	X	D	-0,17	0,02	X	-0,07	X	-0,06

NA = single validation within dataset, dAUC calculated, X = Model not validated in dataset due to missing predictors, D = Development dataset, dAUC not calculated.


3

Chapter 3

The added value of C-reactive protein in diagnosing pneumonia in primary care: an individual patient data meta-analysis of 5,308 adult patients

Margaretha C. Minnaard
Joris de A.H. Groot
Rogier M. Hopstaken
Alwin Schierenberg
Niek J. de Wit
Johannes B. Reitsma
Berna D.L. Broekhuizen
Saskia F. van Vugt
Arie Knuistingh Neven
Aleida W. Graffelman
Hasse Melbye
Timothy H. Rainer
Johann Steurer
Anette Holm
Ralph Gonzalez
Geert-Jan Dinant
Alma C. van de Pol
Theo J.M. Verheij

In revision for CMAJ

Abstract

3

Objectives

C-reactive protein (CRP) is increasingly used in the diagnostic work-up for community acquired pneumonia (CAP) in primary care. The added diagnostic value of CRP beyond signs and symptoms is however unclear. We aim to quantify the added value of CRP in the individual data of multiple studies.

Methods

Meta-analysis of individual patient data (IPD). Studies on diagnostic accuracy of CRP were eligible if they enrolled adult outpatients with suspected lower respiratory tract infections (LRTI). CRP was added to a basic prediction model that reflects daily clinical practice. Improvement in discrimination and risk classification were first calculated within all individual studies and subsequently meta-analysed across studies.

Results

Authors of 8 studies (n=5308) provided their datasets. In all datasets discrimination improved after addition of CRP to the basic model, with a mean improvement in area under the ROC curve (AUC) of 0.075, ranging from 0.02-0.18. Overall the proportion of patients without CAP correctly classified as low risk of CAP increased from 28% to 36% when CRP was added to the model. The proportion of patients with CAP assigned to the low risk category remained equal (n=4) and the proportion of patients assigned to the intermediate risk category decreased from 56% to 51% and the proportion correctly classified as high risk increased from 63% to 70%.

Conclusion

Adding CRP to the diagnostic workup of CAP in patients with LRTI in primary care enhances discrimination and improves diagnostic risk classification, though a substantial group with diagnostic uncertainty remains.

Background

Measurement of the inflammation marker C-reactive protein (CRP)[1] in (capillary) blood is increasingly applied in the diagnostic work-up for community acquired pneumonia (CAP) in primary care.[2] A point-of-care CRP test enables primary care physicians to obtain CRP test results within a few minutes.[3] However, it is unclear what the true added diagnostic value of CRP to signs and symptoms is. Three previously published systematic reviews reported on the diagnostic value of CRP in diagnosing CAP.[4-6] Two reviews evaluated the diagnostic value of CRP as a single test, hence not its added value on top of what is already known for a patient in clinical practice, their signs and symptoms. To assess the diagnostic value of CRP as single test does not reflect the test approach in daily practice. The authors of those reviews considered CRP not sufficiently sensitive and specific to discriminate between patients with and without CAP in primary care.[5,6] A more recent systematic review,[4] based on four diagnostic studies, addressed not only the diagnostic value of CRP for diagnosing CAP but also its ability to discern bacterial aetiology, predict prognosis and influence antibiotic prescription behaviour. Authors concluded that measurement of CRP has (limited) added value in diagnosing pneumonia.[4] Recently three new large diagnostic studies in primary care have been published that were not included in these reviews.[7-9] The previously performed traditional reviews, based on published data only, were forced to work with different combinations of signs and symptoms (as described in the original studies) to evaluate the added value of CRP. This hampered valid comparison between studies and increased heterogeneity in results, as well as reduced the strength of making recommendations about the added value of CRP. In our IPD meta-analysis the same diagnostic model and the same measures of test performance can be evaluated across different (study) datasets, taking into account heterogeneity between studies, providing more valid, informative and generalizable results about the added value of CRP to commonly used and clinically relevant signs and symptoms for diagnosing CAP in primary care.

Methods

Identification, selection and inclusion of studies

A systematic search to identify eligible studies was done in MEDLINE, EMBASE and the Cochrane Library using an electronic search that incorporated indexing terms and plain text words for the index test (CRP) and the target disease (CAP)(details on the full search strategy are presented in Appendix 3.1). A filter was implemented to identify diagnostic studies.[10,11] All studies examining the diagnostic accuracy of CRP with respect to CAP, e.g. infiltrate seen on chest radiography (CXR) as the reference standard, were eligible for inclusion. Studied subjects had to be adult patients (≥18 years) suspected by their physician of lower respiratory tract infection (LRTI) presenting in a primary healthcare setting. 'Suspected LRTI' had to be defined as presence of clinical criteria

suggesting LRTI. Small differences in definitions of LRTI between studies were accepted. Authors of all eligible studies were requested to provide their datasets. For clarity our study does not aim to assess the value of CRP to guide antibacterial prescribing to patients presenting with symptoms of LRTI.[12-14]

Quality Assessment

Two reviewers (AS, JG) independently applied the QUADAS-2 tool[15] to evaluate the risk of bias and the quality of all diagnostic studies included in the IPD meta-analysis. This tool consists of 4 key domains; patient selection, index test, reference standard, and flow and timing. Discrepancies between the reviewers were solved by discussion or, when necessary, by consulting a third reviewer (TV). In case of unclear or undisclosed information with regard to study quality authors of the concerning studies were contacted. The original data were checked on single variables and simple tables and plots were made in order to assess the reproducibility of the reported measures of accuracy in the original publication.

Statistical analysis

Basic signs and symptoms model. To quantify the added value of CRP to signs and symptoms for diagnosing CAP in primary care, first a basic diagnostic model was developed that mimics daily clinical practice using an a-priori selected set of commonly used and clinically relevant signs and symptoms. These clinical predictors were selected from guidelines on management of LRTI[2,16], literature[9,17-21] and discussion with experts. The final set of predictors consisted of age, dyspnoea, tachypnoea, not previously existing chest signs (all new chest signs reported by the GPs defined as either wheezing, rhonchi, crackles, diminished vesicular breathing, pleural rub or dullness), cough, (increased) sputum production, chest pain, ear-nose-throat symptoms (defined as either sore throat or rhinorrhoea), (current) smoking, fever (either self-measured by patients or during consultation by the GP), and selected relevant comorbidity (defined as either heart failure, diabetes mellitus, COPD, asthma, immunodeficiency, malignancy or renal failure). In the datasets of the eligible studies availability of these 11 predictors was evaluated. If a predictor was missing in ≥30% of subjects or if it was missing completely, it was left out of the analysis in that particular dataset. Patients with missing data on the outcome (CAP) were also excluded from the analysis. All other missing values were imputed using available information of all variables in table 3.1 under the reasonable assumption that data was missing at random.[22]

Extended signs and symptoms model. To assess the added value of CRP, the basic model was then extended with CRP. We used visual inspection to assess whether the inclusion of a non-linear component showed a clear deviation from a linear association in a graph.[23] No deviation from linearity was found and CRP was thus added as a continuous variable into the model. For the basic and extended models see Appendix 3.2.

Table 3.1: Missing values.
Missing values (%) of the pre-specified signs & symptoms and CRP in the eight individual patient datasets.

	Melbye *et al.* '92	Hopstaken *et al.* '03	Flanders *et al.* '04	Graffelman *et al.* '07	Holm *et al.* '07	Rainer *et al.* '09	Steurer *et al.* '11	van Vugt *et al.* '13
Age	-	-	-	-	-	-	3.2	-
Dyspnoea	1.5	-	-	-	-	-	0.6	-
Cough	1.5	0.4	-	-	-	-	2.7	0.1
Sputum production	2.2	0.8	-	-	-	-	2.1	0.1
Chest pain	1.7	-	-	-	-	-	3.2	0.1
Ear-nose-throat symptoms[a]	2.0	-	-	-	NA	-	NA	-
(Current) smoking	2.0	0.8	-	-	-	-	-	-
Fever[b]	2.2	22.6	-	-	4.4	-	20.5	0.7
Relevant comorbidity[c]	-	1.2	-	-	-	-	NA	0.1
Tachypnoea	*97.0*	0.4	28	-	-	-	3.5	2.3
New chest signs[d]	-	-	1.2	-	3.3	NA	*48.3*	0.4
C-reactive protein	-	2.9	-	4.7	0.3	-	1.2	0.1

[a] defined as either sore throat or rhinorrhoea, [b] defined as either self-measured by patients or during consultation by the general practitioner, [c] defined as either heart failure, diabetes mellitus, chronic obstructive pulmonary disease, asthma, immunodeficiency, malignancy or renal failure, [d] defined as either wheezing, rhonchi, crackles, diminished vesicular breathing, pleural rub or dullness, NA: not applicable, - not present in dataset

IPD meta-analysis approach. A two-step approach[24,25] was used in this IPD review to meta-analyse two measures of added value of CRP: improvement in discrimination and improvement in classification of patients. These two measures of interest and their precision (i.e. standard error) were first calculated within each of the individual studies and then examined and meta-analysed in the second step. This two-step approach was chosen because (a) it acknowledges the hierarchical nature of the data (i.e., patients and procedures from a single study are more alike and more consistent than across different studies); (b) it is transparent as the methods within each step are comparable to the ones used in an individual study to calculate added value (first step) or comparable to standards methods for meta-analysis (second step); (c) results from individual studies (after first step) can be examined for (in)consistencies.

Outcome measures

Improvement in discrimination. In the first step of the two-step meta-analysis approach basic multivariate prediction models were fitted in each dataset. The ability of these basic models to discriminate between patients with and without CAP was characterized by calculating the area under the receiver operating characteristic curve (AUC) with 95 % confidence intervals (CIs). The change in AUC (delta AUC) between the basic model and the extended model was then calculated for each dataset together with an estimate of its precision using the method of Delong.[26] In the second step, a forest plot was constructed to visualize the improvement in AUC within each dataset and acknowledging the heterogeneity between studies a random effects model was used to obtain a pooled estimate across all studies. The generic inverse variance method was used to calculate the pooled estimate and 95% confidence interval.[27,28]

Improvement in diagnostic risk classification. To give an overall impression how patients were classified by the basic models and the shift in classification (in absolute numbers) after adding CRP, the predicted probability of CAP in all patients was calculated in each dataset for the basic model and the extended model. Three relevant diagnostic risk groups were defined: low risk (predicted probability <2.5%), intermediate risk (predicted probability 2.5-20%) and high risk of CAP (predicted probability >20%), as applied in the most recent study on diagnosing CAP.[9] Using these two cut-offs a 2-by-3 table was constructed for the basic model and similarly for the extended model after adding CRP to each dataset and corresponding measures of accuracy (sensitivities and specificities) were calculated. A fixed effects bivariate regression model was used to simultaneously obtain pooled estimates of sensitivity and specificity.[27-29] To illustrate the general improvement in risk classification when using CRP these pooled sensitivities and specificities and the median prevalence were used to calculate the classification of patients in low, intermediate and high risk for both models in a hypothetical dataset of 1000 patients.[29] Data were analysed using SPSS (version 20.0.0 for Windows, SPSS, Inc. Chicago, Illinois, USA), Microsoft office excel (version 2014) and "R" (version 3.1.1, including the "lme4", "msm" and "rmeta" packages).

Results

Identification, selection and inclusion of studies

Our searches identified a total of 3690 studies, of which eventually 18 met our in- and exclusion criteria (see figure 3.1 for flowchart). The authors of eight of these 18 studies (n=5308 patients)[7-9, 19, 20, 30-32] were willing to provide the individual patient data and were included in our IPD meta-analysis. The median prevalence of CAP across these studies was 13%. The mean age of all individual patients combined was 49 years (with standard deviation (SD) of 18 years). Details on patient characteristics are given in table 3.2.

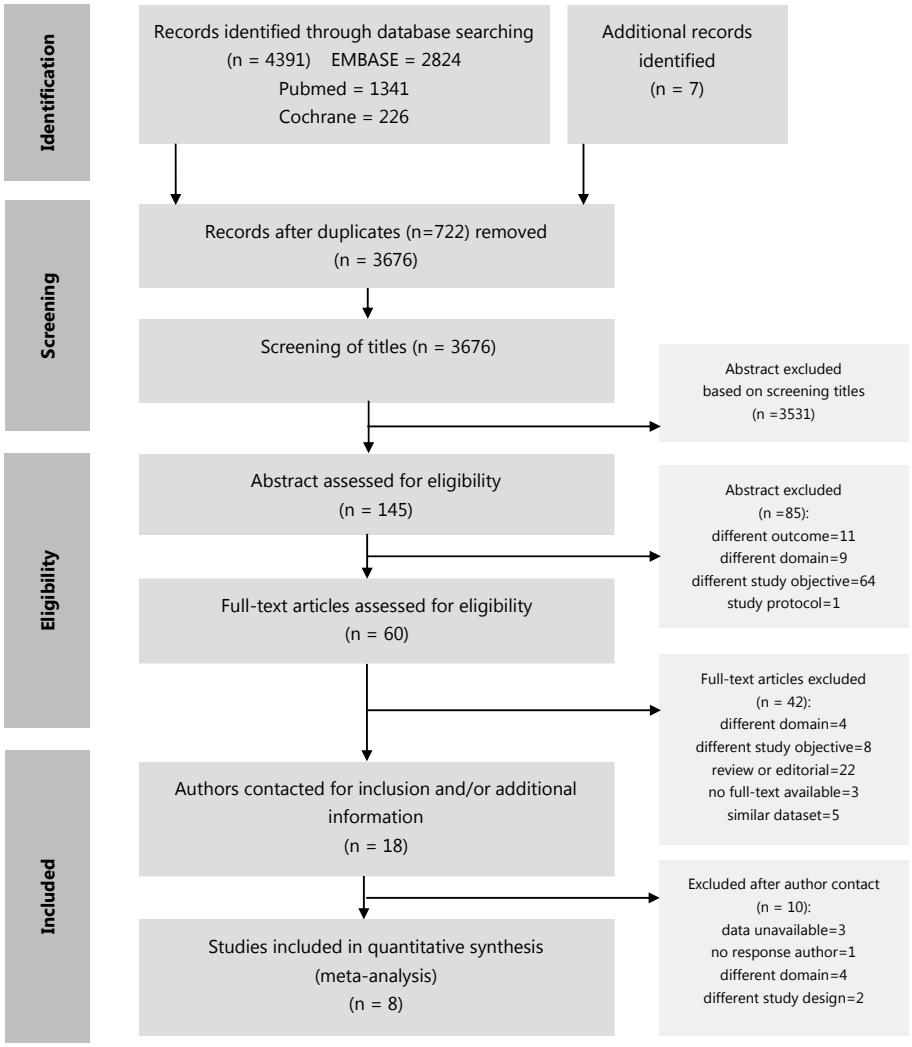


Figure 3.1: PRISMA flow diagram.
PRISMA flow diagram of the selection process of IPD used for external validation of prediction models.[38]

Quality assessment

Quality assessment of the eligible studies identified potential risk of bias and/or applicability concern in patient selection in 5 studies.[7, 9, 20, 30, 31] In the studies by Melbye *et al*[20] and Flanders *et al*,[30] acquisition of CXR (the preferred reference standard) was left up to the physicians' discretion. In the study by Melbye *et al*[20] CXR was ordered when CRP exceeded 60 mg/L, when the physician reported LRTI and CRP exceeded 20 mg/L, in addition to a 25% random sample of the rest of the patients. The study of Rainer *et al*[7] reported CXR results, but these were missing in the provided dataset; therefore

Table 3.2: Patient characteristics.
Patient characteristics per included study. Numbers in table are counts and cases (% per dataset, or specified otherwise)

	Dataset							
	Melbye *et al.* [92]	Hopstaken *et al.* [03]	Flanders *et al.* [04]	Graffelman *et al.* [07]	Holm *et al.* [07]	Rainer *et al.* [09]	Steurer *et al.* [11]	van Vugt *et al.* [13]
Patient characteristics								
Setting	OHD	GP	ED/AC	GP	GP	ED	GP, ED	GP
Number of patients	402	243	168	129	364	561	621	2820
Pneumonia	20 (5%)	32 (13%)	20 (12%)	26 (20%)	48 (13%)	241 (43%)	127 (21%)	140 (5%)
Age, mean (SD)	33 (14)	52 (16)	40 (16)	50 (14)	50 (16)	53 (22)	47 (16)	50 (17)
Gender, male	166 (41%)	115 (47%)	69 (41%)	61 (47%)	179 (49%)	297 (53%)	308 (50%)[1]	1128 (40%)
Duration illness, days (SD)	10 (14)	Categorized	7 (5)	9 (6)	--	17 (9)	7 (10)	10 (10)
(Current) smoker	225 (56%)	81 (33%)	19 (11%)	46 (36%)	165 (45%)	94 (17%)	181 (29%)	777 (28%)
Symptoms								
Cough	363 (91%)	223 (92%)	168 (100%)	127 (98%)	355 (98%)	493 (88%)	602 (97%)	2816 (100%)
Sputum production	352 (88%)	194 (80%)	93 (55%)	102 (79%)	295 (81%)	431 (77%)	302 (49%)	2239 (79%)
Dyspnoea	277 (69%)	188 (77%)	85 (51%)	98 (76%)	263 (72%)	312 (56%)	223 (36%)	1594 (57%)
Coryza	323 (80%)	93 (38%)	115 (69%)	76 (59%)	--	282 (50%)	--	2012 (71%)
Sore throat	290 (72%)	95 (39%)	109 (65%)	50 (39%)	--	283 (50%)	--	--
Chest pain	214 (53%)	145 (60%)	67 (40%)	29 (23%)	234 (64%)	225 (40%)	179 (29%)	1304 (46%)
Diarrhoea	--	19 (8%)	24 (14%)	31 (24%)	--	53 (9%)	--	199 (7%)
Fever[a]	146 (37%)	110 (45%)	102 (61%)	109 (84%)	145 (40%)	470 (84%)	402 (65%)	997 (35%)
Signs								
New chest signs[b]	113 (28%)	206 (83%)	56 (33%)	144 (99%)	129 (35%)	--	262 (42%)	1105 (39%)
Heart rate p.m., mean (SD)	79 (13)	--	85 (19)	82 (11)	81 (15)	98 (18)	--	77 (12)
Respiratory rate p.m.,mean (SD)	--	Categorized	18 (4)	21 (4)	19 (4)	19 (3)	17 (6)	17 (4)
Temperature, C° (SD)	37.3 (0.7)	37.5 (0.8)	37.3 (0.8)	37.9 (0.7)	37.4 (0.6)	37.8 (1.1)	37.4 (1)	36.7 (0.6)
Additional test								
C-reactive protein in mg/L, median (range from 25th – 75th percentile)	5 (5-27)	23 (8-82)	15 (0-35)	50 (17-102)	14 (10-35)	96 (20-258)	36 (15-89)	7 (3-21)

Abbreviations: OHD=Out of Hours Department, GP=General Practice, ED=Emergency Department, AC=Ambulatory Clinic, SD=standard deviation, p.m.= per minute [1] Data from original publication "--"= Variable missing [a] defined as either self-measured by patients or during consultation by the physician (temperature ≥38°Celcius) [b] defined as either wheezing, rhonchi, crackles, diminished vesicular breathing, pleural rub or dullness present at auscultation.

the discharge diagnosis (which was primarily based on CXR results) as provided in the dataset was used. The full QUADAS-2 table is presented in Appendix 3.3.

Performance of the signs and symptoms model (Basic model)

The pre-specified basic model could not be fitted completely in all available datasets. In three datasets one predictor was systematically missing and in one dataset three predictors were systematically missing (table 3.2). For these datasets models without the systematically missing predictors were fitted. The highest percentage of sporadically missing values per predictor was 23%. In one original dataset such missing values had already been imputed in the original analysis using hot-stack imputation.[30] The AUCs for the basic model varied between 0.68 and 0.92 (table 3.3).

Table 3.3: Area under the receiver operating characteristic curve.
Area under the receiver operating characteristic curve (AUC) for the basic and the extended models in the eight different datasets

		AUC basic model (95%CI)	AUC extended model (95%CI)	AUC difference	p-value
Melbye *et al.* '92	n=402	0.77 (0.66-0.89)	0.88 (0.79-0.96)	0.11	0.01*
Hopstaken *et al.* '03	n=243	0.68 (0.59-0.78)	0.86 (0.79-0.94)	0.18	<0.001*
Flanders *et al.* '04	n=168	0.92 (0.85-1.00)	0.97 (0.94-1.00)	0.05	0.155
Graffelman *et al.* '07	n=129	0.76 (0.65-0.86)	0.78 (0.68-0.88)	0.02	0.317
Holm *et al.* '07	n=364	0.78 (0.71-0.86)	0.83 (0.76-0.90)	0.05	0.016*
Rainer *et al.* '09	n=561	0.75 (0.71-0.79)	0.80 (0.76-0.83)	0.05	<0.001*
Steurer *et al.* '11	n=621	0.70 (0.65-0.75)	0.85 (0.81-0.88)	0.15	<0.001*
van Vugt *et al.* '13	n=2820	0.73 (0.68-0.77)	0.79 (0.75-0.83)	0.06	<0.001*

*Abbreviations: AUC= area under the receiver operating characteristic curve, CI=confidence interval, *significant p-value*

Added value of CRP (extended model)

Improvement in discrimination. The improvement in AUC when adding CRP to the basic model ranged from 0.02 to 0.18, and was statistically significant in 6 out of 8 datasets (figure 3.2 and table 3.3). The pooled estimate of the improvement in AUC when adding CRP was 0.075 with a 95% confidence interval of 0.044 to 0.107 (for forest plot see figure 3.3).

Improvement in diagnostic risk classification (pooled results). At the low threshold of 2.5%, the pooled sensitivities for the basic and the extended model were both 0.97. The pooled specificities at this low threshold were 0.28 and 0.36 respectively. At the high threshold of 20% the pooled sensitivities were 0.63 and 0.70, and the pooled specificities were 0.87 and 0.90. We used these pooled sensitivities and specificities together

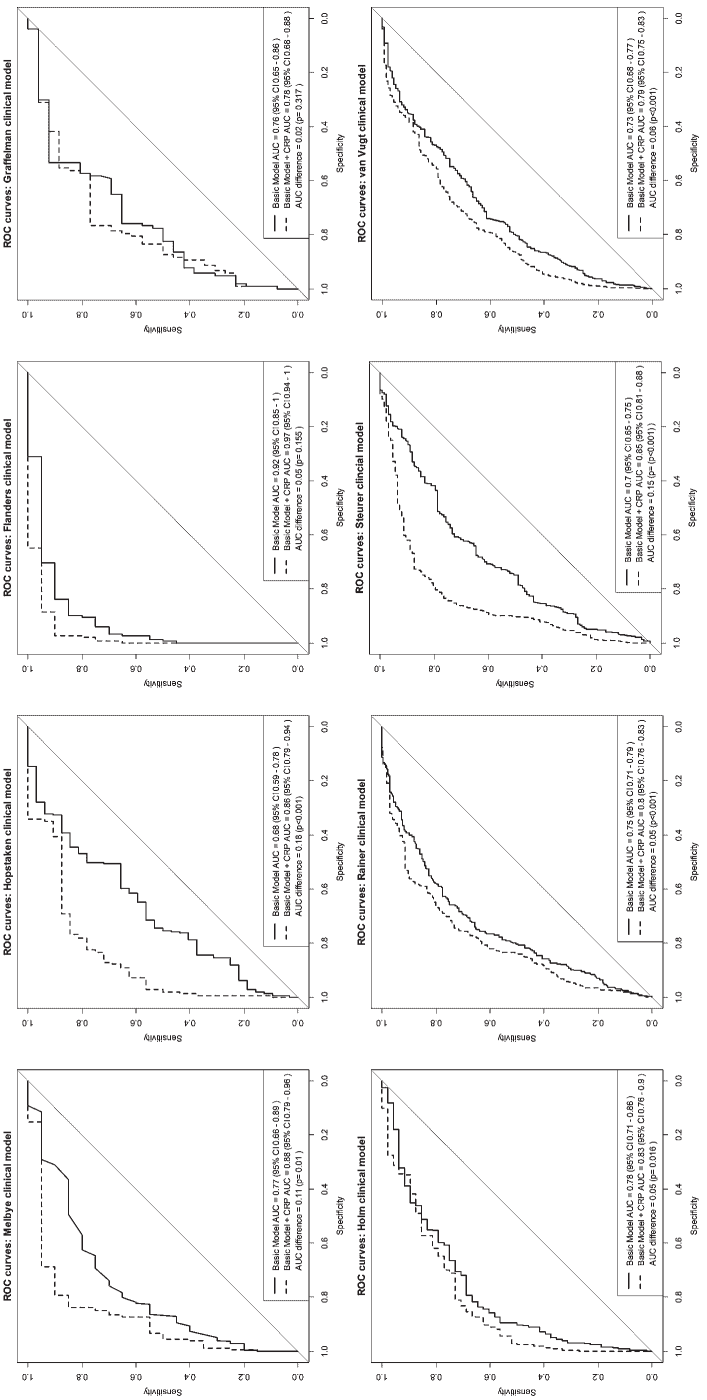


Figure 3.2: Receiver operating characteristic (ROC) curves.
Receiver operating characteristic (ROC) curves with Area under the curve (AUC) for basic model (solid line) and basic model extended with C-reactive protein (dotted line) for eight different datasets, and difference in AUCs.

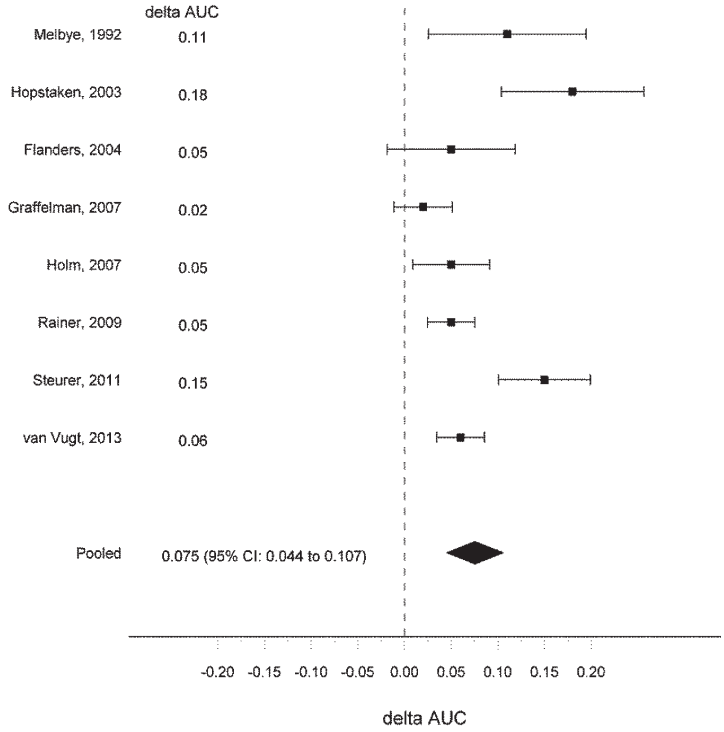


Figure 3.3: Forest plot.
Forest plot showing the improvement in AUC when adding CRP to the basic model in eight individual datasets, together with a pooled estimate with 95% CI using a random effects model.

with the median prevalence of CAP to calculate the summary classification table based on 1000 hypothetical patients (table 3.4).

This illustration shows that using the basic model 248 (25%) of all patients were classified in the low risk group. Four patients out of these 248 low risk patients (2%) are expected to have CAP ("false negatives, missed cases"). Adding CRP to the basic model leads to an increase of the low risk group from 248 (25%) to 317 (31%), whereas the number missed cases with CAP within this low risk group would not increase (n=4). In other words, the proportion of false negatives will decrease from 4 out of 248 (2%) to 4 out of 317 (1%) by adding CRP. Using the basic model 557 (56%) of the 1000 patients were classified in the intermediate risk group. Using the extended model the intermediate risk group decreased to 505 (51%). Using the basic model the high risk group consisted of 195 (20%) and decreased to 178 (18%) patients when using the extended model. The proportion of patients in the high risk group correctly classified as CAP cases increased from 82 of 195 (42%) to 91 of 178 (51%). The proportion of patients in the high risk group who don't have CAP ("false positives") decreased, from 113 of 195

Table 3.4: Reclassification table.
Reclassification table illustrating the improved diagnostic risk classification after adding C-reactive protein (CRP) to the diagnostic work-up of patients suspected of community acquired pneumonia (CAP) in a hypothetical cohort of 1000 patients, based on a relatively low CAP prevalence of 13%.

		Observed CAP		
		Yes	**No**	**Total**
Basic model	Low <2.5%	4*	244*	248
	Intermediate 2.5%-20%	44	513	557
	High >20%	82	113	195
	Total risk group	130	870	1000
Predicted risk of CAP				
Extended model	Low <2.5%	4	313	317
	Intermediate 2.5%-20%	35	470	505
	High >20%	91	87	178
	Total risk group	130	870	1000

Numbers in this table are based on the median prevalence of CAP of 13% across studies and the pooled sensitivities and specificities of the basic and the extended model at low (2.5%) and high (20%) thresholds. At low threshold the pooled sensitivities for the basic and the extended model were both 0.97. The pooled specificities were 0.28 and 0.36 respectively. At high threshold the pooled sensitivities were 0.63 and 0.70, and the pooled specificities were 0.87 and 0.90.

**Calculation example: (1-0.97)*0.13*1000=4 observed CAP patients will get a predicted risk of <2.5% using the basic model. 0.28*(1-0.13)*1000=244 patients without CAP will get a predicted risk of <2.5% using the basic model.*

(58%) with the basic model to 87 of 178 (49%) with the extended model. The proportion of patients with CAP correctly classified as high risk of CAP increased from 82 of 130 (63%) to 91 of 130 (70%) (see table 3.4).

Discussion

Main results

Our IPD meta-analysis including 5308 adults with LRTI symptoms demonstrated that CRP added to signs and symptoms improves the discrimination between those with and without CAP. The pooled analysis further showed that adding CRP to signs and symptoms also improves diagnostic risk classification by increasing the number of patients classified as low risk without increasing the number of "missed" patients with CAP (false negatives) within this category. The number of patients classified as high risk was reduced when adding CRP, but the total number of patients with CAP (true positives) within this category was higher. This means that the proportion of patients that will be falsely categorized as having a high risk for CAP ("false positives") decreases

after adding CRP to the diagnostic process from 58 to 49%. Although CRP does increase diagnostic accuracy, even with the use of CRP a considerable proportion (51%) of patients will be classified as having intermediate risk.

Interpretation of results

Performance of the basic model varied considerably between the included studies (i.e. the AUC varied between 0.68 and 0.92). These heterogeneous results are not surprising considering that patients of the various included studies were selected in different countries, with slightly different primary care settings, the way signs and symptoms have been recorded, and the way CRX has been applied and interpreted (verification problems). The prevalence per study differed (range 5-43%). This clearly underlines that the data from the different studies cannot be analyzed as a single dataset. We therefore used the two-step approach to investigate the true added value of CRP. We found that in two of eight datasets increase in AUC was not statistically significant.[30, 31] Possible explanation is that the AUC of the basic model was already high (>0.75), leaving less room for improvement. We used reclassification tables to give an impression how classifications change by adding CRP in the work-up of all patients with suspected LRTI. Addition of CRP to the model reduced the number of high risk patients without missing patients with CAP, even increasing the number of patients correctly classified with CAP. These numbers change in absolute values by using different thresholds to define low, intermediate and high risk of CAP (data not shown), but the improvement in predicting CAP with CRP remains with different thresholds. The reclassification of diagnostic risk in the hypothetical dataset of 1000 patients was based on the median prevalence found in all 8 datasets combined (13%). Although we think this is a representative figure for the prevalence of CAP in patients presenting in primary care, the prevalence may vary across different communities. We therefore recalculated the reclassification at two different prevalences. At a relatively low CAP prevalence of 5%[9,20], the number of patients classified as false negatives did not change after adding CRP to the model, the number of patients reclassified as true negatives and true positives increased, and a decrease in number of patients reclassified as false positives and in the intermediate diagnostic risk group was seen. This indicates that also at a lower prior risk the addition of CRP improves diagnostic accuracy (Appendix 3.4). At a CAP prevalence of 20% results are comparable (Appendix 3.5).

Strengths and limitations

This is the first meta-analysis using individual patients data investigating the added value of CRP for diagnosing CAP. Study patients were included in primary health care (i.e. non-referred) and therefore the results are representative for daily primary care practice. In the diagnostic risk classification we used predicted risk thresholds of 2.5% and 20%, assuming that when diagnostic risk is low CAP is absent and when diagnostic risk is high CAP is present and has to be treated. This assumption may be debated. We argue this is how physicians are able to use diagnostic risk classification of CAP

in daily practice: a low predicted probability to rule out, a high predicted probability to confirm CAP and an intermediate risk group in which diagnostic doubt remains. Another limitation is that in these analyses CRP is conducted in all patients, while this not necessarily how this test is embedded in daily practice. The strategy of CRP testing varies by region/country. Moreover, CXR although frequently and most practically used in diagnostic studies as reference test, is commonly considered as a suboptimal imperfect gold standard because it does not provide 100% diagnostic certainty on CAP, particularly when used after single-reading.

Comparison with literature

van der Meer *et al.*[6] considered in their review CRP as a single test and found sensitivities ranging from 10-98% and specificities ranging from 44-99%. The positive and negative test results used to calculate sensitivity and specificity were based on dichotomized CRP results (e.g. either CRP <20mg/L as negative test result and CRP ≥20mg/L as positive test result). In our meta-analysis we did not dichotomize CRP results, but added CRP as a continuous predictor to a diagnostic model. We did use dichotomized predicted probabilities as thresholds for calculating sensitivity and specificity (e.g. a predicted probability of CAP <2.5% as a negative test result and a predicted probability of CAP ≥20% as a positive test result). Therefore pooled sensitivity and specificity of this meta-analysis cannot be compared to values of van der Meer *et al.*[6] They conducted also subgroup analysis on 3 homogeneous studies and found a summarized AUC of 0.84 (95% CI 0.78-0.97). Although our meta-analysis also included these 3 studies,[19, 20, 30] next to 5 other studies, for the calculation of the AUC we did not use the different diagnostic models as published in the original studies, but a similar, comparable model across all studies. We found AUCs ranging from 0.68-0.92 for the basic model and ranging from 0.78-0.97 after addition of CRP, which is comparable to results of van der Meer *et al.*[6] We found a slightly different median prevalence (13%, range 5-42%) as compared to the review of Falk and Fahey[5] in their review (14.6%, range 5-89%). Falk and Fahey calculated likelihood ratio's (LRs) at three different thresholds of CRP level, to indicate how much a given CRP test result will raise or lower the pre-test probability of CAP to the post-test probability and they pooled the (heterogeneous) LRs. They concluded that, considering CRP as single test, it may be of value in ruling out CAP in situations where the pre-test probability is higher than 10%. In daily practice CRP is not used as single test and authors suggest examining the incremental value of signs and symptoms alongside CRP testing. Nevertheless they conclude that in primary care settings where the probability of CAP (i.e. prevalence) appears to be between 5 and 10%, additional diagnostic testing with CRP is unlikely to alter the probability of CAP sufficiently to change subsequent management decisions. However we found an improvement in diagnostic risk classification after adding CRP in a hypothetical cohort with 1000 patients. Engel *et al.* concluded in their SR that the additional value of implementing point-of-care CRP measurement in the management of RTI in primary care is limited.[4] In their SR two studies were evaluated in which diagnostic value of CRP was

combined with either a prediction rule[19] or the physicians' presumptive diagnosis of CAP.[32] In the study of Hopstaken *et al.* the AUC increased after adding CRP to the prediction rule. In the study of Holm the positive predictive value increased (from 0.24 to 0.32) and the negative predictive value decreased (from 0.94 to 0.91) after adding CRP to the physicians' presumptive diagnosis. Both studies were included in our analysis, and together with 6 other studies provide more precise estimates of the added value of CRP.

Implications and conclusions

Addition of CRP in the diagnostic workup of patients with LRTI suspected of CAP in primary care enhances diagnostic discrimination and improves classification of patients in terms of risks of CAP. However, even when adding CRP to the diagnostic workup a substantial group will be labelled as having intermediate risk of CAP, in which challenges in clinical decision making remain.

References

1. Clyne B, Olshaker JS. The C-reactive protein. J Emerg Med. 1999;17(6):1019–25.
2. Woodhead M, Blasi F, Ewig S, et al. Guidelines for the management of adult lower respiratory tract infections--full version. Clin Microbiol Infect. 2011;17 Suppl 6:E1–59.
3. Minnaard MC, van de Pol AC, Broekhuizen BDL, et al. Analytical performance, agreement and user-friendliness of five C-reactive protein point-of-care tests. Scand J Clin Lab Invest. 2013;73(8):627–34.
4. Engel MF, Paling FP, Hoepelman AI, et al. Evaluating the evidence for the implementation of C-reactive protein measurement in adult patients with suspected lower respiratory tract infection in primary care: a systematic review. Fam Pract. 2012 Aug;29(4):383–93.
5. Falk G, Fahey T. C-reactive protein and community-acquired pneumonia in ambulatory care: Systematic review of diagnostic accuracy studies. Fam Pract. 2009 Feb;26(1):10–21.
6. Van der Meer V, Knuistingh Neven A, van den Broek PJ, et al. Diagnostic value of C reactive protein in infections of the lower respiratory tract: systematic review. BMJ. 2005 Jul 2;331(7507):26.
7. Rainer TH, Chan CPY, Leung MF, et al. Diagnostic utility of CRP to neopterin ratio in patients with acute respiratory tract infections. J Infect. 2009 Feb;58(2):123–30.
8. Steurer J, Held U, Spaar A, et al. A decision aid to rule out pneumonia and reduce unnecessary prescriptions of antibiotics in primary care patients with cough and fever. BMC Med. 2011 Jan;9:56.
9. 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 Jan;346:f2450.
10. Haynes RB, McKibbon KA, Wilczynski NL, et al. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. BMJ. 2005;330(7501):1179.
11. Wilczynski NL, Haynes RB. EMBASE search strategies for identifying methodologically sound diagnostic studies for use by clinicians and researchers. BMC Med. 2005;3:7.
12. Aabenhus R, Jensen-Jens US, Jørgensen KJ, et al. Biomarkers as point-of-care tests to guide prescription of antibiotics in patients with acute respiratory infections in primary care. Cochrane Database Syst Rev. 2014 Nov 6;11:CD010130.
13. Cooke J, Butler C, Hopstaken RM, et al. Narrative review of primary care point-of-care testing (POCT) and antibacterial use in respiratory tract infection (RTI). BMJ open Respir Res. 2015 Jan;2(1):e000086.
14. Huang Y, Chen R, Wu T, et al. Association between point-of-care CRP testing and antibiotic prescribing in respiratory tract infections: A systematic review and meta-analysis of primary care studies. Br J Gen Pract. 2013 Nov;63(616):787–94.
15. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529–36.

16. NICE Short Clinical Guidelines Technical Team (2008): Respiratory tract infections –antibiotic prescribing. Prescribing of antibiotics for self-limiting respiratory tract infections in adults and children in primary care. London: National Institute for Health and Clinical Excellence; 2008. Guideline number 69.

17. Diehr P, Wood RW, Bushyhead J, et al. Prediction of pneumonia in outpatients with acute cough--a statistical approach. J Chronic Dis. 1984 Jan;37(3):215–25.

18. Heckerling PS, Tape TG, Wigton RS, et al. Clinical prediction rule for pulmonary infiltrates. Ann Intern Med. 1990 Nov 1;113(9):664–70.

19. Hopstaken RM, Muris JW, Knottnerus JA, et al. Contributions of symptoms, signs, erythrocyte sedimentation rate, and C-reactive protein to a diagnosis of pneumonia in acute lower respiratory tract infection. Br J Gen Pract. 2003 May;53(490):358–64.

20. Melbye H, Straume B, Aasebø U, et al. Diagnosis of pneumonia in adults in general practice. Relative importance of typical symptoms and abnormal chest signs evaluated against a radiographic reference standard. Scand J Prim Health Care. 1992 Sep;10(3):226–33.

21. Singal BM, Hedges JR, Radack KL. Decision rules and clinical prediction of pneumonia: evaluation of low-yield criteria. Ann Emerg Med. 1989 Jan;18(1):13–20.

22. Rubin D. Multiple Imputation for Nonresponse in Surveys. Hoboken, New Jersey: John Wiley & Sons, Inc. 1987.

23. Harrell FE, Lee KL, Mark DB. Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. Stat Med. 1996;15(4):361–87.

24. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010;340:c221.

25. Stewart GB, Altman DG, Askie LM, et al. Statistical Analysis of Individual Participant Data Meta-Analyses: A Comparison of Methods and Recommendations for Practice. PLoS One. 2012;7(10):1–8.

26. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988 Sep;44(3):837–45.

27. Higgins J, Thompson S, Deeks J, et al. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. J Health Serv Res Policy. 2002 Jan;7(1):51–61.

28. The Cochrane Collaboration. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane – handbook.org.

29. Schouten HJ, Geersing GJ, Koek HL, et al. Diagnostic accuracy of conventional or age adjusted D-dimer cut-off values in older patients with suspected venous thromboembolism: systematic review and meta-analysis. BMJ. 2013;346(May):f2492.

30. Flanders SA, Stein J, Shochat G, et al. Performance of a bedside C-reactive protein test in the diagnosis of community-acquired pneumonia in adults with acute cough. Am J Med. 2004;116(8):529–35.

31. Graffelman AW, le Cessie S, Knuistingh Neven A, et al. Can history and exam alone reliably predict pneumonia? J Fam Pract. 2007 Jun;56(6):465–70.

32. Holm A, Pedersen SS, Nexoe J, et al. Procalcitonin versus C-reactive protein for predicting pneumonia in adults with lower respiratory tract infection in primary care. Br J Gen Pract. 2007;57(540):555–60.

Appendix 3.1:
Search strategies for PubMed, EMBASE and the Cochrane Library.

General:
Reference date: 21-08-2012. No language or pre-set filters used.
Filters in Pubmed and EMBASE: Haynes et al. and Wilcyinski *et al.*
Mesh term were first mapped and subsequently added to syntax.

PubMed:
(sensitiv*[Title/Abstract] OR sensitivity and specificity[MeSH Terms] OR diagnos*[Title/Abstract] OR diagnosis[MeSH:noexp] OR diagnostic *[MeSH:noexp] OR diagnosis, differential[MeSH:noexp] OR diagnosis[Subheading:noexp]) AND (c-reactive[tiab] OR reactive protein[tiab] OR CRP[tiab] OR C-Reactive Protein[Mesh]) AND (Pneumonie[tiab] OR Pneumonia[tiab] OR Pneumoniae[tiab] OR Pneumonitis[tiab] OR Pneumonias[tiab] OR Airway infection[tiab] OR Airway infections[tiab] OR Airway inflammation[tiab] OR Airway inflammations[tiab] OR Lower respiratory tract infection[tiab] OR Lower respiratory tract infections[tiab] OR Lower respiratory infection[tiab] OR Lower respiratory infections[tiab] OR Lower airway infection[tiab] OR Lower airway infections[tiab] OR Lower airway inflammation[tiab] OR Respiratory infection[tiab] OR Respiratory infections[tiab] OR Respiratory inflammation[tiab] OR Respiratory inflammations[tiab] OR Respiratory tract infection[tiab] OR Respiratory tract infections[tiab] OR Respiratory tract inflammation[tiab] OR Respiratory tract inflammations[tiab] OR LRTI[tiab] OR LRTIS[tiab] OR RTI[tiab] OR RTIS[tiab] OR respiratory tract illness[tiab] OR Pneumonia[Mesh] OR Respiratory Tract Infections[Mesh])

EMBASE:
('pneumonie':ab,ti OR 'pneumonia':ab,ti OR 'pneumoniae':ab,ti OR 'pneumonitis':ab,ti OR 'pneumonias':ab,ti OR 'airway infection':ab,ti OR 'airway infections':ab,ti OR 'airway inflammation':ab,ti OR 'airway inflammations':ab,ti OR 'lower respiratory tract infection':ab,ti OR 'lower respiratory tract infections':ab,ti OR 'lower respiratory infection':ab,ti OR 'lower respiratory infections':ab,ti OR 'lower airway infection':ab,ti OR 'lower airway infections':ab,ti OR 'lower airway inflammation':ab,ti OR 'respiratory infection':ab,ti OR 'respiratory infections':ab,ti OR 'respiratory inflammation':ab,ti OR 'respiratory inflammations':ab,ti OR 'respiratory tract infection':ab,ti OR 'respiratory tract infections':ab,ti OR 'respiratory tract inflammation':ab,ti OR 'respiratory tract inflammations':ab,ti OR 'lrti':ab,ti OR 'lrtis':ab,ti OR 'rti':ab,ti OR 'rtis':ab,ti OR 'respiratory tract illness':ab,ti OR 'pneumonia'/exp OR 'respiratory tract infections'/exp) AND ('specificity' OR predict* OR 'diagnosis':lnk) AND ('c-reactive':ab,ti OR 'reactive protein':ab,ti OR 'crp':ab,ti OR 'c-reactive protein'/exp)

Cochrane:

#1:(pneumonie:ab,ti) OR (pneumonia:ab,ti) OR (pneumoniae:ab,ti) OR (pneumonitis:ab,ti) OR (pneumonias:ab,ti) OR (airway infection:ab,ti) OR (airway infections:ab,ti) OR (airway inflammation:ab,ti) OR (airway inflammations:ab,ti) OR (lower respiratory tract infection:ab,ti) OR (lower respiratory tract infections:ab,ti) OR (lower respiratory infection:ab,ti) OR (lower respiratory infections:ab,ti) OR (lower airway infection:ab,ti) OR (lower airway infections:ab,ti) OR (lower airway inflammation:ab,ti) OR (respiratory infection:ab,ti) OR (respiratory infections:ab,ti) OR (respiratory inflammation:ab,ti) OR (respiratory inflammations:ab,ti) OR (respiratory tract infection:ab,ti) OR (respiratory tract infections:ab,ti) OR (respiratory tract inflammation:ab,ti) OR (respiratory tract inflammations:ab,ti) OR (lrti:ab,ti) OR (lrtis:ab,ti) OR (rti:ab,ti) OR (rtis:ab,ti) OR (respiratory tract illness:ab,ti)

#2: (pneumonia[Mesh])

#3: (respiratory tract infections[Mesh]))

#4: (C-Reactive Protein[Mesh])

#5: (c-reactive:ab,ti) OR (reactive protein:ab,ti) OR (CRP:ab,ti) OR

Final syntax cochrane: (#1 OR #2 OR #3) AND (#4 OR #5)

Appendix 3.2:
Signs and symptoms basic and extended model.

Basic model:

Estimated risk of pneumonia = 1 / (1+ exp -(α + β x [age] + β_1 x [dyspnoea] + β_2 x [new chest signs[a]] + β_3 x [cough] + β_4 x [increased sputum production] + β_5 x [ear-nose-throat symptoms[b]] + β_6 x [current smoking] + β_7 x [fever[c]] + β_8 x [comorbidity[d]])

Extended model:

Estimated risk of pneumonia = 1 / (1+ exp -(α + β x [age] + β_1 x [dyspnoea] + β_2 x [new chest signs[a]] + β_3 x [cough] + β_4 x [increased sputum production] + β_5 x [ear-nose-throat symptoms[b]] + β_6 x [current smoking] + β_7 x [fever[c]] + β_8 x [comorbidity[d]] + β_9 x [CRP[e]])

[a] *new chest signs are defined as either wheezing, rhonchi, crackles, diminished vesicular breathing, pleural rub or dullness.*

[b] *ear-nose-throat symptoms are defined as either sore throat or rhinorrhea.*

[c] *fever is defined as temperature ≥ 38.5ºC either self-measured by patients or during consultation by the GP.*

[d] *relevant comorbidity is defined as either heart failure, diabetes mellitus, COPD, asthma, immunodeficiency, malignancy or renal failure.*

[e] *C-reactive protein*

Appendix 3.3:
Overview of methodological quality.

3

Methodological quality of included validation datasets according to QUADAS-2 assessment.

	Risk of bias				Applicability concerns		
Dataset	***Patient selection***	***Index test***	***Reference standard***	***Flow and timing***	***Patient selection***	***Index test***	***Reference standard***
Melbye *et al.*	😐±	☺+	☺+	☹–	😐±	☺+	☺+
Hopstaken *et al.*	☺+	☺+	☺+	☺+	☺+	☺+	☺+
Flanders *et al.*	😐±	☺+	☺+	☹–	😐±	☺+	☺+
Graffelman *et al.*	☺+	☺+	☺+	☺+	😐±	☺+	☺+
Holm *et al.*	☺+	☺+	☺+	☺+	☺+	☺+	☺+
Rainer *et al.*	☺+	☺+	☺+	☺+	☹–	☺+	☺+
Steurer *et al.*	☺+	☺+	☺+	☺+	☺+	☺+	☺+
Van Vugt *et al.*	☺+	☺+	☺+	😐±	☺+	☺+	☺+

"☺+" = Low risk, "😐±" = Intermediate risk, "☹-"= High risk

Appendix 3.4:
Reclassification table (CAP prevalence 5%).

Reclassification table illustrating the improved diagnostic risk classification after adding C-reactive protein (CRP) to the diagnostic work-up of patients suspected of community acquired pneumonia (CAP) in a hypothetical cohort of 1000 patients, based on a relatively low CAP prevalence of 5%.

		Observed CAP		
		Yes	**No**	**Total**
Basic model	Low <2.5%	2	266	268
	Intermediate 2.5%-20%	16	560	576
	High >20%	32	124	156
	Total risk group	50	950	1000
Predicted risk of CAP				
Extended model	Low <2.5%	2	342	344
	Intermediate 2.5%-20%	13	513	526
	High >20%	35	95	130
	Total risk group	50	950	1000

Numbers in this table are based on a prevalence of CAP of 5% and the pooled sensitivities and specificities of the basic and the extended model at low (2.5%) and high (20%) thresholds. At low threshold the pooled sensitivities for the basic and the extended model were both 0.97. The pooled specificities were 0.28 and 0.36 respectively. At high threshold the pooled sensitivities were 0.63 and 0.70, and the pooled specificities were 0.87 and 0.90.

Appendix 3.5:
Reclassification table (CAP prevalence 20%).

3

Reclassification table illustrating the improved diagnostic risk classification after adding C-reactive protein (CRP) to the diagnostic work-up of patients suspected of community acquired pneumonia (CAP) in a hypothetical cohort of 1000 patients, based on a relatively low CAP prevalence of 20%.

		Observed CAP		
		Yes	**No**	**Total**
Basic model	Low <2.5%	6	224	230
	Intermediate 2.5%-20%	68	472	540
	High >20%	126	104	230
	Total risk group	200	800	1000
Predicted risk of CAP				
Extended model	Low <2.5%	6	288	294
	Intermediate 2.5%-20%	54	432	486
	High >20%	140	80	220
	Total risk group	200	800	1000

Numbers in this table are based on a prevalence of CAP of 20% and the pooled sensitivities and specificities of the basic and the extended model at low (2.5%) and high (20%) thresholds. At low threshold the pooled sensitivities for the basic and the extended model were both 0.97. The pooled specificities were 0.28 and 0.36 respectively. At high threshold the pooled sensitivities were 0.63 and 0.70, and the pooled specificities were 0.87 and 0.90.


4

Chapter 4

Analytical performance, agreement and user-friendliness of five C-reactive protein point-of-care tests

Margaretha C. Minnaard
Alma C. van de Pol
Berna D.L. Broekhuizen
Theo J.M. Verheij
Rogier M. Hopstaken
Sanne van Delft
Antoinette M.J. Kooijman-Buiting
Joris A.H. de Groot
Niek J. de Wit

Scand J Clin Lab Invest. 2013 Dec;73(8):627-34

Abstract

Background

Point-of-care (POC) C-reactive protein (CRP) testing is increasingly used in primary care to assist general practitioners (GPs) in the diagnostic workup for various complaints. The present study compares analytical performance, agreement and user-friendliness of five of these POC CRP tests.

Methods

The following five POC CRP tests were evaluated: Afinion and NycoCard Reader II (both Alere), Eurolyser Smart 700|340 (Eurolyser), QuikRead go and QuikRead 101 (both Orion Diagnostica). Results were compared with those of a standard immunoturbidimetric method performed on a routine analyser (Olympus AU 2700, Beckman Coulter). Analytical performance and agreement with the laboratory standard for the five different POC tests were analysed. Subsequently, user-friendliness of the POC tests was assessed.

Results

Within-day CVs varied from 2.6% (QuikRead go) – 19.4% (Eurolyser Smart 700|340) for low CRP values (<20 mg/L), and 1.1% (QuikRead go) – 17.5% (Eurolyser Smart 700|340) for high values (>100 mg/L). Between-day CVs varied from 4.6% (Afinion) – 30.5% (Eurolyser Smart 700|340) for low values and 4.0% (QuikRead go) – 18.0% (Eurolyser Smart 700|340) for high values. For all POC tests the overall mean difference with the laboratory standard was < 4.2 mg/L. However, with high CRP values (>100 mg/L) agreement with the laboratory standard systematically decreased for all POC tests. Regarding user-friendliness Afinion and Eurolyser Smart 700|340 were judged easiest to operate.

Conclusions

Analytical performance, agreement, and user-friendliness of the POC CRP tests varied considerably, yet overall four devices showed acceptable results and are considered suitable for use in general practice.

Introduction

C-reactive protein (CRP) is an acute phase protein synthesized by the liver in response to inflammation within 6-8 hours. Point-of-care (POC) C-reactive protein (CRP) testing is increasingly used in primary care to assist general practitioners (GPs) in the diagnostic workup for various complaints, including acute cough and abdominal pain.[1,2] CRP testing, be it conventional or POC is currently being implemented in Dutch primary care to adequately differentiate between mild and severe respiratory tract infections and guide appropriate prescription of antibiotics.[2-4] Also, CRP is used in the diagnostic work-up of abdominal complaints, to increase diagnostic certainty in possible diverticulitis or appendicitis.[1,5,6] CRP measurement is mainly useful to exclude severe appendicitis or diverticulitis in patients with abdominal pain. Discussion about utility of CRP testing in management decisions (e.g. guiding antibiotic prescription after differentiating between viral and bacterial infection and referral to hospital) and cost-effectiveness is ongoing[7-11], however CRP testing is currently widely implemented. It is expected that POC CRP use in primary care will increase even further in the near future because there is an urgent need for simple diagnostic procedures that can support clinical decision making in order to limit unnecessary drug prescription and referral. Currently, several different POC CRP test systems are commercially available. Studies directly comparing these systems are scarce and it is unclear which of the assays has best test characteristics for use in everyday primary care practice.[10,12-18] None of the previous studies evaluated user-friendliness of the different POC CRP tests. Therefore, we evaluated the analytical performance, agreement and user-friendliness of five different POC CRP tests, with a laboratory CRP test as the reference standard.

Materials and Methods

Patient samples

Anonymous blood samples were collected by Saltro Diagnostic Center, providing laboratory services, and other diagnostic services, for approximately 350 GPs and 820.000 inhabitants.

Residual stored material from routinely performed laboratory blood tests was pooled to obtain lithium heparin plasma pools [19;20] with CRP concentrations of approximately 10 mg/L, 15 mg/L, 20 mg/L, 25 mg/L, 50 mg/L, 90 mg/L, 100 mg/L and 110 mg/L. These plasma pools were used to determine the analytical performance of the five different POC CRP analysers and to assess the correlation and agreement with the laboratory standard.[21,22] According to the Medical-Ethics Committee of the University Medical Center Utrecht informed consent for this study with pooled anonymous patient material was not required.

CRP tests and procedures

For the current study all five quantitative POC CRP test systems suitable for serum and capillary blood commercially available in the Netherlands (2012) were evaluated. An immunoturbidimetric assay on a routine laboratory analyser (Olympus AU2700, Beckman Coulter) was used for reference testing. This is an instrument calibrated with CRP latex calibrators traceable to IFCC standard CRM470 with a within day precision (CV, n=80) of 3,22% (mean=6,59 mg/l),1,02% (mean=62,84mg/l), and a between day precision (CV, n=20) of 3,79% (mean=6,59 mg/l) and 1,92% (mean=62,84mg/l), with linearity between 0,2 and 480 mg/l. The following five POC CRP test systems were evaluated:

Afinion AS100 Analyzer (Alere / Axis Shield, Oslo, Norway). This test is based on an immunometric membrane 'flow through' assay. It is a solid phase, sandwich-format, immunometric assay. The sample mixture is aspirated through a membrane coated with anti-CRP antibodies and the CRP in the sample is concentrated onto this membrane. A solution containing anti-CRP antibodies conjugated with ultra-small gold particles is then aspirated through the membrane. The gold antibody conjugate binds to the immobilized CRP on the membrane, which will turn red-brown. Excess gold antibody conjugate is removed from the membrane by a washing solution. The color intensity of the membrane is proportional to the amount of CRP in the sample. The test cartridge contains all reagents necessary for one sample analysis. Measurement range in serum is 5 – 160 mg/L.

NycoCard Reader II (Alere / Axis-Shield, Oslo, Norway) The NycoCard Reader II is the predecessor of the Afinion and the mechanism for measuring CRP is similar for both analysers. Sample and reagent mixing has to be performed manually. Measurement range of serum is 5 -120 mg/L.

Eurolyser Smart 700/340 CRP (Eurolyser, Salzburg, Austria) The Eurolyser Smart 700|340 CRP is a immunoturbidimetric test. Kinetic determination of the concentration of CRP is done by photometric measurement at 700 nanometer of antigen-antibody reaction between antibodies to human CRP bound to polystyrene particles and CRP present in the sample. The test cartridge contains all reagents necessary for one sample analysis. Measurement range in serum is 1 – 120 mg/L.

QuikRead go (Orion Diagnostica, Espoo, Finland) The QuikRead go is an immunoturbidimetric assay in which nanoparticles coated with anti-human CRP fragment antigen binding (F(ab)2) fragments react with CRP present in the sample. Turbidity change of the solution resulting from immunoreactions is measured with the QuikRead go instrument. Measurement range in serum is 5 – 120 mg/L when using a sample volume of 20 uL and 5-200 mg/L when using a sample volume of 12 uL.

QuikRead 101 (Orion Diagnostica, Espoo, Finland) The QuikRead 101 is the predecessor of the QuikRead go and the mechanism for measuring CRP is similar for both analysers. Sample and reagent mixing has to be performed manually. Measurement range in serum is 8 – 160 mg/L.

All POC tests were performed according to the manufacturers' instructions in a laboratory setting by laboratory analysts blinded to outcomes of the reference test. Afinion, NycoCard Reader II, QuikRead go and Eurolyser Smart 700|340 tests were all performed within the same timeframe in one sequence of testing and verified with the AU2700 test result. Due to logistic issues, tests with the QuikRead 101 were performed on a different occasion with newly created plasma pools. These new pools were additionally tested with the Olympus AU2700, the conventional laboratory system.

Data analysis

Analytical performance. To measure within-day and between-day coefficients of variation (CVs) a protocol was used based on the clinical and laboratory standards institute EP5 and EP9 guidelines.[23, 24] CVs of each CRP assay were quantified testing 20 aliquots of the low (~20 mg/L) and high (~100 mg/L) level plasma pools. A CV of <15%, was considered to be adequate for POC CRP tests, as no commonly accepted thresholds exist.[25, 26] To indicate to what extent analytical distribution of the test relates to the biological distribution, sigma scores were calculated using the formula (total allowable error– bias)/CV.[27]

Agreement. From each of the eight serum pools with ascending concentrations 10 aliquots were tested to assess agreement between the POC tests and the laboratory reference test using Bland-Altman plots.[28, 29] Based on previous studies, we expected that the agreement between the POC tests and the reference test would depend on the magnitude of the CRP value.[12-15] To accommodate this expectation we added skewed limits of agreement to the Bland-Altman plots.[28] To allow for comparable figures, we show the CRP range from 0 to 120 mg/L for all devices. All reported results, however, are based on all available measurements. Finally, the differences and the 95% CI of the mean differences between the POC test results and the laboratory standard were calculated for all CRP values and for the ranges <20 mg/L, 20-100 mg/L, and >100mg/L.

User-friendliness. User-friendliness of the POC tests was assessed using a standardized questionnaire derived from a published survey by Geersing *et al*.[30] Data on test characteristics was collected from manufacturers' information sheets as set on January 2012. Time required for complete analysis was measured twice on independent moments by different researchers. Susceptibility to flaws in the testing procedure was scored by 20 GPs and GP assistants (GPAs) who were unfamiliar with POC testing. The GP(A)s were instructed on how to use the POC systems and directly hereafter performed one CRP

test with all five devices in a random order and filled out the questionnaire on liability to flaws in the different procedures.

Results

Data analysis

Analytical performance. Within-day and between-day variation calculated for the different POC analysers are shown in table 4.1. The within-day and between-day mean (standard deviation) CRP levels measured by the laboratory reference were both 21 (0.4) for low (~20 mg/L) CRP serum pools and 108 (0.8) and 108 (1.2) for the high (~110 mg/L) plasma pools. Within-day CVs for low level CRP ranged from 2.6% at a mean CRP value of 19 mg/L (QuikRead go) to 19.4% at a mean CRP value of 26 mg/L (Eurolyser smart 700|340). Within-day CVs for high level CRP ranged from 1.1% at a mean CRP value of 95 mg/L (QuikRead go) to 17.5% at a mean CRP value of 122 mg/L (Eurolyser smart 700|340). Between-day CVs were 4.6% (mean 23 mg/L; Afinion) to 30.5% (mean 34 mg/L; Eurolyser smart 700|340); and 4.0% (mean 106 mg/L; QuikRead go) to 18% (mean 136 mg/L; Eurolyser smart 700|340), for low and high levels, respectively.

Agreement. Figure 4.1 shows the agreement of the POC tests with the laboratory test by Bland-Altman plots with 95% limits of agreement. Over the whole range of CRP levels, Afinion, NycoCard Reader II, QuikRead go and QuikRead 101 (figures 4.1A, 4.1B, 4.1D and 4.1E) showed better agreement with the laboratory test than the Eurolyser Smart 700|340 (figure 4.1C). The skewed 95% limits of agreement in the Bland-Altman plots showed that the agreement between all five POC test and the laboratory test decreased at high CRP levels. Overall mean differences and mean differences for the CRP ranges <20, 20-100 and >100mg/L and accompanying 95% CIs are shown in table 4.2. For all POC tests the overall mean difference with the laboratory standard was < 4.2 mg/L. However, with high CRP values (>100 mg/L) agreement with the laboratory standard systematically decreased for all POC tests, resulting in wider 95% CIs around the mean differences between the result POC test results and the laboratory standard, up to -53.2 to 55.0 accompanying a mean difference of 0.9 mg/L for CRP> 100mg/L measured by the Eurolyser.

User-friendliness

Device characteristics. Device details are set out in table 4.3. Data-connectivity is possible unilaterally from device to electronic patient file for Afinion, Eurolyser Smart 700|340 and QuikRead go. Measured data on these devices can also be stored on the device itself. Test kits of all POC devices are preferably stored at temperatures between 2 and 8 degrees Celsius. To store test kits at room temperature is possible but reduces storage life. Sample storage at room temperature is possible for NycoCard Reader II test kits (1 hour), QuikRead go and QuikRead 101 test kits (both 2 hours). Most devices

Table 4.1: Within- and between-day variation.
Within- and between-day results for low CRP values (~20 mg/L) and high CRP values (~100 mg/L) with mean and range, in mg/L and CV, in %.

	Within-day variation[a]				Between-day variation[b]			
	mean in mg/L (range)	CV %	mean in mg/L (range)	CV %	mean in mg/L (range)	CV %	mean in mg/L (range)	CV %
Reference standard								
Olympus AU2700	21 (20-22)	1.8	108 (106-109)	0.8	21 (20-21)	1.9	108 (106-112)	1.1
POC test device								
Afinion	21 (18-24)	7.4[c]	107 (95-117)	5.3	23 (21-25)	4.6	115 (102-146)	7.3
NycoCard Reader II	28 (20-37)	13.3	112 (82-120)	9.8	25 (15-33)	16.9	114 (99-120)	6.0
Eurolyser Smart	26 (14-36)	19.4	122 (76-152)	17.5	34 (26-99)	30.5	136 (94-160)	18.0
QuikRead go	19 (18-20)	2.6	95 (93-96)	1.1	25 (23-32)	8.3	106 (100-114)	4.0
QuikRead 101[d]	18 (16-20)	5.4	107 (96-123)	5.7	19 (16-21)	9.3	92 (82-107)	6.3

[a] *For the within-day variation both low (~20 mg/L) and high (~100 mg/L) CRP was measured 20 times within one day.*
[b] *For the between-day variation both low (~20 mg/L) and high (~100 mg/L) CRP was measured 2 times a day for 10 consecutive days.*
[c] *Calculation example: within-day CV at low level CRP on Afinion: CV = (standard deviation/mean) * 100 = (1.55/21) *100=7.4%.*
[d] *New pools with comparable reference values for this POC test device were created.*

Table 4.2: Mean differences.
Mean differences (with 95% confidence intervals) between all five different Point-of-care (POC) C-reactive protein (CRP) tests and the laboratory test (Olympus UA2700).

	Point-of-care test device				
	Afinion	NycoCard Reader II	Eurolyser	QuikRead go	QuikRead 101
CRP < 20 mg/L mean difference (mg/L) (95% CI)	-1.1 (-3.2; 1.0)	0.3 (-4.4; 5.0)	1.9 (-10.2; 14.1)	0.2 (-1.2; 1.5)	-1.4 (-3.2; 0.4)
CRP 20-100 mg/L mean difference (mg/L) (95% CI)	-0.3 (-6.4; 5.8)	3.0 (-6.2; 12.2)	7.8 (-27.7; 43.3)	2.3 (-3.6; 8.3)	-6.5 (-15.6; 2.7)
CRP > 100 mg/L mean difference (mg/L) (95% CI)	14.7 (-21.1; 50.5)	-10.7 (-30.4; 9.1)	0.9 (-53.2; 55.0)	-8.9 (-21.4; 3.5)	3.2 (-8.4; 14.8)

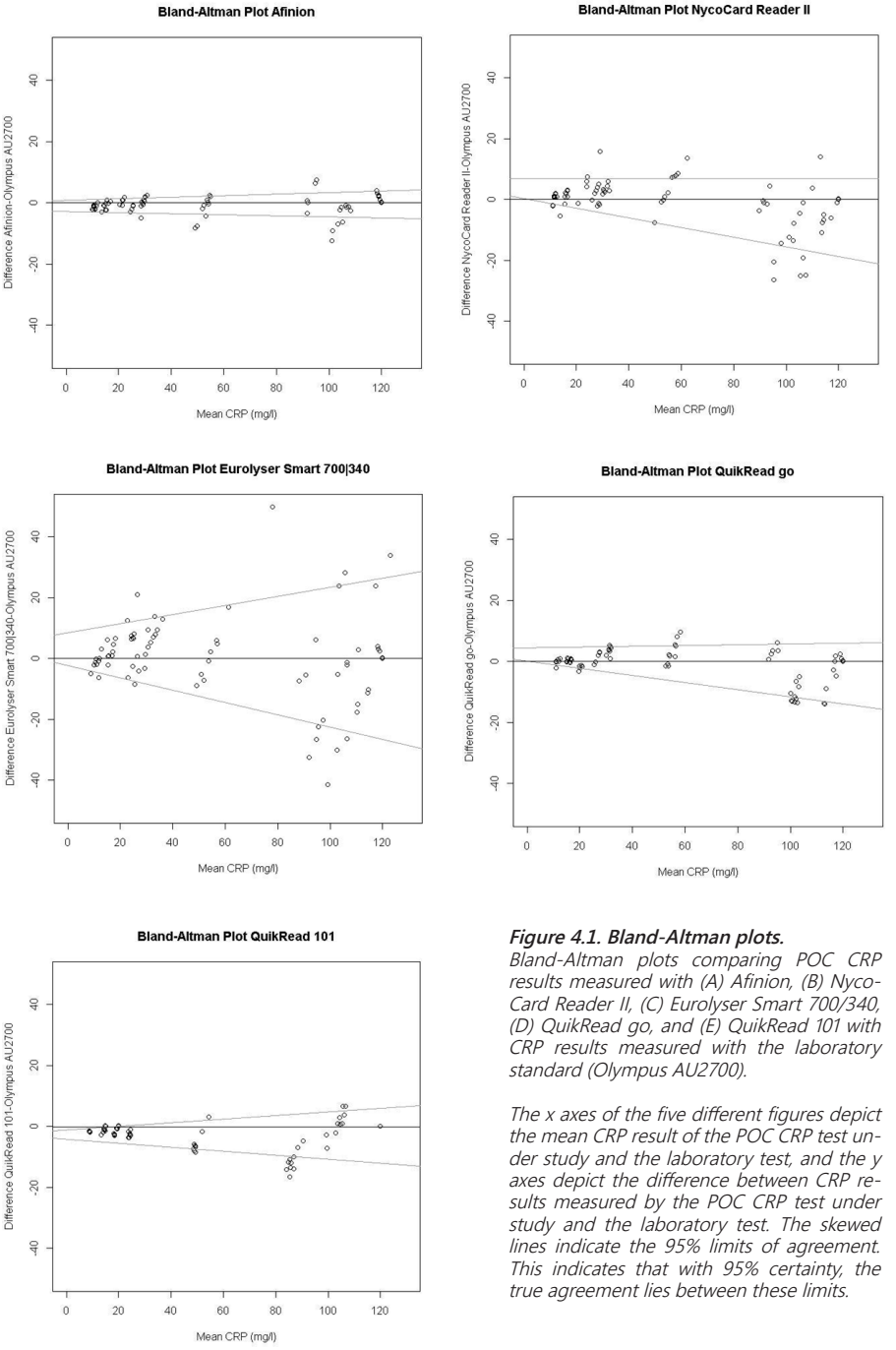


Figure 4.1. Bland-Altman plots.
Bland-Altman plots comparing POC CRP results measured with (A) Afinion, (B) NycoCard Reader II, (C) Eurolyser Smart 700/340, (D) QuikRead go, and (E) QuikRead 101 with CRP results measured with the laboratory standard (Olympus AU2700).

The x axes of the five different figures depict the mean CRP result of the POC CRP test under study and the laboratory test, and the y axes depict the difference between CRP results measured by the POC CRP test under study and the laboratory test. The skewed lines indicate the 95% limits of agreement. This indicates that with 95% certainty, the true agreement lies between these limits.

Table 4.3. Detailed information on five point-of-care C-reactive protein tests.

	Afinion	NycoCard Reader II	Eurolyser	QuikRead go	QuikRead 101
Device characteristics					
Size of the analyser, length x width x height, in cm, weight	17 x 32 x 17 5 kg	17 x 20 x 10 0.5 kg	14 x 14.5 x 26 3.4 kg	28 x 16 x 16 1.5 kg	14 x 22 x 8 1 kg
Connectivity [a], yes/no	Yes	No	Yes	Yes	No
Print function, yes/no	Yes	No	Yes	Yes	No
Storage function [b], yes/no	Yes	No	Yes	Yes	No
Test kit storage time	2-8ºC: expiry date 15-25ºC: 6 weeks	2-8ºC: expiry date 15-25ºC: 6 weeks	2-8ºC: expiry date > 8ºC: no information	2-8ºC: expiry date 18-25ºC: 3 months	2-8ºC: expiry date 18-25ºC: 3 months
Sample storage time	No information	Sample with dilution liquid: 1 hour	No information	2-8ºC: 3 days > 8ºC: 2 hours	2-8ºC: 3 days > 8ºC: 2 hours
Ability to determine other biomarker(s), yes/no, n, specification	Yes, 6, HbA1c, Albumin/Creatinin ratio (ACR), Lipid panel (HDL, LDL, Triglycerides, Cholesterol)	Yes, 3, HbA1c, D-dimer, U-Albumin	Yes, 9, high sensitive CRP, PT-INR, D-dimer, HbA1c, microalbumin, lipoprotein(a), homocystein, fecal occult blood (FOB), ferritin	No [c]	Yes, 3, strep A, fecal occult blood (FOB), U-Albumin
Retail price, analyser, €	>3000	<1000	>3000	1000-1500	1000-1500
Retail price, test kit, €	5.13 [76.98/15 units]	3.67 [88.14/24 units]	5.10 [163.40/32 units]	3.30 [166/50 units]	3.18 [159/50 units]
Test performance time **Time required for:**					
Analyser warm-up	4 min	25 sec	5 sec	50 sec	30 sec
Blank measurement	-	15 sec	-	-	15 sec
Preparation for analysis [d]	35 sec	3 min 20 sec	50 sec	30 sec	1 min 50 sec
Analysis	3 min 45 sec	3 min	4 min 20 sec	3 min 30 sec	1 min
Calibration (=blank measurement), yes/no, time	No, automatic	Yes, 15 sec	No, automatic	No, automatic	Yes, 15 sec
Total time (with analyser warm-up)	8 min 15 sec	7 min 15 sec	5 min 15 sec	4 min 50 sec	3 min 50 sec
Total time (without analyser warm-up)	4 min 15 sec	6 min 50 sec	5 min 10 sec	4 min	3 min 20 sec
Liability to flaws in procedure, small/moderate/large, %					
Blood application on test kit	100/0/0	8/25/67	44.5/44.5/11	25/50/25	17/75/8
Buffer application on test kit	75/25/0	0/58/42	67/22/11	25/50/25	17/67/17
Test kit placement in analyser	100/0/0	50/25/25	100/0/0	67/33/0	58.5/33/8.5
Loss of material	100/0/0	16.5/16.5/57	89/11/0	42/50/8	25/67/8
Overall liability to flaws, most frequent answer, small/moderate large	small	large	small	moderate	moderate

[a] Connectivity refers to the possibility of the device to be connected to electronic patient files, [b] Storage function refers to the possibility to store patient data on the device itself, [c] Not yet but is announced, [d] Time varies with experience

can test other biomarkers in addition to CRP, such as hemoglobin A1c (HbA1c) (Eurolyser Smart 700|340, Afinion, NycoCard Reader II) and D-dimer (NycoCard Reader II, Eurolyser Smart 700|340). Retail prices for the analyzers vary between approximately 1000 (NycoCard Reader II) and 3500 euro (Eurolyser Smart 700|340). Retail prices for test kits vary between approximately 3 (QuikRead 101, QuikRead go) and 6 euro (Afinion, Eurolyser Smart 700|340) per test kit.

Test performance time. Analyser warm-up time varied considerably from 5 seconds (Eurolyser Smart 700|340) to 4 minutes (Afinion). Preparation for analysis varies with experience. Subsequently estimated time required for complete analysis (without analyser warm up) varied approximately from approximately 3 minutes (QuikRead 101) to 7 minutes (NycoCard Reader II).

Susceptibility to flaws. The 20 GPs and GPAs considered Afinion and Eurolyser Smart 700|340 as least susceptible to flaws in the procedures. Using the Afinion test kit the chance of flaws in blood application was smallest. There was no need to perform a separate action to add buffer in the test kit in both Afinion and Eurolyser. Test kits for Afinion, QuikRead go and QuikRead 101 could only be placed in the analyser in one way, which evidently led to the smallest chance of flaws. With Afinion loss of sample material was least likely. On the Afinion and Eurolyser Smart 700|340 analyzers test performance took fewest separate actions.

Discussion

In our study on five POC CRP tests the Afinion, NycoCard Reader II, QuikRead go, and QuikRead 101 systems showed adequate analytical performance and agreement with the laboratory standard. For all CRP POC tests, however, agreement with the reference test decreased at higher CRP ranges (>100mg/L). Regarding user-friendliness, Afinion and Eurolyser Smart 700|340 had the smallest susceptibility to flaws in test procedures.

This is the first study reporting on analytical performance and user-friendliness of five different POC CRP tests. The results enable physicians to make a deliberate choice between different available devices.

There are a few limitations to our study, which should be taken into account. First, in our study all POC CRP tests were performed in a laboratory setting. This laboratory setting might not be representative of the general practice setting in which the POC CRP test will eventually be used. Therefore, future research should focus on the robustness of performance of the different POC CRP tests after implementation in the primary care setting. Secondly, instead of blood obtained by finger prick, we used pools of lithium heparin plasma samples to validate the POC CRP tests. In our opinion, this method was sufficient for our goal to assess analytical performance and general agreement between various POC CRP tests and the laboratory test.[28] As a next step, future studies using capillary blood obtained by finger prick are needed to assess differences between different test devices in daily practice are warranted. Moreover to provide well-grounded sensitivity, specificity, positive and negative predictive values of

POC tests for relevant clinical conditions, further studies using patient data including diagnostic outcome will be necessary. Thirdly, due to sample storage time and device delivery time it was indispensable to create a second set of lithium heparin plasma pools to evaluate the QuikRead 101. Theoretically this may have affected the results, but since the levels of the two different pools of lithium heparin plasma samples were comparable we expect the differences to be minimal. Finally not all devices are able to detect the same range of CRP values. Afinion and QuikRead 101 allow for measurement >120 mg/L while the other devices do not, and this influences the obtainable results. In practice the maximum measurable level of CRP may result in an underestimation of CRP level. The clinical impact of such underestimation regarding a high threshold of 120 mg/L is likely to be small, since a high CRP (>100 mg/L) will often make physicians more inclined to initiate antibacterial treatment in LRTI. Yet the exact relevant clinical cut off points may vary by disorder or complaint and evidence for value of these cut off points is limited. Implications of errors of all devices in daily practice are therefore still uncertain and should be explored in further research.

Although we found no previous studies on the performance of Afinion, Eurolyser Smart 700|340, and QuikRead go, there were several studies evaluating the analytical performance of the NycoCard Reader II and QuikRead 101,[12-18, 31, 32] of which only two studies directly compared the two systems.[15, 18] Monteny *et al.* studied febrile children in primary practice[15] and Zecca *et al.* included infants with suspected sepsis.[18] Their results compare well to the overall agreement found in the present study. Also, both previous studies showed that agreement decreases with higher CRP concentrations.

The results of this study show that there is considerable variation in agreement between the different CRP POC test in comparison with a laboratory reference standard, yet that most POC systems perform adequately. The choice of individual physicians for a POC CRP, however, can vary according to clinical context and personal preferences for certain test characteristics and also costs will probably be important when deciding which test to use.

References

1. de Wit NJ, Berger MY, Vogelenzang R, et al. [Summary of the Dutch College of General Practitioners (NHG) practice guideline on 'Diverticulitis']. Ned Tijdschr Geneeskd 2012;156:A4140.
2. Verlee L, Verheij TJ, Hopstaken RM, et al. [Summary of NHG practice guideline 'Acute cough']. Ned Tijdschr Geneeskd 2012;156:A4188.
3. Cals JW, Butler CC, Hopstaken RM, et al. Effect of point of care testing for C reactive protein and training in communication skills on antibiotic use in lower respiratory tract infections: cluster randomised trial. BMJ 2009;338:b1374.
4. Hopstaken RM, Muris JW, Knottnerus JA, et al. Contributions of symptoms, signs, erythrocyte sedimentation rate, and C-reactive protein to a diagnosis of pneumonia in acute lower respiratory tract infection. Br J Gen Pract 2003;53:358-64.
5. Andersson M, Andersson RE. The appendicitis inflammatory response score: a tool for the diagnosis of acute appendicitis that outperforms the Alvarado score. World J Surg 2008;32:1843-9.
6. Kaser SA, Fankhauser G, Glauser PM, et al. Diagnostic value of inflammation markers in predicting perforation in acute sigmoid diverticulitis. World J Surg 2010;34:2717-22.
7. Engel MF, Paling FP, Hoepelman AI, et al. Evaluating the evidence for the implementation of C-reactive protein measurement in adult patients with suspected lower respiratory tract infection in primary care: a systematic review. Fam Pract 2012;29:383-93.
8. Falk G, Fahey T. C-reactive protein and community-acquired pneumonia in ambulatory care: systematic review of diagnostic accuracy studies. Fam Pract 2009;26:10-21.
9. van der Meer V, Knuistingh Neven A, van den Broek PJ, et al. Diagnostic value of C reactive protein in infections of the lower respiratory tract: systematic review. BMJ 2005;331:26.
10. http://madox.org/horizon-scanning-reports/20110017/point-of-care-test-poct-for-c-reactive-protein-crp
11. Cals JW, Ament AJ, Hood K, et al. C-reactive protein point of care testing and physician communication skills training for lower respiratory tract infections in general practice: economic evaluation of a cluster randomized trial. J Eval Clin Pract 2011;17:1059-69.
12. Dahler-Eriksen BS, Lassen JF, Petersen PH, et al. Evaluation of a near-patient test for C-reactive protein used in daily routine in primary healthcare by use of difference plots. Clin Chem 1997;43:2064-75.
13. Dinant GJ, Costongs R, Leclercq RM, et al. Reliability of C-reactive protein measurement in general practice in The Netherlands. Scand J Clin Lab Invest 1994;54:113-7.
14. Esposito S, Tremolati E, Begliatti E, et al. Evaluation of a rapid bedside test for the quantitative determination of C-reactive protein. Clin Chem Lab Med 2005;43:438-40.
15. Monteny M, ten Brinke MH, van Brakel J, et al. Point-of-care C-reactive protein testing in febrile children in general practice. Clin Chem Lab Med 2006;44:1428-32.
16. Seamark DA, Backhouse SN, Powell R. Field-testing and validation in a primary care setting of a point-of-care test for C-reactive protein. Ann Clin Biochem 2003;40:178-80.

17. Urdal P, Borch SM, Landaas S, et al. Rapid immunometric measurement of C-reactive protein in whole blood. Clin Chem 1992;38:580-4.

18. Zecca E, Barone G, Corsello M, et al. Reliability of two different bedside assays for C-reactive protein in newborn infants. Clin Chem Lab Med 2009;47:1081-4.

19. http://www.skup.dk/20130404/flx/aproevninger/?mode=liste&category=12.

20. Lippi G, Ippolito L, Russello T, et al. Analytical performance of the new ACL AcuStar HemosIL D-Dimer. Blood Coagul Fibrinolysis 2012;23:164-7.

21. Reix N, Massart C, Gasser F, et al. Should functional sensitivity of a new thyroid stimulating hormone immunoassay be monitored routinely? The ADVIA Centaur(R) TSH3-UL assay experience. Clin Biochem 2012.

22. Töpfer G, Lutze G, Sauer K, et al. Einfluß der Citratkonzentration in Blutentnahmeröhrchen auf hämostaseologische Meßgrößen. J Lab Med 2001;25:162-8.

23. Zhang DJ, Elswick RK, Miller WG, et al. Effect of serum-clot contact time on clinical chemistry laboratory results. Clin Chem 1998;44:1325-33.

24. Clinical and Laboratory Standards Institute (CLSI). Evaluation of precision performance of clinical chemistry devices; approved guideline. NCCLS Document EP5-A. Wayne (PA): CLSI;1992. Wayne (PA): CLSI 1992.

25. Clinical and Laboratory Standards Institute (CLSI). Method comparison and bias estimation using patient samples; approved guideline. CLSI Document EP9-A. Wayne (PA): CLSI 1992.

26. Clark GH, Fraser CG. Biological variation of acute phase proteins. Ann Clin Biochem 1993;30 (Pt 4):373-6.

27. Macy EM, Hayes TE, Tracy RP. Variability in the measurement of C-reactive protein in healthy subjects: implications for reference intervals and epidemiological applications. Clin Chem 1997;43:52-8.

28. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet 1995;346:1085-7.

29. Dewitte K, Fierens C, Stockl D, et al. Application of the Bland-Altman plot for interpretation of method-comparison studies: a critical investigation of its practice. Clin Chem 2002;48:799-801.

30. Geersing GJ, Toll DB, Janssen KJ, et al. Diagnostic accuracy and user-friendliness of 5 point-of-care D-dimer tests for the exclusion of deep vein thrombosis. Clin Chem 2010;56:1758-66.

31. Hjortdahl P, Landaas S, Urdal P, et al. C-reactive protein: a new rapid assay for managing infectious disease in primary health care. Scand J Prim Health Care 1991;9:3-10.

32. Hobbs FD, Kenkre JE, Carter YH, et al. Reliability and feasibility of a near patient test for C-reactive protein in primary care. Br J Gen Pract 1996;46:395-400.


5

Chapter 5

The added diagnostic value of five different C-reactive protein point-of-care test devices in detecting pneumonia in primary care: a nested case-control study

Margaretha C. Minnaard
Alma C. van de Pol
Joris A.H. de Groot
Niek J. de Wit
Rogier M. Hopstaken
Sanne van Delft
Herman Goossens
Margareta Ieven
Christine Lammens
Paul Little
Chris C. Butler
Berna D.L. Broekhuizen
Theo J.M. Verheij

Scand J Clin Lab Invest. 2015 Jul;75(4):291-5

Abstract

Background

The results obtained from various point-of-care (POC) test devices for estimating C-reactive protein (CRP) levels in a laboratory setting differ when compared to a laboratory reference test. We aimed to determine whether such differences meaningfully affect the accuracy and added diagnostic value in predicting radiographic pneumonia in adults presenting with acute cough in primary care.

Methods

A nested case control study of adult patients presenting with acute cough in 12 different European countries (the GRACE Network). Venous blood samples from 100 patients with and 100 patients without pneumonia were tested with five different POC CRP tests and a laboratory analyser. Single test accuracy values and the added value of CRP to symptoms and signs were calculated.

Results

Single test accuracy values showed similar results for all five POC CRP tests and the laboratory analyser. The area under the curve of the different POC CRP tests and the laboratory analyser (range 0.79-0.80) were all comparable and higher than the clinical model without CRP (0.70). Multivariable odds ratios were the same (1.2) for all CRP tests.

Conclusions

Five POC CRP test devices and the laboratory analyser performed with similar accuracy in detecting pneumonia both as single test, and when used in addition to clinical findings. Variability in results obtained from standard CRP laboratory and POC test devices do not translate into clinically relevant differences when used for prediction of pneumonia in patients with acute cough in primary care.

Introduction

C-reactive protein (CRP) is an acute phase protein synthesized by the liver in response to inflammation. CRP helps identify patients at high or low risk of pneumonia.[1-6] CRP point of care (POC) testing is attractive to clinicians as results can be available within four minutes and thus results can alter management decisions during the patient's first presentation. Currently, there are several POC CRP devices available, but studies directly comparing these systems are scarce. In a comparative laboratory study using anonymous routine blood samples we found variability between the POC CRP devices compared to results obtained from a conventional laboratory analyser.[7] However, the extent to which variability of POC CRP test results translates into differences in accuracy to detect the clinical target condition pneumonia is unknown. Furthermore, accuracy of the added value of test results in addition to clinical evaluation alone should be considered.[8] We therefore set out to determine and compare the diagnostic test accuracy of five available POC CRP test devices and a laboratory test device for pneumonia in samples obtained from patients presenting to primary care with acute cough as part of a prospective cohort initiated by the GRACE (Genomics to combat Resistance against Antibiotics in Community-acquired LRTI in Europe) network (www.grace-lrti.org).[9]

Methods

Design

The present diagnostic accuracy study was conducted using a nested case-control design.[10] From the GRACE networks' cohort we randomly selected 100 patients with pneumonia present (cases) and 100 patients with pneumonia absent (controls), as categorized by chest radiographs (case control ratio 1:1). Equal number of cases and controls is the most efficient distribution when evaluating accuracy. With a total sample size of 200, differences in area under the receiver operating characteristic (ROC) curve (AUC) of the magnitude (depending on correlation) of 0.05 could be detected with a power of 80%.

Setting and study population

Adult out-patients presenting with acute cough were included in the GRACE study between 2007 and 2010 (www.grace-lrti.org).[9] The GRACE Network consisted of 16 primary care national research networks with 294 general practitioners (GPs) in 12 different European countries, details on study patients are described elsewhere.[6,9] Medical ethics committees of the participating study sites approved the study. All recruited patients provided written informed consent.

Data-collection

Venous blood samples were taken on day one (the day of presentation) and were transported to and stored at -80°C in the central laboratory in Antwerp. Conventional labo-

ratory CRP values were measured with Vitros 5.1 FS (Ortho Diagnostics; 2007-2009) and Dimension Vista Systems (Siemens; 2009 onward)[6] further referred to as laboratory analyser. POC CRP levels were determined using all five quantitative POC CRP test systems suitable for serum and capillary blood detection commercially available in the Netherlands in 2011: "Afinion", "NycoCard Reader II" (both Alere/Axis Shield), "Eurolyser Smart 700|340" (Eurolyser), "QuikRead Go" and "QuikRead 101" (both Orion Diagnostica) as previously described.[7] All tests were performed according to the manufacturer's instructions (March 2012) in a laboratory setting by laboratory analysts blind to patient characteristics. Chest radiographs were taken within seven days of the initial presentation to determine whether the patients had pneumonia (presence or absence). Results were registered on a standardized study form by local service radiologists blind to patients' clinical information (but not to previous radiographs).[6] Pneumonia was considered present if the radiologist recorded 'lobar pneumonia or bronchopneumonia'. All other outcomes on the standardized study form (e.g. 'acute bronchitis' or 'normal x-ray') were categorized as 'pneumonia absent'. Recruiting GPs systematically documented information on patient history and physical examination. A previously developed clinical diagnostic symptoms and signs model ("basic model") was fitted with the following diagnostic items yielding optimal results: absence of runny nose, breathlessness, crackles and diminished breath sounds on auscultation, tachycardia (>100/min), and temperature >37.8°C.[6]

Data-analysis

Demographic variables of the GRACE population and our sub sample were summarized using descriptive statistics. Mean and standard deviations were used for continuous demographic variables. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of the five different POC CRP tests for pneumonia were calculated, using most commonly used thresholds of 20 and 100 milligram/liter (mg/L).[11-14] Because a nested case-control study design was used, controls and cases were weighted by the sampling fraction to obtain valid estimates of the various diagnostic accuracy measurements.[10] Thus, by multiplying the number of cases by (total with pneumonia/100, i.e. the sampling fraction) and the number of controls by (total without pneumonia/100, i.e. the sampling fraction) new 2 by 2 tables were constructed to calculate corrected sensitivity, specificity, PPV and NPV. Also, univariate odds ratios (ORs) with 95% confidence intervals (95% CIs) were calculated using logistic regression modelling. The added value of the CRP tests to symptoms and signs model ("basic model"; for details see box 1) was assessed by ROC curve analysis. Also, multivariate ORs (with 95% CIs) for the different POC CRP tests were calculated. Furthermore, as a general measure of discrimination, the area under the curve (AUC) was calculated for both the basic model, and for the basic model plus continuous POC CRP test values (extended models 1-6). Data were analysed using SPSS (Version 20.0 for Windows) and R (Version 2.15.0).

Box 5.1: Basic and extended model

Basic model = symptoms and signs model, defined as 1/(1 + exp-(-3.984 + 0.446 * breathlessness + 0.698 * absence of runny nose + 0.596 * diminished vesicular breathing + 1.404 * crackles + 0.961 * tachycardia + 0.980 * temperature>37.8°C))

Extended model = basic model complemented with CRP measured with laboratory analyser.

Results

Descriptive statistics

Patient characteristics of the total GRACE cohort and the case-cohort sample are shown in table 5.1. There were no differences between the used sample and the total GRACE cohort in patient characteristics, apart from the (set) distribution of pneumonia present and absent.

Table 5.1: Baseline characteristics.
Baseline characteristics of diagnostic variables in 2820 patients presenting with acute cough in primary care (the "GRACE" cohort) and a case control sample (n=200). Figures are numbers (percentages) unless mentioned otherwise.

Diagnostic variable	GRACE[a] cohort no pneumonia n=2680	GRACE[a] ohort pneumonia n=140	Sample no pneumonia n=100	Sample pneumonia n=100
Mean age, years (SD)	50 (17)	54 (15)	51 (15)	54 (16)
Male gender, n (%)	1066 (40)	62 (44)	43 (43)	45 (45)
CRP[b], mg/L (SD)	17 (28)	69 (83)	19 (28)	62 (81)
History of breathlessness present, n (%)	1498 (56)	96 (69)	53 (53)	66 (66)
Diminished vesicular breathing (auscultation), n (%)	331 (12)	31 (22)	5 (5)	24 (24)
Crackles (auscultation), n (%)	220 (8)	45 (32)	6 (6)	34 (34)
Tachycardia (pulse > 100 bpm[b]), n (%)	94 (4)	17 (12)	6 (6)	11 (11)
Tachypnoea (respiratory rate > 24/min), n (%)	49 (2)	6 (4)	2 (2)	5 (5)
Body temperature (temp > 37.8 ºC), n (%)	136 (5)	22 (16)	3 (3)	14 (14)

[a] *Genomics to combat Resistance against Antibiotics in Community-acquired lower respiratory tract infection in Europe*
[b] *mean C-reactive protein as measured by the laboratory analyser; [c]beats per minute*

Single test accuracy values

Single test accuracy measures for the different POC CRP test analysers at thresholds of 20 and 100 mg/L are shown in table 5.2 and 5.3.

*T**able 5.2:** Single test accuracy measures at threshold of 20 mg/L.*

	Sensitivity,% (95% CI)	Specificity,% (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)
Laboratory Analyser	61.4 (53.2-69.1)	76.0 (74.3-77.5)	11.8 (9.6-14.3)	97.4 (96.6-98.0)
Afinion	55.0 (45.2-64.4)	73.0 (63.6-80.7)	9.6 (7.8-11.9)	96.9 (96.0-97.6)
NycoCard Reader II	54.0 (44.3-63.4)	75.0 (65.7-82.5)	10.1 (8.2-12.5)	96.9 (96.1-97.6)
Eurolyser Smart	48.0 (38.5-57.7)	79.0 (70.0-85.8)	10.7 (8.5-13.3)	96.7 (95.8-97.3)
QuikRead go	52.0 (42.3-61.5)	72.0 (62.5-79.9)	8.8 (7.1-11.0)	96.6 (95.7-97.3)
QuikRead 101	49.0 (39.4-58.7)	74.0 (64.6-81.6)	9.0 (7.1-11.2)	96.5 (95.6-97.2)

CI=confidence interval, PPV=positive predictive value, NPV=negative predictive value

***Table 5.3:** Single test accuracy measures at threshold of 100 mg/L.*

	Sensitivity,% (95% CI)	Specificity,% (95% CI)	PPV, % (95% CI)	NPV, % (95% CI)
Laboratory Analyser	24.3 (17.9-27.8)	97.7 (97.0-98.2)	35.4 (26.6-45.4)	96.1 (95.3-96.8)
Afinion	20.0 (13.3-28.9)	99.0 (94.6-99.9)	51.1 (38.2-63.8)	95.9 (95.1-96.6)
NycoCard Reader II	20.0 (13.3-28.9)	98.0 (93.0-99.4)	34.3 (24.9-45.1)	95.9 (95.1-96.6)
Eurolyser Smart	19.0 (12.5-27.8)	99.0 (94.6-99.9)	49.8 (36.9-62.8)	95.9 (95.1-96.6)
QuikRead go	20.0 (13.3-28.9)	99.0 (94.6-99.9)	51.1 (38.2-63.8)	95.9 (95.1-96.6)
QuikRead 101	19.0 (12.5-27.8)	99.0 (94.6-99.9)	49.8 (36.9-62.8)	95.9 (95.1-96.6)

CI=confidence interval, PPV=positive predictive value, NPV=negative predictive value

***Table 5.4:** Single test accuracy measure: odds ratio with 95% confidence intervals.*

	Univariate OR (95% CI)
Laboratory Analyser	1.2 (1.1-1.3)
Afinion	1.2 (1.1-1.3)
NycoCard Reader II	1.2 (1.1-1.4)
Eurolyser Smart	1.2 (1.1-1.4)
QuikRead go	1.2 (1.1-1.3)
QuikRead 101	1.2 (1.1-1.3)

OR=odds ratio, CI=confidence interval

All single test accuracy measures showed similar results for all different CRP analysers, both POC and laboratory analysers. Sensitivity decreased at high threshold (100 mg/L). Specificity and PPV increased at high threshold (100 mg/L). Values of NPV showed the same results for all POC tests on both thresholds. Univariate ORs were comparable for the different POC CRP test analysers and the CRP laboratory analyser (table 5.4).

Added value of POC CRP testing

The symptoms and signs model without CRP has an area under the ROC curve (AUC) of 0.70 (with 95% CI; 0.65-0.75).[6] Figure 1 shows the ROC curves for both the basic symptoms and signs model (basic model) and the basic model extended with the different CRP POC tests (extended model 2 to 6). AUCs are shown in table 5.5 for the basic model and the various extended models. The increase in AUC for all extended models shows, that all POC CRP tests have limited, but added predictive value which is equal for all POC tests and comparable to CRP measured with laboratory analysers.

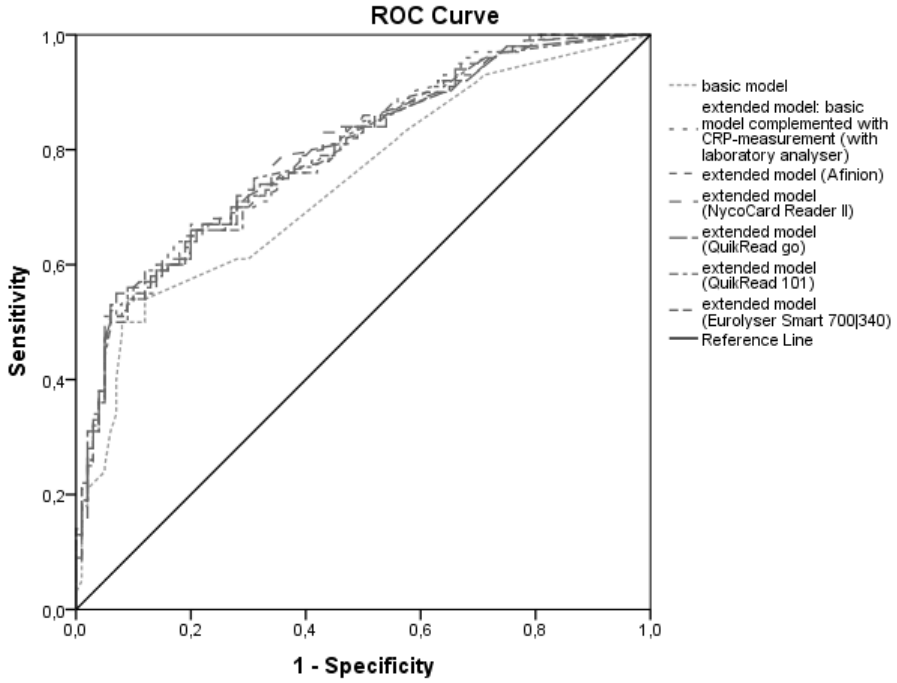


Figure 5.1: ROC curves.
Multivariate receiver operating characteristic (ROC) curves of basic symptoms and signs model and extended models (of laboratory analyser and five different point-of-care C-reactive protein tests.

Table 5.5: *Area under the curve and multivariate odds ratios with 95 % confidence interval for the basic symptoms and signs model and 6 extended models.*

	AUC (95% CI)	Multivariate OR (95% CI)
Basic model[a] (symptoms and signs)	0.70 (0.65-0.75)	-
Extended model = Basic model complemented with CRP, measured with:		
(1) laboratory analyser[b]	0.79 (0.75-0.83)	1.2 (1.1-1.3)
(2) Afinion	0.79 (0.73-0.85)	1.2 (1.1-1.3)
(3) NycoCard Reader II	0.80 (0.74-0.86)	1.2 (1.1-1.4)
(4) Eurolyser Smart 700|340	0.80 (0.74-0.86)	1.2 (1.1-1.3)
(5) QuikRead go	0.79 (0.73-0.86)	1.2 (1.1-1.4)
(6) QuikRead101	0.79 (0.73-0.85)	1.2 (1.1-1.3)

AUC=area under the curve, CI=confidence interval, OR=odds ratio

Discussion

Main findings

An efficient case cohort design using the large GRACE dataset enabled us to determine the accuracy and added value of five POC CRP test devices and conventional laboratory CRP test for pneumonia, which did not meaningfully differ, nor as single test neither as added test above clinical findings.

Strengths and limitations of this study

This is the first study to quantify added diagnostic value of results obtained from five different POC CRP test devices in predicting radiographic pneumonia in a study population with acute cough in primary care, the study population realistically representing usual primary care patients with acute cough. Previous studies compared accuracy of POC CRP tests with a laboratory standard, while our study used a clinically relevant outcome, i.e. pneumonia. Some limitations of our study warrant further discussion. First, a possible limitation of the study includes the use of chest radiograph as reference standard for pneumonia as this may have led to misclassification, a limitation encountered in all similar pneumonia studies.[6] We used a formal diagnostic model to demonstrate the value of POC CRP when added to symptoms and signs. This diagnostic model may not completely represent clinical judgement and could therefore have resulted in an over- but also an underestimation of the added value of POC CRP when added to clinical judgement. However, a formal diagnostic model is more reproducible and objective than individual clinical assessment and therefore as a first step more appropriate to determine the added value of POC CRP above symptoms and signs. Also, in our study all actual tests were done in a laboratory setting which might not be representative of the general

practice setting in which the POC CRP tests will eventually be used, although in our opinion, this method was sufficient for our goal to compare five different POC tests mutually.

Interpretation of findings in relation to previously published work

Previously, several other studies evaluated the analytical performance of different POC CRP test devices, but only one study directly compared two different POC CRP test devices in primary practice and unlike in our study, they found variability between the different devices.[15] The main difference between that and our study was the study population. They evaluated the two POC CRP tests in febrile children whereas our study results were obtained in adults with acute cough, which might have led to the difference in result. In a previously conducted comparative laboratory study using anonymous routine blood samples we found considerable variation in agreement between the different POC CRP tests in comparison with a laboratory reference standard based on calibration.[7]

Implications for future research, policy and practice

In the present study we demonstrated that, if plasma samples from patients with acute cough and chest radiographs as reference test are used, all five POC CRP test devices and the laboratory analyser showed similar result. Thus the variability found in the earlier study[7] does not translate into differences in diagnostic accuracy to detect the clinical target condition pneumonia. Future research should focus on the robustness of performance of the different POC CRP tests after implementation in the primary care setting.

Conclusions

The results of this study show GPs that the five POC CRP test devices and the laboratory analyser are equally suitable in predicting radiographic pneumonia both as single test and when added to signs and symptoms.

References

1. Almirall J, Bolibar I, Toran P, et al. Contribution of C- reactive protein to the diagnosis and assessment of severity of community-acquired pneumonia. Chest 2004;125:1335-42.
2. Flanders SA, Stein J, Shochat G, et al. Performance of a bedside C-reactive protein test in the diagnosis of community-acquired pneumonia in adults with acute cough. Am J Med 2004;116:529-35.
3. Holm A, Pedersen SS, Nexoe J, et al. Procalcitonin versus C-reactive protein for predicting pneumonia in adults with lower respiratory tract infection in primary care. Br J Gen Pract 2007;57:555-60.
4. Melbye H, Straume B, Brox J. Laboratory tests for pneumonia in general practice: the diagnostic values depend on the duration of illness. Scand J Prim Health Care 1992;10:234-40.
5. Pepys MB, Hirschfield GM. C-reactive protein: a critical update. J Clin Invest 2003; 111(12):1805-1812.
6. van Vugt SF, Broekhuizen BD, 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:f2450.
7. Minnaard MC, van de Pol AC, Broekhuizen BD, et al. Analytical performance, agreement and user-friendliness of five C-reactive protein point-of-care tests. Scan J Clin Lab Invest 2013;73:627-34.
8. Moons KG. Criteria for scientific evaluation of novel markers: a perspective. Clin Chem 2010;56:537-41.
9. Nuttall J, Hood K, Verheij TJ, et al. Building an international network for a primary care research program: reflections on challenges and solutions in the set-up and delivery of a prospective observational study of acute cough in 13 European countries. BMC Fam Pract 2011;12:78.
10. Biesheuvel CJ, Vergouwe Y, Oudega R, et al. Advantages of the nested case-control design in diagnostic research. BMC Med Res Methodol 2008;8:48.
11. Cals JW, Butler CC, Hopstaken RM, et al. Effect of point of care testing for C reactive protein and training in communication skills on antibiotic use in lower respiratory tract infections: cluster randomised trial. BMJ 2009;338:b1374.
12. Falk G, Fahey T. C-reactive protein and community-acquired pneumonia in ambulatory care: systematic review of diagnostic accuracy studies. Fam Pract 2009;26:10-21.
13. Hopstaken RM, Muris JW, Knottnerus JA, et al. Contributions of symptoms, signs, erythrocyte sedimentation rate, and C-reactive protein to a diagnosis of pneumonia in acute lower respiratory tract infection. Br J Gen Pract 2003;53:358-64.
14. Verheij TJM, Hopstaken RM, Prins JM, et al. The NHG-Guideline acute cough. Huisarts Wet 2011;68-92.
15. Monteny M, ten Brinke MH, van Brakel J, et al. Point-of-care C-reactive protein testing in febrile children in general practice. Clin Chem Lab Med 2006;44:1428-32.

Chapter 6

Effect of C-reactive protein point of care testing on antibiotic prescribing in general practice: an observational study

Margaretha C. Minnaard
Alma C. van de Pol
Rogier M. Hopstaken
Sanne van Delft
Berna D.L. Broekhuizen
Theo J.M. Verheij
Niek J. de Wit

Submitted for publication

Abstract

Background
In clinical trials the potential of point-of-care (POC) C-reactive protein (CRP) tests was demonstrated in decreasing antibiotic prescribing in adults with acute cough in general practice, but effects of implementation are unknown.

Objective
To determine the overall effect of POC CRP testing on antibiotic prescribing rate in general practice.

Methods
In an observational study general practitioners (GPs) were instructed to use POC CRP in adults with acute cough following current guidelines. After routine history taking and physical examination they reported whether they intended to prescribe antibiotics ("pre-test decision"). They reported their revised decision after receiving the POC CRP test result ("post-test decision"). Primary outcome was the difference between "pre-test" and "post-test" antibiotic prescribing rate. Secondary outcome was the percentage of patients for whom the decision to prescribe antibiotics did alter after CRP testing.

Results
A total of 40 GPs enrolled 939 patients, 78% of whom were tested for CRP. Antibiotic prescribing before and after CRP testing did not differ ("pre-test" 31%, "post-test" 28%; 95% confidence interval of difference -7 to 1). GPs changed their decision after POC CRP testing in 200 patients (27%). In 41% of the tested patients the indication for testing was in accordance with the guidelines.

Conclusion
POC CRP testing did not reduce overall antibiotic prescribing when this is used by GPs who already have a low antibiotic prescribing rate. POC CRP did affect GPs to change their decision about antibiotic prescribing in patients with acute cough. In daily primary care practice POC CRP may be overused.

Introduction

Acute cough is one of the main reasons for patients to consult their general practitioner (GP) worldwide.[1–3] In patients presenting with acute cough or other symptoms of lower respiratory tract infections (LRTIs), bronchitis is to be differentiated from community acquired pneumonia (CAP), since bronchitis is usually self-limiting while CAP requires treatment with antibiotics. To support the detection of CAP C-reactive protein (CRP) is increasingly used in the diagnostic workup in general practice. In patients in whom diagnostic uncertainty remains after clinical examination, CRP level has added diagnostic value to detect CAP.[4–6] Several randomized controlled trials (RCTs) and reviews have shown that CRP point-of-care (POC) testing, i.e. testing during GP consultation, can reduce antibiotic prescribing,[7–13] and is cost-effective.[14–16] Based on this evidence guidelines advise POC CRP testing for patients who present with acute cough suspected of CAP.[3,17] In RCTs, however, GPs are explicitly instructed on which patients should be tested and how CRP results should be interpreted. The actual impact of CRP implementation in routine GP practice, where POC CRP tests are used at the discretion of the GP without monitoring and strict protocols on use and interpretation, might differ from the trial situation. It is not known yet whether POC CRP test results affect the GP's decision to prescribe antibiotics or not in daily practice. The present study aims to determine the effect of POC CRP testing on antibiotic prescribing when implemented in general practice, where guidelines are already available. Firstly, we determined the overall effect of POC CRP on antibiotic prescribing rate among participants, and secondly we assessed how often the POC CRP test results did alter the GPs decision on antibiotic prescribing for individual patients. Finally, we assessed guideline adherence for testing and antibiotic prescribing.

Methods

Design

A prospective observational study.

Setting

The study was carried out in nine general practices in the Netherlands, four in the residential area of Rotterdam and five in a new residential area near Utrecht, with in total approximately 60.000 patients registered. The patient population is representative for the Dutch population.[1,18] Four of the practices involved in this study were academic research centers ("academic practices") and five were not routinely involved in research ("non-academic practices").

Study population

From February 2012 to February 2013, 40 participating GPs enrolled all patients aged 18 years and over, presenting in routine general practice with acute cough (less than 24 days). GPs provided care as usual, i.e. decided on diagnostic work-up (including possible POC CRP testing) and management according to their routine practice. GPs

did not receive formal additional training on indications and interpretations of POC CRP test results for study purposes, yet the research team did notify GPs of the existing professional guideline "Acute Cough" from the Dutch College of GPs (2011), in which indications for CRP testing are described.[17] The Act on Medical Research involving human subjects did not apply to this study and therefore an official approval of this study by the Medical Ethics Research Council of the University Medical Center Utrecht was not required.

Measurements

GPs reported characteristics, and their diagnostic and therapeutic management of all patients included on case report forms (CRFs): age, relevant comorbidities, fever reported by patients, auscultation abnormalities (defined by crackles or diminished vesicular breathing), tachypnea (>24 breaths/min), tachycardia (pulse>100/min), hypotension (blood pressure <90/60 mmHg), and mental state of confusion and drowsiness. Relevant comorbidity was defined as heart failure, diabetes mellitus, chronic obstructive pulmonary disease, asthma, impaired immunity, malignancy, severe neurologic illness, renal failure, congenital heart disease or congenital lung disease. On the CRF GPs registered whether they intended primarily to prescribe antibiotics ("pre- test" decision) before they decided whether or not to order a POC CRP test. The POC CRP test was performed by the practice assistant. CRP level measurement was carried out using Alere Afinion® AS100 POC CRP test devices[19] in accordance with the manufacturer's instructions. After disclosure of the POC CRP test result, GPs registered whether they actually prescribed antibiotics ("post-test" decision). Within 28 days follow-up the GPs reported for all included patients whether or not the patient presented for re-consultation, whether or not they had prescribed antibiotics during re-consultation and whether or not they had referred the patient to a pulmonologist because of suspected CAP.

Indications for POC CRP testing

Risk groups for CAP and indications for POC CRP testing were defined according to the guideline "Acute cough" from the Dutch College of GPs.[17] This is the main (consensus) guideline in Dutch general practice. It focuses on two steps in the diagnostic process. Firstly, patients who present with acute cough are classified into three risk categories based on signs and symptoms: [a] low, [b] intermediate and [c] high risk of complicated LRTI (for complete definition of these subgroups see figure 6.1). Secondly, recommendations are given for each risk group category on whether to test POC CRP and to prescribe antibiotics. For patients with acute cough in category [a] the guideline recommends to refrain from both POC CRP testing and antibiotics. For category [c] the guideline recommends not to test, but to prescribe antibiotics or refer to secondary care instantly. For category [b], the group with largest diagnostic uncertainty, the guideline recommends to perform a POC CRP test. If POC CRP results are <20 mg/L the advice is to refrain from antibiotics, for 20-100 mg/L clinical characteristics are leading

and decisive concerning prescribing antibiotics or not, and for results >100 mg/L it is recommended to start antibiotic treatment.

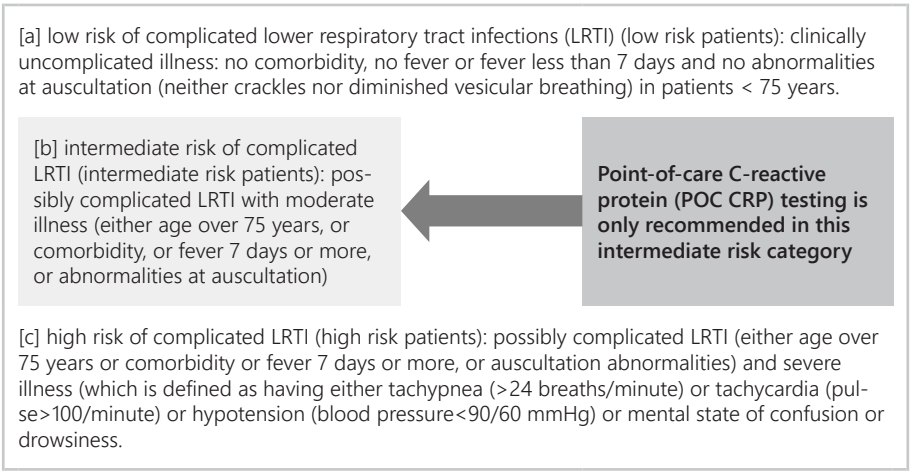


Figure 6.1: Risk categories
Risk categories of patients presenting with acute cough, defined in the national guideline of the Dutch College of General Practitioners[17]

Outcomes

The primary outcome was the difference between "pre-test" and "post-test" antibiotic prescribing rate. Secondary outcomes were the change in decisions of the GPs on antibiotic prescribing in individual patients "pre-test" compared to "post-test", and the direction of these changes ("pre-test yes" to "post-test no" and vice versa). Thirdly, we assessed guideline adherence for requesting POC CRP tests. If GPs requested a POC CRP test in intermediate risk (category [b]) patients, this was considered as adherent, i.e. patients had an indication for testing. POC CRP testing in low/high risk (category [a] and [c]) patients was considered as non-adherent. We also assessed the guideline adherence for antibiotic prescribing for the intermediate risk (category [b]) patients who had a POC CRP test requested that prompted the GP to alter his/her decision on antibiotic prescribing (from "pre-test: yes" to "post-test: no" or vice versa). A change from "pre-test: yes" to "post-test: no" was considered adherent for POC CRP results of <20 mg/L. Alternatively, a change from "pre-test: no" to "post-test: yes" was deemed adherent for POC CRP results >100 mg/L. For intermediate risk (category [b]) patients with POC CRP levels of 20-100 mg/L all changes were considered non-adherent, since for these cases the guideline recommends to rely predominantly on clinical findings; for the purpose of this study the "pre-test" antibiotic prescribing decision ("yes" or "no") was considered to reflect the GPs interpretation of clinical findings. Fourthly we summarized follow-up data as absolute numbers (counts) for all intermediate risk (category [b]) patients.

Data analysis

Clinical characteristics, POC CRP test results, guideline adherence for POC CRP, guideline adherence for antibiotic prescribing and follow-up data were analyzed using descriptive statistics. Percentages were computed to summarize categorical variables. Mean and standard deviations (SDs) were used for continuous variables. To test whether there was a difference between "pre-test" and "post-test" antibiotic prescribing rate the McNemar test was used. Post hoc subgroup analysis of possible differences in antibiotic prescribing rate (on group level) between GPs in academic practices and GPs from non-academic practices was performed. Five patients had missing CRP values and were left out of the analyses. Statistical packages SPSS for Windows 20.0.0 (SPSS, Inc. Chicago, Illinois, USA) were used.

Results

939 patients were included (mean age 47 (SD 15) years, 38% male). For 735 (78%) patients the GPs requested a POC CRP test. Demographics, symptoms, signs and comorbidity are presented in table 6.1.

Antibiotic prescribing

GPs intended to prescribe antibiotics in 290 (31%) patients, after performing the POC CRP test they actually prescribed antibiotics to 262 (28%) patients (difference -3% (95% CI: -7 to 1) (p=0.06; figure 6.2). On individual patient level POC CRP test results prompted changes in antibiotic prescribing decisions in 200 (27%) of all POC CRP tested patients. In 114 (16% of total, 57% of changes) of the tested patients the decision on antibiotic prescribing changed from "pre-test: yes" to "post-test: no" and in 86 (12% of total, 43 % of changes) vice versa (figure 6.3). Post hoc subgroup analysis of the antibiotic prescribing rate showed that in academic practices the "pre-test" antibiotic prescribing rate was 25% and the "post-test" antibiotic prescribing rate 28% (difference: 3% (95%CI -3 to 8)(p=0.19). For non-academic practices the prescribing rate was 38% "pre-test" and 27% "post-test" (difference: -11%, 95%CI -17 to 4)(p<0.001).

Guideline adherence

According to the professional guideline 335 (46%) of the POC CRP tested patients were to be categorized as being at low risk [a], 97 (13%) as being at high risk [c] and 303 (41%) as being at intermediate risk [b] of complicated LRTI. For the definition of these subgroups; see figure 6.1. The intermediate group is the only group for which the guideline recommends testing, hence 303 (41%) of indications were in accordance with the guideline. Details on CRP levels and antibiotic prescribing in patients with an intermediate risk [b] are shown in figure 6.3. For 93 (31%) of the 303 intermediate risk patients, the POC CRP result prompted the GP to change their decision on antibiotic prescribing. The change in decision on antibiotic prescribing was in line with recommendations of the guideline in 56 out of these 93 (60%) patients (figure 6.4).

Table 6.1: Baseline characteristics
Baseline characteristics of all included patients, all POC CRP tested patients and all non-tested patients.

	All included patients n=939	POC CRP tested patients n=735	Not-tested patients n=204
Mean age, years (SD)	47 (15)	47 (15)	48 (15)
Male gender	356 (38)	274 (37)	82 (40)
CRP, mean in mg/L (SD)	24 (33)	24 (33)	-
<20 mg/L	514 (55)	514 (70)	-
20-100 mg/L	190 (20)	190 (26)	-
>100 mg/L	31 (3)	31 (4)	-
HISTORY			
Breathlessness (reported)	379 (40)	321 (44)	85 (28)
Allergic to antibiotics	50 (5)	43 (6)	7 (3)
Fever (reported) 7 days or more	38 (4)	31 (4)	7 (3)
Comorbidity	326 (35)	264 (36)	62 (30)
Heart failure	16 (2)	12 (2)	4 (2)
Diabetes mellitus	72 (8)	49 (7)	23 (11)
COPD	74 (8)	59 (8)	15 (7)
Asthma	122 (13)	107 (15)	15 (7)
Immunocompromised	17 (2)	16 (2)	1 (0.5)
Malignancy	21 (2)	18 (2)	3 (2)
Neurology	14 (2)	11 (2)	3 (2)
Renal failure	8 (1)	7 (1)	1 (0.5)
Congenital	4 (0.5)	2 (0.3)	2 (1)
PHYSICAL EXAMINATION			
Abnormal auscultation (crackles or diminished vesicular breathing)	156 (17)	129 (18)	27 (13)
Tachycardia (pulse > 100 bpm)	82 (9)	62 (8)	20 (10)
Tachypnea (respiratory rate > 24/min)	59 (6)	43 (6)	16 (8)
Hypotension (SBP<90 and DBP <60 mmHg)	3 (0.3)	3 (0.4)	-
Confusional mental state	1 (0.1)	-	1 (0.5)
Sedation	2 (0.2)	1 (0.1)	1 (0.5)

All characteristics are shown as absolute numbers, between brackets in % unless otherwise specified.

Abbreviations: CRP: C-reactive protein, mg/L: milligram per liter, bpm: beats per minute, SBP: systolic blood pressure, DBP: diastolic blood pressure, mmHg: millimeters mercury.

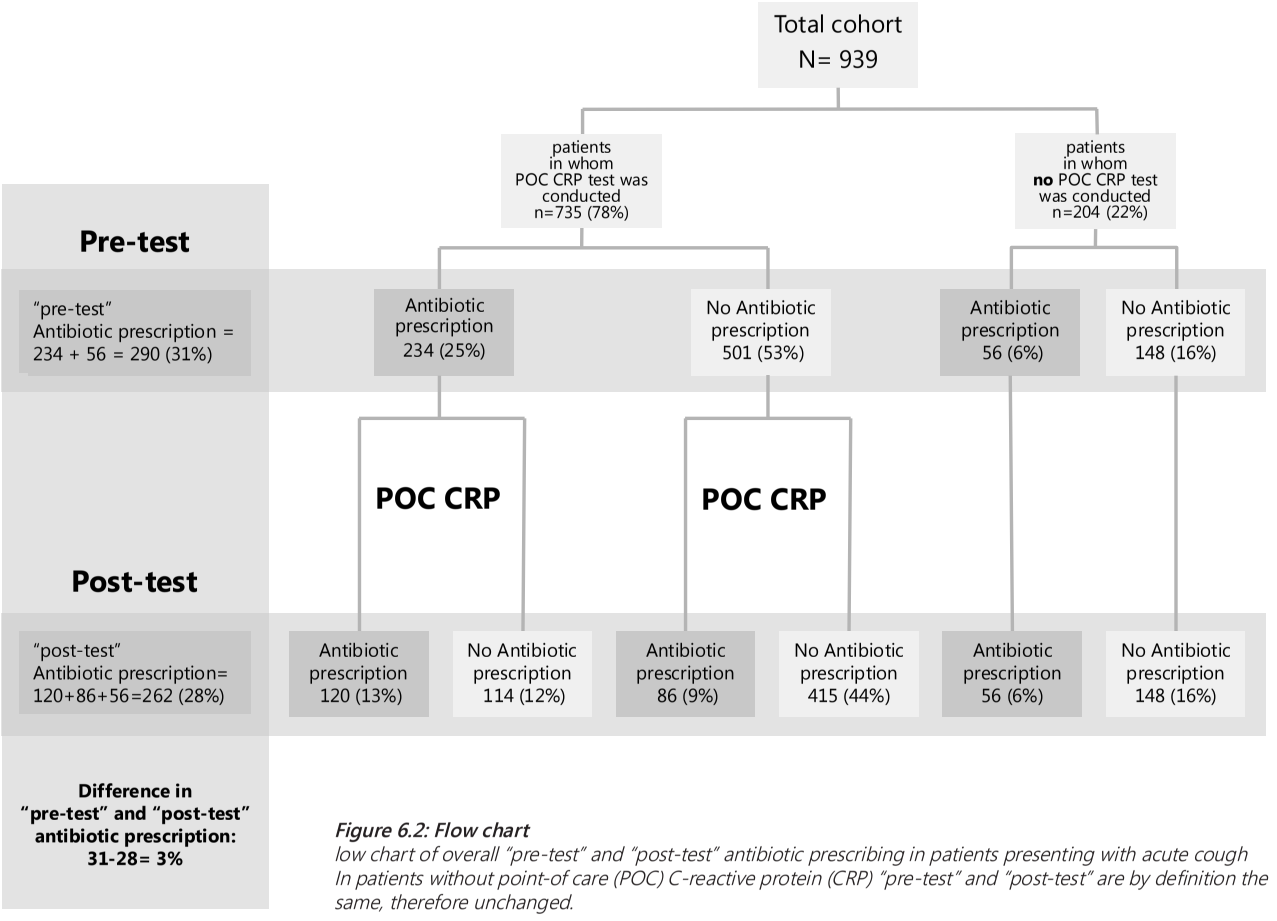


Figure 6.2: Flow chart
low chart of overall "pre-test" and "post-test" antibiotic prescribing in patients presenting with acute cough In patients without point-of care (POC) C-reactive protein (CRP) "pre-test" and "post-test" are by definition the same, therefore unchanged.

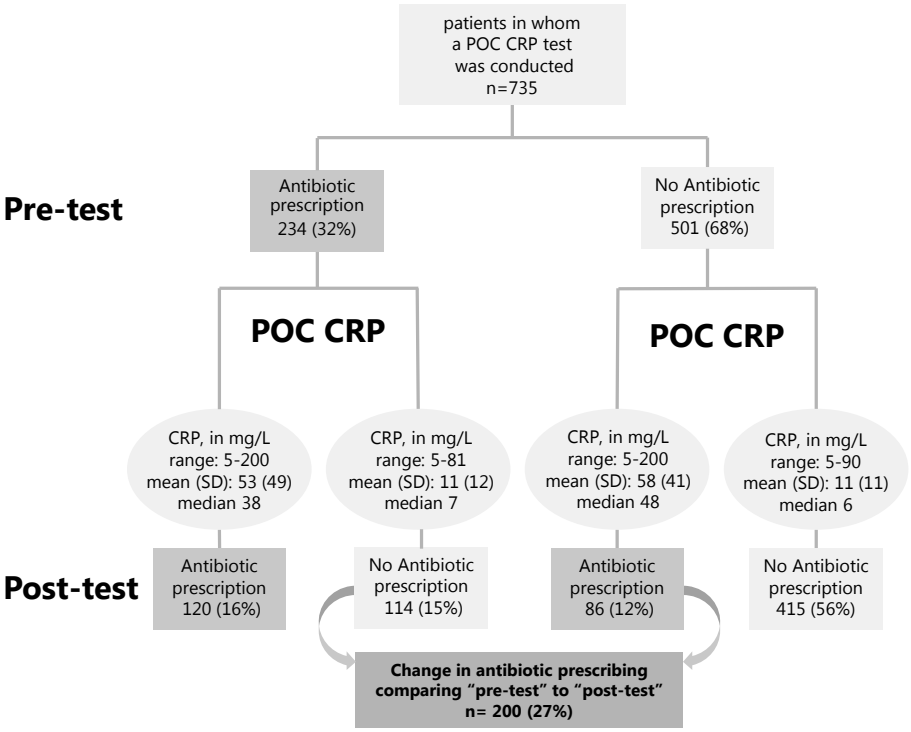
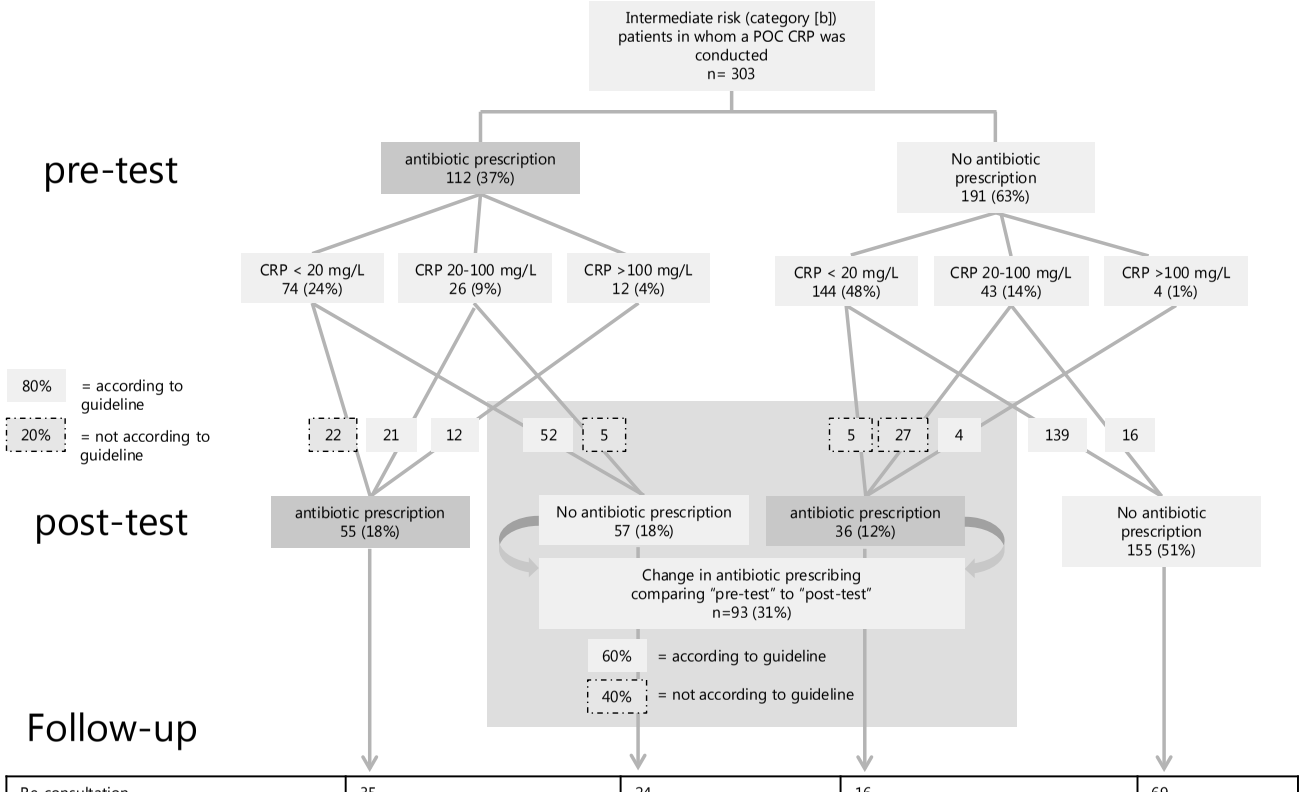


Figure 6.3: Flow chart
Flow chart of individual changes in "pre-test" and "post-test" antibiotic prescribing decisions for point-of-care C-reactive protein tested patients. Abbreviations: POC: point-of-care, CRP: C-reactive protein

Follow up

A total of 207 patients with intermediate risk [b] presented for re-consultation.144 presented for re-consultation, 48 received a (second) prescription of antibiotics at re-consultation and 5 were referred to the pulmonologist because of suspected CAP (figure 6.4).



Re-consultation	35	24	16	69
Prescription at reconsultation	4	10	2	32
Referral (because of suspected pneumonia)	2	1	0	2
missings	-	3	1	4

Figure 6.4 *(For caption see next page)*

Figure 6.4: Antibiotic prescribing decisions and follow up in intermediate risk patiens
Antibiotic prescribing decisions in relation to C-reactive protein levels in patients with intermediate risk of complicated lower respiratory tract infection. C-reactive protein (CRP) categories are depicted for patients with intermediate risk of complicated lower respiratory tract infection (LRTI; see figure 6.1), and related to the FPs decisions on antibiotic prescribing "pre-test" and "post-test" (i.e. before and after being made aware of the POC CRP test results). The intermediate risk group is the group of patient for whom POC CRP testing is recommended by the guideline "Acute cough" of the Dutch College of General Practitioners.[17] Please view the methods section of this chapter for detailed information on recommendations of the "Acute Cough" guideline regarding antibiotic prescribing according to different CRP categories.

Discussion

Summary of results

POC CRP testing in nine general practices in the Netherlands did not result in a clear drop in antibiotic prescribing for patients presenting with acute cough. 41% of POC CRP tests were indicated according to present guidelines. The decision to prescribe antibiotics in individual patients did change in 27% of tested patients. In the group of patients who were indicated for CRP testing according to the guidelines 40% of changes in decisions on antibiotic prescribing after testing was not according to guideline recommendations.

Strengths and limitations

Several limitations need to be addressed. Firstly, the antibiotic prescribing rate in our study was relatively low compared to international data from previous studies,[9–13] which limits the potential of POC CRP in reducing overprescribing of antibiotics. Yet, the prescribing rate in our study is representative for that in respiratory tract infections in Dutch general practice, which decreased over recent years.[20] We performed a post hoc sensitivity analysis comparing practices with relatively high (38%; in our case mostly non-academic practices) and relatively low prescribing rates (26%; in this study mainly the academic practices), and we found that indeed in practices with higher prescribing rates implementation of POC CRP did diminish antibiotic use "post-test", the difference in the non-academic practices was -11% (95% CI -17 to -4). Secondly, GPs did complete CRFs for most acute cough patients for whom they requested a POC CRP test, yet for acute cough patients in whom they did not request POC CRP they often forgot the CRF. Therefore, the 78% POC CRP test rate in the study population of the present study is likely to be an overestimation. To estimate the true POC CRP test rate we extracted anonymous registration data of all patients presenting with acute cough in 4 of the participating practices during 6 months of the inclusion period. 1473 eligible patients were identified of whom only 342 (23%) were POC CRP tested, i.e. a much smaller proportion than in the study sample. We did not evaluate patient characteristics like preference for antibiotics which also may have influenced the GP in prescribing (or not).

Comparison with literature

Our finding that there was no overall reduction in antibiotic prescribing rate "post-CRP testing" differs from RCTs which reported a decrease in antibiotic use in CRP tested LRTI patient groups.[7,9–11] This difference could partly be explained by the low antibiotic prescribing rate in practices in our study. Both "pre-test" and "post-test" prescribing rates in our study were approximately 30%, which is comparable to the rate of the CRP-tested intervention groups from the previous trials which ranged from 31-38%.[7,9–11,13] The difference in results could also be due to differences in training. As our goal was to study daily practice we did not offer formal training to GPs when implementing POC CRP testing. This might have contributed to the lack of impact on antibiotic prescribing rates on group level.

Conclusion

POC CRP testing did not reduce overall antibiotic prescribing in this observational study, where POC CRP is used by GPs who already have a low antibiotic prescribing rate. Better targeting antibiotic prescribing by reducing diagnostic uncertainty and thus reducing antibiotic prescribing is an important rationale for implementing POC CRP in general practice. In addition, given the low adherence rate, apparently other factors than mentioned in the guideline prompt GPs to use POC CRP tests. Future studies should focus on other indications for POC CRP than lower respiratory tract infections and a better and effective use of tests like POC CRP in daily practice.

Acknowledgements

We would like to thank all participating general practitioners.

References

1. Hak E, Rovers MM, Kuyvenhoven MM, et al. 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.
2. Macfarlane J, Holmes W, Gard P, et al. Prospective study of the incidence, aetiology and outcome of adult lower respiratory tract illness in the community. *Thorax*. 2001;56(2):109–114.
3. Woodhead M, Blasi F, Ewig S, et al. Guidelines for the management of adult lower respiratory tract infections--full version. *Clin Microbiol Infect*. 2011;17 Suppl 6:E1–59.
4. Flanders S a., Stein J, Shochat G, et al. Performance of a bedside C-reactive protein test in the diagnosis of community-acquired pneumonia in adults with acute cough. *Am J Med*. 2004;116(8):529–535.
5. Hopstaken RM, Muris JW, Knottnerus JA, et al. Contributions of symptoms, signs, erythrocyte sedimentation rate, and C-reactive protein to a diagnosis of pneumonia in acute lower respiratory tract infection. *Br J Gen Pract*. 2003;53(490):358–364.
6. 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:f2450.
7. Andreeva E, Melbye H. Usefulness of C-reactive protein testing in acute cough/respiratory tract infection: an open cluster-randomized clinical trial with C-reactive protein testing in the intervention group. *BMC Fam Pract*. 2014;15(1):80.
8. Cals JW, Schot MJ, de Jong SA, et al. Point-of-care C-reactive protein testing and antibiotic prescribing for respiratory tract infections: a randomized controlled trial. *Ann Fam Med*. 2010;8(2):124–33.
9. Cals JWL, Butler CC, Hopstaken RM, et al. Effect of point of care testing for C reactive protein and training in communication skills on antibiotic use in lower respiratory tract infections: cluster randomised trial. *BMJ*. 2009;338:b1374.
10. Huang Y, Chen R, Wu T, et al. Association between point-of-care CRP testing and antibiotic prescribing in respiratory tract infections: A systematic review and meta-analysis of primary care studies. *Br J Gen Pract*. 2013;63(616):787–794.
11. Little P, Stuart B, Francis N, et al. Effects of internet-based training on antibiotic prescribing rates for acute respiratory-tract infections: A multinational, cluster, randomised, factorial, controlled trial. *Lancet*. 2013;382(9899):1175–1182.
12. Llor C, Cots JM, López-Valcárcel BG, et al. Interventions to reduce antibiotic prescription for lower respiratory tract infections: Happy Audit study. *Eur Respir J*. 2012;40(2):436–441.
13. Aabenhus R, Jensen-Jens US, Jørgensen KJ, et al. Biomarkers as point-of-care tests to guide prescription of antibiotics in patients with acute respiratory infections in primary care. *Cochrane Database Syst Rev*. 2014 Nov 6;11:CD010130.
14. Cals JW, Ament AJ, Hood K, et al. C-reactive protein point of care testing and physician communication skills training for lower respiratory tract infections in general practice: Economic evaluation of a cluster randomized trial. *J Eval Clin Pract*. 2011;17(6):1059–1069.

15. Dahler-Eriksen BS, Lauritzen T, Lassen JF, et al. Near-patient test for C-Reactive protein in general practice: Assessment of clinical, organizational, and economic outcomes. *Clin Chem*. 1999;45(4):478–485.

16. Oppong R, Jit M, Smith RD, et al. Cost-effectiveness of point-of-care C-reactive protein testing to inform antibiotic prescribing decisions. *Br J Gen Pract*. 2013;63(612):465–471.

17. Verlee L, Verheij TJM, Hopstaken RM, et al. [Summary of NHG practice guideline "Acute cough"]. *Ned Tijdschr Geneeskd*. 2012;156:A4188.

18. Katier N, Uiterwaal CSPM, de Jong BM, et al. The Wheezing Illnesses Study Leidsche Rijn (WHISTLER): rationale and design. *Eur J Epidemiol*. 2004;19(9):895–903.

19. Minnaard MC, van de Pol AC, Broekhuizen BDL, et al. Analytical performance, agreement and user-friendliness of five C-reactive protein point-of-care tests. *Scand J Clin Lab Invest*. 2013;73(8):627–34.

20. Van den Broek d'Obrenan J, Verheij TJM, Numans ME, et al. Antibiotic use in Dutch primary care: Relation between diagnosis, consultation and treatment. *J Antimicrob Chemother*. 2014;69(6):1701–1707.


7

Chapter 7

Selective recruitment in observational research: an empirical example of an implementation study on C-reactive protein point-of-care testing in primary care

Margaretha C. Minnaard
Janna van der Zand
Alma C. van de Pol
Niek J. de Wit
Alwin Schierenberg
Rogier M. Hopstaken
Sanne van Delft
Theo J.M. Verheij
Berna D.L. Broekhuizen

Submitted for publication

Abstract

Background

Failure to recruit all eligible study patients can lead to biased results. Little is known on selective patient recruitment in studies on implementation of diagnostic devices.

Objectives

The aim of this secondary analysis was to measure recruitment of patients in an implementation study in primary care on use of point-of-care (POC) C-reactive protein (CRP) and to evaluate recruitment bias and its impact on the study endpoint.

Methods

In a cross sectional observational study on POC CRP implementation and related antibiotics prescribing we compared included patients with all eligible patients to assess representativeness of the included subjects. Eligible patients were adults presenting with acute cough in primary care between March and September 2012. The frequency of POC CRP testing and the proportion of prescribed antibiotics were compared between recruited and non-recruited patients. As measure of bias odds ratios (ORs) with accompanying 95% confidence intervals (CIs) for the association between CRP level (< 20 mg/L or not) and antibiotic prescribing were computed.

Results

Of all 1473 eligible patients 348 (24%) were recruited. In recruited patients POC CRP tests were conducted and antibiotics prescribed more frequently as compared to non-recruited patients (81% vs. 6% and 44% vs. 29%, respectively). The ORs were 18.2 (95% CI 9.6-34.3), 30.5 (95% CI 13.2-70.3) and 3.8 (95% CI 0.9-14.8) in respectively all eligible patients, the recruited and the non-recruited patients.

Conclusion

Selective recruitment resulted in an overestimation of POC CRP test use and antibiotic prescribing.

Background

Recruitment bias is a systematic error that can affect the validity of a study. Failure to recruit all eligible study patients is common[1–4], but recruitment selection does however not necessarily lead to a biased estimate of the study outcome. To evaluate selection in recruitment and possible bias, recruited patients should be compared to the total group of eligible patients. If these groups differ too much in baseline characteristics and studied associations, recruitment bias is likely and this should be taken into account when interpreting the results. Many studies reported on statistical solutions to adjust for recruitment bias in non-randomised therapeutic interventions which are widely embedded in daily usual care, such as influenza vaccination.[5–8] Little is known however about recruitment in studies on implementation of new interventions. We recently completed a study on the impact of the implementation of C-reactive protein (CRP) point-of-care (POC) testing on antibiotic prescribing for acute cough patients in primary care in the Netherlands. CRP is an inflammation marker which is increasingly being measured in primary care at POC, i.e. during consultation to help to assess the severity of lower respiratory tract infection (LRTI) merely aiming to result in better antibiotic stewardship in patients presenting with acute cough.[9] We found that the POC CRP test results did not reduce overall antibiotic prescribing rate, but did influence general practitioners (GPs) to change their decision about antibiotic prescribing in patients with acute cough. Of the included study patients 78% was tested for POC CRP, which was higher than we expected because POC CRP testing is only indicated in a subgroup of patients. Because of suspected selection bias we conducted a secondary analysis on the recruitment of patients to estimate the impact of selective patient inclusion on the conclusions of the original study.

Methods

Design

Secondary analysis of a cross sectional non-randomised implementation study.

Setting and study population

Data of the original study were collected from February 2012 to February 2013 in nine health care centres in the Netherlands. Participating GPs were instructed to recruit all consecutive adult patients presenting with acute or newly presented cough (defined as cough less than 24 days) during day care consulting hours, and to report their management (i.e. whether they ordered POC CRP testing or not) and their intention to prescribe antibiotics both before and after POC CRP testing on a case report form. All patients received 'usual care', i.e. the GPs independently decided on the diagnostic work-up (including use of POC CRP test) and management.

For the secondary analysis we collected data retrospectively from March to September 2012. The secondary analysis was performed with data from four of the nine participating health care centres. The four centres are academic primary health care centres in Leidsche Rijn, a rapidly growing residential area near the city of Utrecht in the Netherlands. In these four health care centres 20 GPs serve approximately 35,000 patients in total.[10] These centres participate in the Utrecht Primary Care Network, and routine care data can easily be extracted from this research database.

Measurements

We retrospectively identified anonymised data of all potentially eligible patients that consulted the GPs between March to September 2012 from their electronic medical records (EMR), using International Classification of Primary Care (ICPC) codes[11]: R05 (cough), R78 (acute bronchitis/bronchiolitis), R81 (pneumonia), R74 (acute upper respiratory tract infection), R95 (COPD/emphysema) and R96 (asthma). In case of ICPC codes R74, R95 or R96, we analysed the full EMR consultation texts and only included subjects if the included consultation represented acute or newly presented cough (defined as cough less than 24 days). Patients that were actually included in the study by the GPs were labelled as "recruited patients". For all patients the following data was extracted from their EMR: age, gender, relevant comorbidity (defined by ICPC code R95 [COPD], R96 [asthma], K77 [heart failure] and T90 [diabetes mellitus]), results of the infection parameters CRP, erythrocyte sedimentation rate (ESR) and leukocyte count (LC), referral to the pulmonologist and the prescription of antibiotics by the GP within the 4 weeks after consultation. The latter was extracted from the EMRs using the Anatomical Therapeutic Chemical (ATC) classification code:[12] J01A (Tetracyclines), J01C (Beta-lactam Antibacterials and Penicillins) and J01F (Macrolides, Lincosamides and Streptogramins). CRP level was dichotomized into <20 mg/L (low) and ≥20mg/L (high), in concordance with the guidelines, that recommend to withhold antibiotics in patients with CRP<20 mg/L.[13,14] To evaluate possible mechanisms for selective recruitment all participating GPs of the four academic primary health care centres completed a questionnaire (Appendix 7.1) enquiring about their main reasons not to have included all eligible patients, after data-collection of the original study was completed.

Outcomes

The primary outcome was the size of selection, reflected by proportion of all eligible patients who were actually included in the implementation study. Secondary outcomes were the proportion of patients who underwent POC CRP measurement, the proportion of patients who were prescribed antibiotics and the association between POC CRP level and antibiotic prescribing.

Data analysis

In order to evaluate selective recruitment and subsequent bias we first compared the actual recruited patients with all potentially eligible patients to determine the repre-

sentativeness of the recruited subjects for the whole group of eligible patients. Because the recruited patients are a subgroup of all eligible patients these groups partly overlap. Therefore a direct comparison of these groups with formal statistical testing was not applicable; hence we tested for differences between recruited with non-recruited patients. Distributions of patient characteristics (age, gender, relevant comorbidity) and diagnostic and therapeutic management (blood tests, antibiotic prescription and referral to pulmonologists) were computed for all eligible patients, the recruited and the non-recruited patients, and compared between the recruited and the non-recruited patients. Means of continuous variables and proportions of categorical variables were compared using 95% confidence intervals (CIs). Furthermore the proportion and accompanying 95% CIs of being POC CRP tested and of receiving antibiotics were calculated and reported (without formal statistical testing) for all eligible patients and the recruited patients, and compared between recruited and non-recruited with a p-value of 0.05 used as significance level. As a measure of bias, the association between CRP level (dichotomized at 20 mg/litre) and antibiotic prescribing was expressed as univariate odds ratio (OR) with accompanying 95% CIs. First we computed this association for the three groups separately. Second we compared the ORs between recruited and non-recruited patients using the log likelihood ratio test (χ^2 test) for the differences between models with and without an interaction term for group (recruited versus non-recruited). We used this interaction term to estimate whether the association between CRP results and antibiotic prescribing (yes or no) differed between recruited and non-recruited patients. Responses on the questionnaire were summarized. Data were analysed using SPSS for Windows 20.0.0 (SPSS, Inc. Chicago, Illinois, USA).

Ethical approval

The act on medical research involving human subjects does not apply to this study and therefore an official approval of this study by the Medical Ethics Research Committee of the University Medical Center Utrecht was not required.

Results

During the study period 1473 patients were eligible. Three hundred forty-eight (24%) had actually been recruited by the 20 participating GPs. Table 7.1 shows patient characteristics of all eligible patients, and of recruited and non-recruited patients. These patient groups did not differ in age, gender, the number of relevant comorbidities (heart failure, COPD, asthma and diabetes mellitus) and referrals to the pulmonologist. In 342 (23%, 95%CI 21-25) of all eligible patients POC CRP test results were available. 280 (81%, 95% CI 76-84) of the recruited patients had POC CRP test results versus 62 (6%, 95% CI 4-7) of the non-recruited ($p<0.00$). Of all eligible patients 477 (32%, 95% CI 30-35) got an antibiotic prescription.

Table 7.1: patient characteristics.
characteristics of all eligible patients, divided into recruited and non-recruited patients

		All eligible patients n=1473	Recruited patients n=348	Non-recruited patients n=1125	p-value[b]
Mean age, years	(SD, 95%CI)	48 (16, 47-49)	48 (15, 46-49)	48 (16, 47-49)	0.76
Male gender,	n (%, 95%CI)	607 (41, 39-44)	131 (38, 33-43)	476 (42, 40-45)	0.12
Comorbidity					
Heart failure,	n (%, 95%CI)	32 (2, 2-3)	7 (2, 1-4)	25 (2, 2-3)	0.81
COPD,	n (%, 95%CI)	103 (7, 6-8)	28 (8, 6-11)	75 (7, 5-8)	0.38
Asthma,	n (%, 95%CI)	204 (14,12-16)	54 (16, 12-20)	150 (13, 12-16)	0.30
Diabetes Mellitus,	n (%, 95%CI)	132 (9, 8-11)	29 (8, 6-12)	103 (9,8-11)	0.64
Blood tests performed					
CRP,	n (%, 95%CI)	342 (23, 21-25)	280 (81, 76-84)	62 (6, 4-7)	0.00
ESR,	n (%, 95%CI)	63 (4, 3-6)	7 (2, 1-4)	56 (5, 4-6)	0.02
LC,	n (%, 95%CI)	70 (5, 4-6)	8 (2, 1-5)	62 (6, 4-7)	0.01
Antibiotic prescription,	n[a] (%, 95%CI)	477 (32, 30-35)	152 (44, 39-49)	325 (29, 26-32)	0.00
Referral to pulmonologists,	n[a] (%, 95%CI)	44 (3, 2-4)	15 (4, 3-7)	29 (3, 2-4)	0.10

Abbreviations: SD: standard deviation, COPD: chronic obstructive pulmonary disease, CRP:C-reactive protein, ESR: erythrocyte sedimentation rate, LC: leucocyte count., [a] during 28 days of follow up [b] accompanying the comparison of recruited versus non-recruited patients.

Table 7.2: Antibiotic prescribing.
Antibiotic prescribing in point-of-care C-reactive protein (CRP) tested patients related to CRP level, in all eligible, recruited and non-recruited patients

	all eligible patients			Recruited patients			Non-recruited patients		
	antibiotics prescribed	no antibiotics prescribed	total	antibiotics prescribed	no antibiotics prescribed	total	antibiotics prescribed	no antibiotics prescribed	total
CRP≥20 mg/L n (%)	82 *(85%)*	14 *(15%)*	96 *(100%)*	77 *(92%)*	7 *(8%)*	84 *(100%)*	5 *(42%)*	7 *(58%)*	12 *(100%)*
CRP<20 mg/L n (%)	60 *(24%)*	186 *(76%)*	246 *(100%)*	52 *(26%)*	144 *(74%)*	196 *(100%)*	8 *(16%)*	42 *(84%)*	50 *(100%)*
Total n (%)	142 *(41%)*	200 *(59%)*	342 *(100%)*	129 *(46%)*	151 *(54%)*	280 *(100%)*	13 *(21%)*	49 *(79%)*	62 *(100%)*

Table 7.3: Questionnaires
Results from questionnaires completed by participating general practitioners to evaluate possible mechanisms for selection in recruitment.

Information on participating General practitioners, n=15	
Age, mean (SD)[a]	38 (8)
male, n (%)	1 (8)
PhD, n (%)	3 (32)
Years of experience as general practitioner, mean (range)	7 (0-22)
POC CRP used for other diseases (non LRTI),n (%)[b]	6 (46)
Percentage included patients according to GPs[c], %	89
Information on study	
Presentation, n (%)	3 (20)
Newsletter, n (%)	7 (47)
Poster, n (%)	1 (7)
Colleague, n (%)	10 (67)
No information, n (%)	1 (7)
Provided information was sufficient, n (%)	15 (100)
Opinions of participating GPs 0=disagree, 7=agree (median score with range 0-7)	
Did not encounter eligible patients	1
Forget to recruit due to busy schedule	5
No cooperation of patient	1
More information needed during recruitment timeframe	2
POC CRP is easy to perform[a]	6
Enough reminders provided during recruitment timeframe	6

[a] missing data: n=1, [b] missing data: n=2, [c] missing data: n=6

A total of 152 (44%, 95% CI 39-49) of the recruited patients got an antibiotic prescription versus 325 (29%, 95%CI 26-32) of the non-recruited (p<0.00), (table 7.1). The OR for the association between CRP level and antibiotic prescribing was 18.2 (95% CI 9.6-34.3) and 30.5 (95% CI 13.2-70.3) and 3.8 (95% CI 0.9-14.8) in respectively all eligible patients, the recruited and the non-recruited patients (table 7.2). The OR of the recruited and of the non-recruited patients differed significantly (p-value <0.01). In total 15 (75%) GPs completed the questionnaire. They reported as main reason not to have included all eligible patients that, due limited time per consultation, they frequently forgot to recruit eligible patients. Questionnaire responses are summarized in table 7.3.

Discussion

Main results

In our study on implementation of POC CRP tests in adults presenting with acute cough in primary care, participating GPs recruited only 24% of all eligible patients. POC CRP tests were conducted more frequently and antibiotics were prescribed more often in recruited patients than in non-recruited patients (81% versus 6% and 44% versus 29%). The ORs between CRP level and antibiotic prescribing were different for recruited and non-recruited patients.

Interpretation of findings

Time constraints and forgetting to recruit patients as indicated by the participating GPs, are the most indicated reasons for the low recruitment rate. Another reason could be that patients refused undergoing POC CRP testing, therefore leading to selective recruitment. Confounding by (test) indication could also be an explanation for a higher proportion of POC CRP testing and antibiotic prescribing in recruited subjects, assuming that GPs recruited predominantly patients for whom they were planning on POC CRP testing and/or prescribing antibiotics. Such a more active treatment attitude in recruited patients has been described previously by Grobbee *et al*.[15] Misinterpretation of the study protocol, i.e. only including patients when CRP testing was considered, could also be a supplementary explanation. One could also argue that patients in the recruited group had more severe symptoms, which could explain that they were prescribed more antibiotics compared to all eligible patients.

Strengths and limitations

This is to the best of our knowledge the first study which evaluated recruitment in an observational study on implementation of a diagnostic device in routine primary care. Some aspects need to be taken into consideration when interpreting the results. Firstly, because we conducted a retrospective study, retrieval of all information on signs and symptoms and their impact on recruitment was not possible. Nevertheless we think that missing data because of insufficient registration was limited because the participating GPs have been well trained to record data in a standardised manner. Secondly we chose to compare the actually recruited patients with all eligible patients, although this does not meet requirements for formal statistical testing because these groups partly overlap. We did this because we aimed to assess the representativeness of the recruited group for all eligible patients. Thirdly, it was not possible to detect the indication for all antibiotics prescribed. Antibiotics may have been prescribed for other diseases than lower respiratory tract infections. However, since we used ATC codes corresponding to the index consultation (therefore linked to the ICPC codes) the risk of misinterpretation of antibiotic prescribing was small.

Comparison with literature

Suboptimal recruitment caused by physician related factors such as time constraints and forgetting to recruit as suggested by the questionnaires in the present study has been reported before.[16] Under-recruitment is also problematic in randomised controlled trials (RCTs); in the UK less than one-third of the trials recruited to target[17,18] and some trials were even forced to stop due to under-recruitment.[19] RCTs, in which patient informed consent is required, reported inclusion rates ranging from 38% to 61%.[1,4,20] Despite randomization, patient characteristics often differ between the study arms of a RCT and there is debate on whether statistical testing for baseline imbalances is appropriate to detect possible recruitment bias in trials and whether statistical techniques should be used to adjust for these imbalances.[21,22] Our study was not designed as a RCT and patients' informed consent was not requested, which eliminates potential for consent bias. Previous observational studies on non-randomized interventions, for example influenza vaccination[7], have shown relevant bias regarding the preferences in patients receiving the interventions, where patients with worse prognosis were more likely to receive the intervention (confounding by indication).[5,7,8]

Conclusion

In a POC CRP implementation study GPs recruited only a minority of eligible patients which resulted in an overestimation of POC CRP testing and antibiotic prescribing in recruited patients and also a biased estimate of the association between CRP level and antibiotic prescribing.

References

1. Bijkerk CJ, Muris JWM, Knottnerus JA, et al. Randomized patients in IBS research had different disease characteristics compared to eligible and recruited patients. *J Clin Epidemiol*. 2008;61(11):1176–1181.
2. Herber OR, Schnepp W, Rieger MA. Recruitment rates and reasons for community physicians' non-participation in an interdisciplinary intervention study on leg ulceration. *BMC Med Res Methodol*. 2009;9:61.
3. Page MJ, French SD, McKenzie JE, et al. Recruitment difficulties in a primary care cluster randomised trial: investigating factors contributing to general practitioners' recruitment of patients. *BMC Med Res Methodol*. 2011;11(1):35.
4. Parkes J, Donnelly C, Mcdowell B, et al. Recruitment bias and characteristics of participants with severe cerebral palsy in a cross-sectional survey. *J Adv Nurs*. 2012;68(2):368–378.
5. Fireman B, Lee J, Lewis N, et al. Influenza vaccination and mortality: Differentiating vaccine effects from bias. *Am J Epidemiol*. 2009;170(5):650–656.
6. Groenwold RHH, Van Deursen AMM, Hoes AW, et al. Poor quality of reporting confounding bias in observational intervention studies: a systematic review. *Ann Epidemiol*. 2008;18(10):746–751.
7. Hak E, Verheij TJM, Grobbee DE, et al. Confounding by indication in non-experimental evaluation of vaccine effectiveness: the example of prevention of influenza complications. *J Epidemiol Community Health*. 2002;56(12):951–955.
8. Ye Y, Kaskutas LA. Using propensity scores to adjust for selection bias when assessing the effectiveness of Alcoholics Anonymous in observational studies. *Drug Alcohol Depend*. 2009;104(1-2):56–64.
9. 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:f2450.
10. Katier N, Uiterwaal CSPM, de Jong BM, et al. The Wheezing Illnesses Study Leidsche Rijn (WHISTLER): rationale and design. *Eur J Epidemiol*. 2004;19(9):895–903.
11. Soler JK, Okkes I, Wood M, et al. The coming of age of ICPC: Celebrating the 21st birthday of the International Classification of Primary Care. *Fam Pract*. 2008;25(4):312–317.
12. WHO Collaborating Centre for Drug Statistics Methodology. Guidelines for ATC Classification and DDD assignment 2013. Norwegian Institute of Public Health, Oslo, Norway. www.whocc.no/filearchive/publications/1_2013guidelines.pdf
13. Woodhead M, Blasi F, Ewig S, et al. Guidelines for the management of adult lower respiratory tract infections--full version. *Clin Microbiol Infect*. 2011;17 Suppl 6:E1–59.
14. Verlee L, Verheij TJM, Hopstaken RM, et al. [Summary of NHG practice guideline "Acute cough"]. *Ned Tijdschr Geneeskd*. 2012;156:A4188.
15. Grobbee DE, Hoes AW. Confounding and indication for treatment in evaluation of drug treatment for hypertension. *BMJ*. 1997;315(7116):1151–1154.

16. Parkes J, Kerr C, McDowell BC, et al. Recruitment bias in a population-based study of children with cerebral palsy. *Pediatrics*. 2006;118(4):1616–1622.

17. Graffy J, Grant J, Boase S, et al. UK research staff perspectives on improving recruitment and retention to primary care research; nominal group exercise. *Fam Pract*. 2009;26(1):48–55.

18. McDonald AM, Knight RC, Campbell MK, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. *Trials*. 2006;7:9.

19. Peters-Lawrence MH, Bell MC, Hsu LL, et al. Clinical trial implementation and recruitment: Lessons learned from the early closure of a randomized clinical trial. *Contemp Clin Trials*. 2012;33(2):291–297.

20. Taylor KSM, Gordon JC, Harris CE, et al. Recruitment bias resulted in poorer overall health status in a community-based control group. *J Clin Epidemiol*. 2008;61(9):890–895.

21. Eldridge S, Kerry S, Torgerson DJ. Bias in identifying and recruiting participants in cluster randomised trials: what can be done? *BMJ*. 2009;339(oct09_1):b4006.

22. Giraudeau B, Ravaud P. Preventing bias in cluster randomised trials. *PLoS Med*. 2009;6(5).

Appendix 1:
Questionnaire for participating General Practitioners.

Please fill out all questions below, mark your response.

Age: ____ years
sex: male/female
obtained PhD-degree: yes/no
how many years of experience as a general practitioner do you have: ____ years

Did you use POC CRP for other diseases than LRTI: yes/no
what is the percentage eligible patients you have included for the study?

Did you receive and understand information on the study given by:
presentation: yes/no
newsletter: yes/no
poster: yes/no
a colleague: yes/no
no information: yes/no

Please indicate if you agree or disagree on the next statements:

I did not encounter eligible patients
totally disagree 1 2 3 4 5 6 7 totally agree

I forgot to recruit due to busy schedule
totally disagree 1 2 3 4 5 6 7 totally agree

Patients did not cooperate
totally disagree 1 2 3 4 5 6 7 totally agree

I needed more information on the study during the inclusion timeframe
totally disagree 1 2 3 4 5 6 7 totally agree

The point of care test for measuring C-reactive protein is well performable
totally disagree 1 2 3 4 5 6 7 totally agree

I was provided with enough reminders during the recruitment timeframe
totally disagree 1 2 3 4 5 6 7 totally agree

Chapter 8

General Discussion

8

This thesis aimed to evaluate optimal use of the point-of-care (POC) C-reactive protein (CRP) in primary care and its impact in day-to-day practice. More specifically our aims were to determine: [1] the (added) diagnostic value of CRP in the diagnostic work-up of community acquired pneumonia (CAP) in primary care, [2] the suitability of POC CRP test devices for primary care and [3] the impact of the POC CRP test result on the GP's decision to prescribe antibiotics.

Main conclusions of this thesis

The first part of this thesis showed that a prediction model including coryza, dyspnoea, crackles, diminished breath sounds, fever, and tachycardia, had the best diagnostic performance in diagnosing CAP. Moreover, adding CRP to signs and symptoms in the diagnostic workup of CAP improved both discrimination and risk classification. Nevertheless, in the majority of patients uncertainty about the diagnosis remained after adding CRP, because their predicted risk of CAP after application of the risk model remained in the 'intermediate' range. The second part of this thesis showed that five different POC CRP test devices show considerable variation in analytical performance and agreement with a laboratory reference test. These differences do not translate into differences in accuracy of predicting (radiographic) CAP, both as single test as well as in combination with available clinical findings. In the third part of this thesis we concluded that implementation of POC CRP testing did not reduce overall antibiotic prescribing for patients presenting with acute cough. Moreover, only a minority of patients who underwent POC CRP testing had an indication for testing according to guideline recommendations. The POC CRP test result did influence the GPs' decision making to prescribe antibiotics in nearly one third of all cases. Finally, only a quarter of all eligible patients were recruited in our implementation study, which resulted in overestimation of POC CRP test use, antibiotic prescribing and in an overestimation of the effects of POC CRP test use on antibiotics prescribing. In this chapter we will discuss these conclusions, put them in a broader perspective and elaborate on possible implementation in daily practice.

Are diagnostic models for CAP helpful in daily practice?

First, the answer to this question above depends upon whether a diagnosis is in itself helpful in daily practice. Because the link between diagnosis and prognosis and that between diagnosis and treatment effect is often not straightforward, information on prognosis and/or treatment effect may be much more helpful in daily practice than predicting diagnosis.[1] In the case of CAP for instance, it was shown recently that on average patients with a radiologically proven CAP do clearly benefit from antibiotic treatment.[2] At the same time we know that a substantial proportion of CAP has a viral

aetiology.[3,4] Thus, CAP is a diagnosis that has clear therapeutic implications for the average patient, but for a more targeted approach predicting bacterial CAP would be a better option in the future. Furthermore, the utility (and second to that, implementation and use) of prediction models is of course hampered by their imperfection. We found a considerable proportion of patients in whom the estimated risk of CAP remained in the 'intermediate risk' range after adding CRP. This means that for this large group of patients the GP still remains uncertain about the presence of CAP, and the true incremental value of CRP may therefore be questioned. We should realise, however, that tests will never reduce diagnostic uncertainty to zero. Last but not least, the usefulness of prediction models in daily clinical practice depends on whether they are easy to apply.[5] Most prediction models are rather complex, requiring detailed calculations during consultation. Simplifying a model leads to loss of information and therefore a less accurate prediction. Unless models become available that combine simple application with sufficient diagnostic accuracy, their implementation success in daily practice will probably remain disappointing. ICT supported prediction models, linked to routine registration systems in primary care, may help to implement more complex prediction models in the near future.

Methodological issues in the external validation of diagnostic models

We used individual patient data from all available diagnostic studies for CAP published so far to validate previously published diagnostic models for CAP, to evaluate model performance in alternative settings and patientgroups than those used for the development of the models. In this process we encountered some problems worth mentioning. First of all, due to heterogeneity in the data, and the fact that not all predictors (i.e. signs and symptoms) were present in all datasets, it was not possible to pool all data and to validate all models in the pooled dataset. This underlines the urgency to use comparable clinical registration forms, measurements and data collection methods in comparable fields of research. A second problem we encountered was the fact that the study populations included in the IPD meta-analysis differed considerably in size; in our analysis one study population was even larger than that of all other studies together. One may argue that the largest developmental dataset also has the largest impact on the results of the external validation analysis. To prevent this we evaluated model performance on a relative scale and used weighing factors correcting for the size of the study populations in the different datasets. Moreover, the model from the largest development study does not automatically perform well in other settings with different case mix.

Introduction of POC CRP test devices in daily practice

Calibration of POC CRP test devices with a laboratory reference test showed clear differences between these test devices. However, the POC CRP tests did not differ in their capacity to diagnose a clinically relevant patient outcome, in this case CAP. Considering the fact that all four POC CRP test devices showed comparable diagnostic properties, with acceptable agreement, user-friendliness and costs are leading while choosing a device for use in daily care. User-friendliness will support correct use in primary care centers. Costs are also important, especially the price per test-unit, while differences in the price of the POC test hardware will hardly contribute to the annual costs per practice. The development of new POC CRP test devices is continuously ongoing and new tests are allowed to the European market on condition that European Requirements (98/79/EC) for a communautés Européenes (CE) are met.[6] It is striking that neither adequate diagnostic accuracy nor optimal test efficacy is included in these requirements before tests are allowed market access. This is in sharp contrast with new drugs for which extensive registration requirements must be met, including evidence on clinical effectiveness and safety, before market access. As more and more GPs consider using POC CRP test devices, risks associated with its use, such as risk of infection or inadequate functioning of equipment, should not be ignored. This calls for guidelines regarding the use of these devices.[7] In the Netherlands recently a multidisciplinary guideline on implementation of POC tests in daily practice was developed.[8] We agree with experts who claim that POC CRP testing is to be accredited according to ISO norms.[9,10] Implementation of these quality procedures might well be too complicated and time consuming for individual practices, and collaboration with an experienced diagnostic partner like a regional primary care- or hospital laboratory could be a solution.

Implementation of POC CRP in primary care; premature?

Chapter 6 showed that GPs did not use and interpret POC CRP tests according to the guidelines in the majority of patients. Previous reports have shown that in general, GPs comply with professional guidelines in approximately 60-70% of their management decisions.[11] In the remaining part of decisions GPs will sometimes legitimately deviate from guideline recommendations, and in other cases not. If the guideline deviation is structural, as found in our study, this might be explained by the fact that either [i] the advice in the guideline is too complicated to implement in daily practice, or [ii] that the recommendation itself is still under debate and that therefore GPs may be ambivalent to comply. In the first case [i] GPs should be better facilitated by training and support. Selection criteria should be simplified and integrated ICT decision support systems could improve the use of additional POC tests like CRP. The second case [ii], however, is more complex, as it indicates that the decision to incorporate the CRP test in the LRTI guideline may have been made prematurely. The previous trials on POC CRP evaluated

POC CRP testing in all patients presenting with symptoms of LRTI, and showed reduction of antibiotic prescribing. In the Dutch guideline on "acute cough" POC CRP was however introduced as "selective testing", accompanied by a newly developed flowchart on the indication for testing based on clinical assessment. Our results show disappointing results on the compliance with the recommendations in the guideline, and point at a clear POC test 'overuse'. This may be explained by the fact that the flowchart which advises on selective use of POC CRP is too complex for use in daily practice. Taking into account that the IPD meta-analysis showed that diagnostic doubt remains in the majority of patients after additional testing of CRP, one may wonder if the current emphasis on CRP in the guideline was somewhat premature. Nevertheless POC CRP can certainly be helpful for GPs in their management of patients presenting with LRTI. Some GPs use this test to support their explanation to the patient why they refrain from prescribing antibiotics, although this could be considered as "improper test use". The effect of POC CRP on antibiotic prescribing in our study was smaller than in previous randomized controlled trials (RCTs) in this field.[12–16] This could be partly explained by the pre-existing low antibiotic prescribing rate in practices participating in our study, as both "pre-test" and "post-test" prescribing rates were approximately 30%. A post hoc analysis showed a reduction in antibiotic prescribing for GPs who had a higher "pre-test" prescribing rate. This suggests that also without the use of CRP low antibiotic prescription rates in LRTI can be achieved. This may be related to a stronger emphasis on patient education and communication, as several studies showed positive effects of training in communication about antibiotic prescribing.[12,13,17]

8

The need for implementation studies

Some may argue that evaluating the implementation of health care innovations which have been proven effective in experimental studies is in fact a waste of time and resources. Our study however proves otherwise. Many health care interventions are effective in clinical trials, but often these circumstances are artificial. Participants are optimally trained, patients carefully selected, the invention optimally supported, and the process closely monitored. This is not comparable to the 'real world' of clinical practice where the intervention will be used in. Therefore we consider it mandatory to closely monitor the actual use of new health care interventions in daily practice in a systematic manner. Here we chose an observational study design to evaluate the implementation of POC CRP. By providing the GPs with case-record-forms we created a setting which still could be interpreted as somewhat unnatural. We argue nonetheless this was as close as we could get to monitor what really happens when the POC CRP test is used in daily practice and to study "real-life" compliance to guidelines. Alternatively, retrospective data analysis from primary care databases would not have given us a complete insight in the decision making process of the GP for a given patient. We could now use each patient included in our study as its own comparison, and zoom in on the "pre-test" and "post-

test" decision of the GP, which would not have been possible using retrospective data. Our study demonstrates that also implementation research has a substantial risk of biased inclusion, potentially threatening the validity of the conclusions. We agree with experts that possible selection in recruitment should be reported to allow for assessment of generalization and applicability of results.[18] Risk of selection could be minimised by sending the GPs reminders through their electronic patient records when encountering an eligible patient. Previous observational studies on (non-randomized) interventions like the influenza vaccination, have shown that relevant bias can affect effect estimation of vaccination on prognosis, because patients with worse prognosis were more likely to receive the intervention (confounding by indication).[19–21] In our study we also observed selection of patients which led to overestimation of POC CRP testing and antibiotic use, with a subsequent overestimation of the influence of POC CRP on antibiotic prescribing.

Future improvement for LRTI management in primary care

We think there is still room for improvement of LRTI management. Although the Netherlands is already frontrunner in antibiotic prescribing control, improvement is still possible in both reducing as in better targeted antibiotic prescribing, for instance by training and ICT support. The use of CRP as an incremental test could be further subject to research, in particular addressing the question if repeated measurements of CRP could be instrumental to reduce diagnostic uncertainty and to better predict prognosis, both in LRTI and in other common infectious diseases in primary care.

References

1. Naaktgeboren C. Improving the methodological framework for diagnostic studies. University Utrecht. 2015. ISBN 978-90-393-6333-1.
2. Teepe J, Little P, Elshof N, et al. Amoxicillin for clinically unsuspected pneumonia in primary care: subgroup analysis. Eur Respir J. 2015; accepted for publication.
3. Holm A, Nexoe J, Bistrup LA, et al. Aetiology and prediction of pneumonia in lower respiratory tract infection in primary care. *Br J Gen Pract*. 2007;57:547–554.
4. Graffelman a W, Knuistingh Neven A, le Cessie S, et al. Pathogens involved in lower respiratory tract infections in general practice. *Br J Gen Pract*. 2004;54(498):15–19.
5. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201–9.
6. Europese Commissie. Richtlijn 98/79/EG betreffende medische hulpmiddelen voor in vitro diagnostiek. Publicatieblad van de Europese Gemeenschappen,7.12.98, L331/1-37.
7. Janssens PMW. Widespread use of point-of-care testing is irreconcilable with the present-day quest for safety. *Clin Chem Lab Med*. 2010;48(10):1527–8.
8. Hopstaken RM , Kleinveld HA, Balen van JAM, et al. Richtlijn point of care testing (POCT) in de huisartsenzorg. www.nhg.org/actueel/nieuws/nieuwe-richtlijn-point-care-testing-de-huisartsenzorg.
9. ISO 22870:2006 Point-of-care testing (POCT) - Requirements for quality and competence. International Organization of Standards. 2006.
10. ISO 15189:2012 Medical laboratories - Particular requirements for quality and competence. International Organization of Standards. 2012.
11. Grol R, Dalhuijsen J, Thomas S, et al. Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *BMJ*. 1998;317(7162):858–61.
12. Cals JWL, Butler CC, Hopstaken RM, et al. Effect of point of care testing for C reactive protein and training in communication skills on antibiotic use in lower respiratory tract infections: cluster randomised trial. *BMJ*. 2009;338:b1374.
13. Little P, Stuart B, Francis N, et al. Effects of internet-based training on antibiotic prescribing rates for acute respiratory-tract infections: A multinational, cluster, randomised, factorial, controlled trial. *Lancet*. 2013;382(9899):1175–1182.
14. Andreeva E, Melbye H. Usefulness of C-reactive protein testing in acute cough/respiratory tract infection: an open cluster-randomized clinical trial with C-reactive protein testing in the intervention group. *BMC Fam Pract*. 2014;15(1):80.
15. Aabenhus R, Jensen-Jens US, Jørgensen KJ, et al. Biomarkers as point-of-care tests to guide prescription of antibiotics in patients with acute respiratory infections in primary care. *Cochrane Database Syst Rev*. 2014 Nov 6;11:CD010130.
16. Huang Y, Chen R, Wu T, et al. Association between point-of-care CRP testing and antibiotic prescribing in respiratory tract infections: A systematic review and meta-analysis of primary care studies. *Br J Gen Pract*. 2013;63(616):787–794.

17. Cals JW, Schot MJ, de Jong SA, et al. Point-of-care C-reactive protein testing and antibiotic prescribing for respiratory tract infections: a randomized controlled trial. *Ann Fam Med*. 2010;8(2):124–33.

18. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *Int J Surg*. 2014;12(12):1500–24.

19. Hak E, Verheij TJM, Grobbee DE, et al. Confounding by indication in non-experimental evaluation of vaccine effectiveness: the example of prevention of influenza complications. *J Epidemiol Community Health*. 2002;56(12):951–955.

20. Ye Y, Kaskutas LA. Using propensity scores to adjust for selection bias when assessing the effectiveness of Alcoholics Anonymous in observational studies. *Drug Alcohol Depend*. 2009;104(1-2):56–64.

21. Fireman B, Lee J, Lewis N, et al. Influenza vaccination and mortality: Differentiating vaccine effects from bias. *Am J Epidemiol*. 2009;170(5):650–656.

Summary


S

Summary

Cough is a very common symptom in the community, and lower respiratory tract infections (LRTIs) occur very frequently in all seasons (**chapter 1**). The primary goal in the diagnostic work-up of LRTI is to differentiate serious conditions like community acquired pneumonia (CAP) from self-limiting conditions and to prevent unnecessary prescribing of antibiotics. For adequate management of LRTI general practitioners (GPs) primarily rely on history and physical examination ('signs and symptoms'). Prediction models for CAP based on signs and symptoms have been proposed to support this clinical assessment. To further improve diagnostic accuracy additional diagnostic testing is potentially useful. A promising additional test is C-reactive protein (CRP). A point-of-care (POC) test in capillary blood yields CRP results within a few minutes, and can thus be applied in decision making (whether or not to prescribe antibiotics) during consultation. Several studies have shown that POC CRP testing is cost-effective and can reduce antibiotic prescribing. Currently most national and international guidelines provide guidance on indications for POC CRP testing and interpretation of test results in patients who present with acute cough. POC CRP testing was introduced in Dutch primary care after incorporation in the guideline LRTI of the Dutch College of General Practice in 2011. So far, little is known about the compliance of GPs with these guidelines and the impact of POC CRP introduction on prescription of antibiotics for LRTI. Moreover the guideline does not provide guidance on which POC CRP test device to use in primary care. Studies directly comparing all available devices are scarce and the extent to which variability of POC CRP test results translates into differences in diagnostic accuracy for CAP is unknown. With this thesis we aimed to determine the (added) diagnostic value of CRP in the diagnostic work-up of CAP in primary care, the suitability of POC CRP test devices for primary care and the impact of the POC CRP test result on the GP's decision to prescribe antibiotics.

In **chapter 2** all published signs and symptoms models for prediction of CAP in primary care were identified and externally validated in the individual patient data (IPD) of previously performed diagnostic studies. Six prediction models and IPD of eight diagnostic studies (N total=5308, prevalence of CAP 12%) were included. Performance of the six prediction models was compared on discrimination and calibration. Discrimination was measured using the pooled area under the curve (AUC) and delta AUC, representing the difference in discriminative performance between an individual model and the average performance of all models within an IPD dataset. Calibration was measured using calibration plots. Prediction models by van Vugt *et al*. and Heckerling *et al*. demonstrated the highest pooled AUC of 0.79 (95% confidence interval (CI) 0.74-0.85) and 0.72 (95% CI 0.68-0.76), respectively. Other models by Diehr *et al*., Singal *et al*., Melbye *et al*., and Hopstaken *et al*. demonstrated pooled AUCs of 0.65 (95% CI 0.61-0.68), 0.64 (95% CI 0.61-0.67), 0.56 (95% CI 0.49-0.63) and 0.53 (95% CI 0.5-0.56), respectively. The delta AUCs of the models showed a similar ranking. The

models by van Vugt *et al*. and Singal *et al*. showed close agreement between observed and predicted probabilities, other models lacked such correspondence. The model by van Vugt *et al*. demonstrated the highest discriminative accuracy coupled with reasonable-to-good calibration across the IPD of different study populations. This model was therefore regarded as the main candidate for primary care use.

In **chapter 3** we quantified the added value of CRP beyond signs and symptoms in IPD of the eight diagnostic studies. CRP was added to a prediction model that reflects daily clinical practice. Improvement in discrimination and risk classification were first calculated within all individual studies and subsequently meta-analysed across studies. In all datasets discrimination improved after addition of CRP to the basic model, with a mean improvement in area under the ROC curve (AUC) of 0.075, ranging from 0.02-0.18. Overall the proportion of patients without CAP correctly classified as low risk of CAP increased from 28% to 36% when CRP was added to the model. The proportion of patients with CAP assigned to the low risk category remained equal (n=4), the proportion of patients assigned to the intermediate risk category decreased from 56% to 51% and the proportion correctly classified as high risk increased from 63% to 70%. Resuming, adding CRP to the diagnostic workup of CAP in patients with LRTI in primary care enhanced discrimination and improved diagnostic risk classification, though a substantial group with diagnostic uncertainty remained.

In **chapter 4** we evaluated the following five POC CRP test devices: Afinion and NycoCard Reader II (both Alere), Eurolyser Smart 700|340 (Eurolyser), QuikRead go and QuikRead 101 (both Orion Diagnostica). Results were compared with those of a standard immunoturbidimetric method performed on a routine analyser (Olympus AU 2700, Beckman Coulter). Analytical performance and agreement with the laboratory standard for the five different POC tests were analysed and varied considerably. Within-day coefficients of variation (CVs) varied from 2.6% (QuikRead go) – 19.4% (Eurolyser) for low CRP values (<20 mg/L), and 1.1% (QuikRead go) – 17.5% (Eurolyser) for high CRP values (>100 mg/L). Between-day CVs varied from 4.6% (Afinion) – 30.5% (Eurolyser) for low CRP values and 4.0% (QuikRead go) – 18.0% (Eurolyser) for high CRP values. For all POC CRP test devices the overall mean difference with the laboratory standard was < 4.2 mg/L. However, with high CRP values (>100 mg/L) agreement with the laboratory standard systematically decreased for all POC CRP tests. Subsequently, user-friendliness of the POC tests was assessed. Information on device characteristics was collected from manufacturers' information sheets as set on January 2012. Afinion and Eurolyser were judged easiest to operate by 20 GPs and GP assistants who were unfamiliar with POC testing.

We aimed to determine whether the differences that were found in chapter 4 would meaningfully affect the accuracy and added diagnostic value in predicting radiographic pneumonia in adults presenting with acute cough in primary care in **chapter 5**. We conducted a nested case control study of adult patients presenting with acute cough in 12 different European countries (www.grace-lrti.org). Venous blood samples from 100 random patients with and 100 random patients without CAP were tested with

the five POC CRP tests and a laboratory analyser. Single test accuracy values showed similar results for all five POC CRP tests and the laboratory analyser. To assess the added value of CRP to symptoms and signs (clinical model), the area under the receiver operating characteristic curve (AUC) was calculated. All POC CRP test analysers showed comparable added value of CRP above symptoms and signs; the AUC of the different POC CRP test analysers and the laboratory analyser ranged from 0.79-0.80, and were higher than the clinical model without CRP (0.70). Multivariable odds ratios (ORs) were the same (1.2) for all POC CRP test devices. Therefore we concluded that variability in results obtained from standard CRP laboratory and POC test devices did not translated into clinically relevant differences when used for prediction of CAP in patients with acute cough in primary care.

To determine the effect of POC CRP testing on antibiotic prescribing and the extent of guideline adherence when implemented in routine primary care practice, we conducted an observational study (**chapter 6**). From February 2012 until February 2013 GPs were instructed to use POC CRP in adults with acute cough following current guidelines. After routine history taking and physical examination they reported whether they intended to prescribe antibiotics ("pre-test decision"). If they decided to perform POC CRP, they reported their (revised or not) decision after receiving the POC CRP test result ("post-test decision"). Primary outcome was the difference between "pre-test" and "post-test" antibiotic prescribing rate. Secondary outcome was the percentage of patients for whom the decision to prescribe antibiotics altered after POC CRP testing. A total of 40 GPs enrolled 939 patients, 78% of whom were tested for CRP. Antibiotic prescribing before and after CRP testing did not differ ("pre-test" 31%, "post-test" 28%; 95% CI of difference -7 to 1). GPs changed their decision after POC CRP testing in 200 patients (27%). In 41% of the tested patients the indication for testing was in accordance with the guidelines. We therefore concluded that POC CRP testing did not reduce overall antibiotic prescribing when this was used by GPs who already had a low antibiotic prescribing rate. POC CRP did influence GPs to change their decision about antibiotic prescribing in patients with acute cough.

We conducted a secondary analysis of this observational study in **chapter 7** to measure recruitment of patients and to evaluate recruitment bias. We compared recruited patients with all eligible patients to assess representativeness of the included subjects. Eligible patients were adults presenting with acute cough (less than 3 weeks). The frequency of POC CRP testing and the proportion of prescribed antibiotics were compared between recruited and non-recruited patients. As measure of bias ORs with accompanying 95% CIs for the association between CRP level and antibiotic prescribing were computed. Of all 1473 eligible patients 348 (24%) were recruited in the observational study (described in chapter 6). In recruited patients POC CRP tests were conducted and antibiotics prescribed more frequently as compared to non-recruited patients (81% vs. 6% and 44% vs. 29%, respectively). The ORs were 18.2 (95% CI 9.6-34.3), 30.5 (95% CI 13.2-70.3) and 3.8 (95% CI 0.9-14.8) in respectively all eligible patients, the recruited and the non-recruited patients. We concluded that selective

recruitment resulted in an overestimation of POC CRP test use, of antibiotic prescribing and of the association between POC CRP tests and antibiotic prescribing.

Finally, in **chapter 8** we discussed the conclusions of all previous chapters by putting them in a broader perspective. Firstly we discussed whether diagnostic models for CAP are helpful in daily practice. The utility and usefulness of prediction models are hampered by imperfection (i.e. imprecision) and their complexity (i.e. not easy to apply during conventional consultations). Secondly we discussed heterogeneity and size differences of the used datasets in the individual patient data meta-analysis. We elaborated on the introduction of POC CRP test devices in daily practice and concluded that implementation of the POC CRP test devices is not straightforward. GPs should consider collaboration with experienced diagnostic partners. We raised the question whether implementation of POC CRP in primary care is premature. We encountered a considerable proportion of GPs in our study that used the test not according to the guideline, and the effect of POC CRP on antibiotic prescribing was disappointing. Because of the discrepancy in effectiveness compared with previous conducted randomized controlled trials in this field we emphasized the need for implementation studies, to closely monitor the actual use of new health care interventions in daily practice. However, also in implementation research a substantial risk of biased inclusion potentially threatens the validity of the conclusions. Finally room for future improvement for LRTI management in primary care is discussed.


S


N

Nederlandse Samenvatting
Dutch summary


N

Nederlandse Samenvatting | Dutch summary

Hoesten is een veelvoorkomende klacht en lage luchtweginfecties (LLWIs) komen frequent voor in alle seizoenen (**hoofdstuk 1**). Voor huisartsen is het belangrijk om in het diagnostische behandelproces van LLWIs onderscheid te kunnen maken tussen ernstige aandoeningen zoals longontsteking (Engels: community acquired pneumonia, afgekort CAP) en aandoeningen die vanzelf overgaan. Tevens is het belangrijk om onnodig voorschrijven van antibiotica te voorkómen. Huisartsen baseren hun behandeling van LLWIs primair op anamnese en lichamelijk onderzoek. Er bestaan meerdere predictiemodellen om CAP te voorspellen met de informatie die verkregen wordt uit de anamnese en het lichamelijk onderzoek. Mogelijk kan de nauwkeurigheid van deze diagnostische voorspelling verbeterd worden door toevoeging van aanvullende diagnostiek. Een veelbelovende aanvullende test is de bepaling van C-reactief eiwit (Engels: C-reactive protein, afgekort CRP). CRP kan gemeten worden met een sneltest. Deze sneltest meet de concentratie van CRP in capillair bloed met een vingerprik en het resultaat is binnen een paar minuten beschikbaar. De CRP sneltest is daarom goed bruikbaar voor de huisarts om tijdens een consult te beslissen over voorschrijven van antibiotica. Verschillende studies hebben aangetoond dat de CRP sneltest kosten-effectief is en dat het gebruik ervan het aantal antibiotica voorschriften laat dalen. De CRP sneltest werd opgenomen in de richtlijn 'acuut hoesten' van het Nederlands Huisartsen Genootschap (NHG, 2011). Deze (en andere, internationale) richtlijnen adviseren wanneer CRP getest dient te worden, en hoe de uitslagen vervolgens kunnen worden geïnterpreteerd bij patiënten die zich presenteren met acuut hoesten bij de huisarts. Tot dusver is weinig bekend over in hoeverre huisartsen deze richtlijn volgen en wat de invloed is van de introductie van de CRP sneltest in de dagelijkse praktijk op het voorschrijven van antibiotica. De richtlijn geeft geen advies of informatie over welk CRP sneltest apparaat het meest geschikt is voor de huisartsenpraktijk. Studies die de verschillende CRP sneltest apparaten (direct) met elkaar vergeleken zijn schaars en het is onduidelijk in hoeverre verschillen in diagnostische nauwkeurigheid van de sneltesten samenhangen met de validiteit om CAP vast te stellen.

In dit proefschrift beoogden wij om de (toegevoegde) diagnostische waarde van CRP in het diagnostische proces van CAP in de eerstelijns gezondheidszorg te beschrijven, evenals de geschiktheid van de verschillende CRP sneltesten voor de eerstelijns gezondheidszorg. Tevens werd onderzocht wat de invloed is van CRP sneltest resultaten op de beslissing van de huisarts om wel of geen antibiotica voor te schrijven.

In **hoofdstuk 2** werden alle gepubliceerde predictiemodellen voor CAP (gebaseerd op informatie uit anamnese en lichamelijk onderzoek) in de eerste lijn geïdentificeerd en extern gevalideerd in de individuele patiënt gegevens (Engels: individual patient data, afgekort IPD) van eerder uitgevoerde diagnostische studies. Zes predictiemodellen en IPD van acht diagnostische studies (n totaal=5308, prevalentie van CAP 12%) werden geïncludeerd. De prestatie van de zes predictiemodellen werd vergeleken in termen van discriminatie (tussen ziek en niet ziek) en kalibratie. Om de discriminatie

te bepalen gebruikten wij de gepoolde oppervlakte onder de receiver operating characteristic curve (een plot van sensitiviteit tegen 1-specificiteit) (Engels: Area under the ROC curve, afgekort AUC) en de delta AUC. Met de delta AUC wordt het verschil tussen de prestatie van een individueel model en de gemiddelde prestatie van alle modellen binnen een IPD set weergegeven. Om de kalibratie weer te geven gebruikten wij kalibratieplots. De predictiemodellen van van Vugt *e.a*. en Heckerling *e.a*. lieten de hoogste gepoolde AUCs zien van respectievelijk 0.79 (95% betrouwbaarheidsinterval (BI) 0.74-0.85) en 0.72 (95% BI 0.68-0.76). Andere modellen van Diehr *e.a*., Singal *e.a*., Melbye *e.a*. en Hopstaken *e.a*. lieten gepoolde AUCs van respectievelijk 0.65 (95% BI 0.61-0.68), 0.64 (95% BI 0.61-0.67), 0.56 (95% BI 0.49-0.63) en 0.53 (95% BI 0.5-0.56) zien. De delta AUCs van de modellen lieten een vergelijkbare rangschikking zien. De modellen van van Vugt *e.a*. en Singal *e.a*. lieten een nauwe overeenkomst tussen daadwerkelijk waargenomen CAP en voorspelde kans op CAP zien, bij andere modellen ontbrak een dergelijke nauwe overeenkomst. Het model van van Vugt *e.a.* liet de beste discriminatie zien in combinatie met een redelijke-tot-goede kalibratie over alle IPD van de verschillende studiepopulaties. Dit model beschouwen wij daarom als de belangrijkste kandidaat voor gebruik in de eerstelijns gezondheidszorg.

In **hoofdstuk 3** kwantificeerden wij de toegevoegde diagnostische waarde van CRP waarden als aanvulling op anamnese en lichamelijk onderzoek in de IPD van acht diagnostische studies. CRP werd toegevoegd aan een uniform predictiemodel dat de dagelijkse klinische praktijk weerspiegelt. Verbetering in discriminatie (tussen ziek en niet-ziek) en risicoclassificatie werden eerst berekend binnen alle individuele studies en vervolgens voerden wij een meta-analyse uit over alle studies samen. In alle datasets verbeterde de discriminatie na het toevoegen van CRP aan het predictiemodel, met een gemiddelde toename van de AUC van 0.075 (variërend tussen 0.02 en 0.18). Toevoegen van CRP aan het predictiemodel vergrootte de proportie patiënten zonder CAP die correct geclassificeerd waren als laag risico patiënten van 28% naar 36%. De proportie patiënten met CAP die (onterecht) in de laag risicogroep terecht kwamen bleef gelijk (n=4), de proportie patiënten in de intermediaire risicogroep nam af van 56% naar 51% en de proportie patiënten met een correct geclassificeerd hoog risico op CAP nam toe van 63% naar 70%. Het inzetten van CRP als aanvullend onderzoek bij patiënten die zich presenteren met symptomen van LLWIs versterkte de discriminatie tussen patiënten met en zonder CAP en verbeterde de diagnostische risicoclassificatie, hoewel een grote groep met diagnostische onzekerheid (m.a.w. een intermediair risico op CAP) bleef bestaan.

In **hoofdstuk 4** hebben wij de volgende vijf CRP sneltesten geëvalueerd: Afinion en NycoCard Reader II (beiden Alere), Eurolyser Smart 700|340 (Eurolyser), QuikRead go en QuikRead 101 (beiden Orion Diagnostica). De resultaten werden vergeleken met die van een standaard immunoturbidimetrische methode, uitgevoerd op een routine laboratorium analyzer (Olympus AU 2700, Beckman Coulter; de referentie standaard). Analytische prestaties en overeenkomsten met de laboratorium referentie standaard werden geanalyseerd voor de vijf verschillende sneltesten en varieerden aanzienlijk. De

binnen dag variatiecoëfficiënt varieerde van 2.6% (QuikRead go) tot 19.4% (Eurolyser) voor lage CRP waarden (<20 mg/L), en 1.1% (QuikRead go) tot 17.5% (Eurolyser) voor hoge CRP waarden (>100 mg/L). De dag tot dag variatiecoëfficiënt varieerde van 4.6% (Afinion) tot 30.5% (Eurolyser) voor lage CRP waarden en 4.0% (QuikRead go) tot 18.0% (Eurolyser) voor hoge CRP waarden. Voor alle sneltest apparaten was het totale gemiddelde verschil met de laboratorium referentiestandaard <4.2 mg/L. Echter, bij hoge CRP waarden (>100 mg/L) daalde de overeenkomst met de laboratorium referentiestandaard systematisch bij alle CRP sneltesten. Vervolgens werd de gebruiksvriendelijkheid van de verschillende sneltesten beoordeeld. Informatie over apparaat eigenschappen werd verzameld uit gebruiksaanwijzingen van fabrikanten (januari 2012). Afinion en Eurolyser werden als meest gebruiksvriendelijk beoordeeld door 20 huisartsen en assistenten die niet bekend waren met het gebruik van CRP sneltesten.

In **hoofdstuk 5** stelden wij ons als doel om te bepalen of de verschillen die wij vonden in hoofdstuk 4 zich (betekenisvol) zouden doorvertalen naar verschillen in nauwkeurigheid en/of verschillen in de toegevoegde waarde van CRP bij het voorspellen van CAP bij volwassen patiënten die zich bij de huisarts presenteren met acuut (minder dan 3 weken) hoesten. Wij voerden een geneste case-controle studie uit waarbij gegevens werden gebruikt van volwassen eerstelijns patiënten met acuut hoesten in 12 verschillende Europese landen (www.grace-lrti.org). Veneus bloed van 100 willekeurig gekozen patiënten met CAP en 100 willekeurig gekozen patiënten zonder CAP werd getest met de vijf verschillende CRP sneltestapparaten en de laboratoriumtest. Accuratesse-maten van CRP als losstaande test waren vergelijkbaar voor alle vijf de CRP sneltest apparaten en de laboratoriumtest. Om de toegevoegde waarde van CRP als aanvulling op informatie uit anamnese en lichamelijk onderzoek te bepalen (het 'klinische model'), werden de AUC voor het klinsiche model mét CRP en de AUC voor het klinische model zónder CRP berekend. Alle sneltestapparaten lieten een vergelijkbare toegevoegde waarde zien van CRP; de AUCs van de verschillende sneltest apparaten en de laboratoriumtest varieerden tussen 0.79 en 0.80, en waren hoger dan het klinische model zónder CRP (0.70). Multivariabele ORs waren gelijk voor alle test apparaten (1.2). Wij concludeerden dat de verschillen in resultaten die wij zagen in hoofdstuk 4 zich niet doorvertaalden in klinisch relevante verschillen wanneer de CRP sneltest wordt gebruikt om CAP te voorspellen bij volwassen patienten die zich presenteren met acuut hoesten bij de huisarts.

Om het effect van de CRP sneltest op het aantal antibiotica voorschriften te bepalen evenals de mate waarin de de richtlijn gevolgd werd na de implementatie van de CRP sneltest in de dagelijkse praktijk voerden wij een observationele studie uit die beschreven werd in **hoofdstuk 6**. Van februari 2012 tot februari 2013 werden 40 huisartsen uit regio Utrecht en Rotterdam geïnstrueerd om de CRP sneltest volgens de huidige richtlijn te gebruiken bij volwassen patiënten die zich presenteerden met acuut hoesten. Na routinematige anamnese en lichamelijk onderzoek rapporteerden de huisartsen of zij wel of niet van plan waren om antibiotica voor te schrijven (de "pre-test" beslissing). Wanneer zij besloten een CRP sneltest te laten uitvoeren, rapporteerden

zij daarna of zij bij hun beslissing bleven of veranderden van besluit om wel of geen antibiotica voor te schrijven (de "post-test" beslissing). De primaire uitkomstmaat was het verschil tussen het "pre-test" en "post-test" antibiotica voorschrijfpercentage. De secundaire uitkomstmaat was het percentage patiënten bij wie de beslissing om wel of geen antibiotica voor te schrijven veranderde na de CRP sneltest. In totaal includeerden de huisartsen 939 patiënten waarvan 78% getest werden. Het percentage antibiotica voorschriften was niet verschillend ("pre-test" 31% en "post-test" 28%; 95% BI van het verschil -7 tot 1). De huisartsen veranderden hun beslissing na de CRP test in 200 patienten (27%). Bij 41% van de geteste patienten was de indicatie om te testen in overeenkomst met de richtlijn. Wij concludeerden dat de CRP sneltest het aantal antibiotica voorschriften niet verminderde, wanneer deze test gebruikt werd door huisartsen met een algeheel laag antibiotica voorschrijfpercentage. De CRP sneltest beïnvloedde een deel van de huisartsen wel bij hun beslissing om wel of geen antibiotica voor te schrijven. Tevens zagen wij dat de test vaker en anders werd gebruikt dan wordt voorgesteld in de richtlijn.

In **hoofdstuk 7** beschreven wij de secundaire analyse die wij uitvoerden in het kader van de observationele studie beschreven in hoofdstuk 6 om te bepalen of er kans was op vertekening van de studieresultaten door selectieve inclusie van patiënten. We onderzochten voor de periode maart tot september 2012 in vier verschillende huisartspraktijken welke patiënten in aanmerking kwamen om geïncludeerd te worden en bepaalden welk percentage hiervan daadwerkelijk werden geïncludeerd. Patiënten die in aanmerking kwamen om geïncludeerd te worden waren alle volwassen patiënten die zich bij de huisarts presenteerden met acuut hoesten (korter dan 3 weken). Deze groep bestond uit geïncludeerde patiënten en niet geïncludeerde patiënten. Deze twee groepen vergeleken wij met elkaar voor o.a. het percentage CRP testen en het percentage antibiotica voorschriften. Als maat voor vertekening van de resultaten berekenden wij de ORs (met 95% BI) voor de associatie tussen CRP en antibiotica voorschriften. Van alle 1473 patiënten die in aanmerking kwamen om geïncludeerd te worden werden er 348 (24%) daadwerkelijk geïncludeerd in de observationele studie die beschreven is in hoofdstuk 6. Bij de geïncludeerde patiënten werd er vaker CRP bepaald en antibiotica voorgeschreven dan bij de niet-geïncludeerde patiënten (respectievelijk 81% versus 6% en 44% versus 29%). De ORs waren 18.2 (95% BI 9.6-34.3) voor alle patiënten die in aanmerking kwamen, 30.5 (95% BI 13.2-70.3) voor de geïncludeerde patiënten en 3.8 (95% BI 0.9-14.8) voor de niet-geïncludeerde patiënten. Wij concludeerden dat de selectie in geïncludeerde patiënten geresulteerd heeft in een overschatting van het aantal uitgevoerde CRP testen, een overschatting van het aantal antibiotica voorschriften en een overschatting van de associatie tussen CRP en antibiotica voorschriften.

Uiteindelijk bespraken wij in **hoofdstuk 8** de conclusies van de eerdere hoofdstukken en stelden wij deze ter discussie door ze in een breder perspectief te plaatsen. Als eerste bespraken wij of diagnostische predictiemodellen nuttig zijn in de dagelijkse praktijk. Zowel het nut als de bruikbaarheid van predictiemodellen wordt belemmerd doordat ze niet perfect zijn (maar onnauwkeurig in de voorspelling) en door de com-

plexiteit ervan (m.a.w. niet gemakkelijk te gebruiken tijdens reguliere consulten). Ten tweede bespraken wij de heterogeniteit en het verschil in grootte van de gebruikte datasets in de IPD meta-analyse. Wij gingen verder in op de introductie van de CRP sneltest in de dagelijkse huisartsenpraktijk en concludeerden dat implementatie niet zo eenvoudig is als het mogelijk op het eerste gezicht lijkt. Huisartsen zouden wanneer zij de test in hun praktijk invoeren in ieder geval moeten overwegen om samen te werken met ervaren partners in diagnostiek. Mogelijk is de implementatie van de CRP sneltest in de huisartsenpraktijk enigszins voorbarig geweest. Een aanzienlijk deel van de huisartsen in onze studie gebruikten de test niet volgens de richtlijn en het effect op (reductie van) het percentage antibiotica voorschriften was teleurstellend. Vanwege de tegenstrijdigheid in uitkomsten van onze observationele studie (hoofdstuk 6) en de in het verleden uitgevoerde gecontroleerde gerandomiseerde onderzoeken benadrukten wij de noodzaak voor implementatiestudies, zodat het daadwerkelijk gebruik van innovaties in de eerstelijns gezondheidszorg nauwlettend gevolgd kan worden. Echter, ook in implementatieonderzoek bestaat er een wezenlijk risico op vertekende resultaten door selectieve inclusie wat validiteit van de conclusies potentieel bedreigt. Uiteindelijk bespraken wij ruimte voor toekomstige verbeteringen in het beleid voor patiënten met LLWIs in de huisartsenpraktijk.


A

Acknowledgements
Dankwoord


D

Acknowledgements | Dankwoord

Dit proefschrift is tot stand gekomen met hulp en steun van vele mensen, die ik hier graag wil bedanken. Een aantal mensen wil ik graag in het bijzonder danken voor hun betrokkenheid bij mijn promotie, allereerst mijn promotoren en co-promotoren voor hun prettige supervisie.

Prof. Dr. N.J. de Wit, geachte promotor, beste Niek. Bedankt voor de kans en mogelijkheden die je me hebt geboden met dit onderzoekstraject. Je wist me te inspireren en onze besprekingen altijd boeiend en tevens gestructureerd te laten verlopen. Ik waardeer je vertrouwen en steun; fijn dat je altijd de grote lijn in de gaten had en hield en tijd wist vrij te maken wanneer dat nodig was.

Prof. Dr. T.J.M. Verheij, geachte promotor, beste Theo. Bedankt voor de geboden laagdrempelige hulp gedurende het gehele traject. Ik bewonder je kennis en expertise op het gebied van luchtweginfecties in de huisartspraktijk. Je netwerk binnen het Europese eerstelijns onderzoek op het gebied van (luchtweg)infecties is enorm en nogmaals bedankt dat ik via jou deze zeer kundige maar vooral ook vriendelijke groep onderzoekers heb mogen leren kennen op het GRIN congres in Ierland (ook een mooie kans om mijn skills "Iers volksdansen met co-auteurs" te ontdekken).

Dr. A.C. van den Pol, geachte co-promotor, beste Alma. Ik ben zo blij dat je me belde in de zomer van 2011. Bedankt voor al je hulp en betrokkenheid in de afgelopen jaren. Je wist me iedere keer te inspireren en motiveren en je stond altijd voor mij klaar, ondanks dat deze periode voor jou -op z'n zachtst gezegd- behoorlijk turbulent is verlopen. Je bent een groot voorbeeld voor mij!

Dr. B.D.L. Broekhuizen, geachte co-promotor, beste Lidewij. Bedankt voor je altijd snelle reacties op mijn vragen en verzoeken. Ik heb enorm veel bewondering hoe je het onderzoek, je werk in de huisartspraktijk en je gezins- en sociale leven weet te combineren. En altijd vrolijk en vriendelijk.

In de verschillende fasen van het onderzoek en de huisartsenopleiding heb ik (afwisselend) intensief samengewerkt met een grote groep mensen uit verschillende vakgebieden. Ik beschouw het als een voorrecht dat jullie allemaal tijd voor mij hebben gemaakt om me uitleg en hulp te geven.

In het eerstelijns diagnostisch centrum Saltro heb ik mogen samenwerken met Sanne van Delft, eerst als afdelingshoofd klinische chemie en later als wetenschappelijk programmamanager. Ik vond het altijd heel prettig om met je samen te werken, dank voor je persoonlijke introductie in het laboratorium bij Attika en Zakia, en de aangename combinatie van gezelligheid en professionaliteit tijdens het plaatsen van de sneltest-

apparatuur in de huisartsenpraktijken. Tevens wil ik Rogier Hopstaken als vakspecialist point-of-care bedanken voor zijn constructieve bijdragen aan dit proefschrift, zowel vanuit de hoek van het laboratorium en als inspirerend voorbeeld als huisarts-onderzoeker. Het werken met de dataset 'Hopstaken *et al*.' voor de individuele patiënten data (IPD) meta-analyse was een stuk minder 'droog' omdat ik je al kende vanuit het laboratorium. Verder ook mijn dank voor Annette Kooijman-Buiting voor de introductie in de (terminologie van) de klinische chemie, wat voor mij voordat ik aan dit onderzoekstraject begon een volledig onbekend terrein was. Verder ook dank aan de medewerkers van Star-MDC waaronder Linda Nekeman en de POC deskundigen die me in Rotterdam op sleeptouw hebben genomen.

En wat fantastisch dat er binnen het Julius Centrum in het UMCU de mogelijkheid is voor laagdrempelig contact met mensen uit meerdere onderzoeksgroepen en vakgebieden. Beste Joris de Groot, bedankt dat je de afgelopen jaren hebt gefungeerd als spreekwoordelijke rots in de methodologische branding en licht in de duisternis wat betreft het computerprogramma 'R'. Je hebt me veel geleerd over methodologie zowel in de beginfase van de promotie waarbij je mee bent gegaan naar het laboratorium als in een later stadium samen met Hans Reitsma bij de IPD meta-analyse. Mijn speciale dank ook voor Alwin Schierenberg; bedankt voor je belangrijke bijdrage aan de IPD meta-analyse. Wat fijn dat we collega's zijn geworden na jouw stage (waarvan ik het grootste deel met zwangerschapsverlof was, maar dat terzijde).

I would like to acknowledge all collaborators of the IPD meta-analysis, including Hasse Melbye, Tim Rainer, Johann Steurer, Anette Holm, Ralph Gonzales, Willy Graffelman, Arie Knuistingh Neven, Saskia van Vugt, Rogier Hopstaken and Geert-Jan Dinant; thank you for your time, effort and of course the use of your datasets. I would also like to thank all (other) co-authors who contributed to the other articles/manuscripts described in this thesis; including Chris Butler, Paul Little, Herman Goossens, Margareta Ieven and Christine Lammens. I am grateful that we were able to use some of the GRACE venous blood samples. Verder wil ik ook alle huisartsen en praktijkassistenten uit de praktijken in Utrecht en Rotterdam die aan onze observationele studie hebben meegewerkt, bedanken voor de enthousiaste medewerking. Data management medewerkers in het bijzonder Nicole Boekema en Julia Velikopolskaia, bedankt voor jullie medewerking bij (het opzetten van) de dataset in 'research online'.

D

De beoordelingscommissie bestaande uit Prof. Dr. R.A.M.J. Damoiseaux, Prof. Dr. R.J.P.M. Scholten, Prof. Dr. F.G. Schellevis, Prof. Dr. M.J.M. Bonten en Prof. Dr. P. Little wil ik bedanken voor het lezen en kritisch beoordelen van het manuscript van dit proefschrift / I would like to thank you for reading the manuscript of this thesis as part of the appraisal committee.

Collega promovendi en postdoc onderzoekers van het HAG- en infectieoverleg, medeaiotho's (arts-in-opleiding-tot-huisarts-onderzoeker), kamergenoten van 6.118, kamer- en ganggenoten van het van Geuns gebouw; fantastisch dat ik met mijn vragen altijd bij jullie terecht kon. Ook onze pauzes waren altijd een mooie combi van gezelligheid en het uitwisselen van handige onderzoekstips.

Christiana Naaktgeboren en Marieke Bolkenbaas; ik was erg blij met jullie als kamergenoten en vraagbaken voor respectievelijk methodologie en huisartsgeneeskunde-combineren-met-onderzoek; bedankt dat jullie mijn paranimfen willen zijn.

Verder wil ik graag mijn huisartsenopleiders en medewerkers van huisartsenpraktijk 'Tuindorp', het Leidsche Rijn Julius gezondheidscentrum 'Parkwijk' in Utrecht en gezondheidscentrum 'De Nije Veste' in Nijkerk bedanken, waaronder speciale dank voor opleiders Erik de Laat, Hanneke Smits-Pelser, Johan van Langen, Jan Cozijnsen en Hans van Daalen. Het grote voordeel van een onderbroken opleidingstraject (afwisselend in de praktijk werken en bezig zijn met onderzoek) en 'duo-opleiders', is dat ik van jullie allemaal heb mogen leren, ieder met zijn eigen stijl. Jullie zijn allemaal goede voorbeelden voor mij en ik hoop jullie nog vaak als collega's te zullen treffen.

Ook mijn dank aan de SBOH die dit als prettige werkgever mogelijk heeft gemaakt, tevens dank aan al mijn huisartsen-in-opleidings(groeps)genoten; dankzij mijn AIOTHO-traject heb ik veel verschillende collega's(-in-opleiding) ontmoet en veel kunnen leren tijdens de terugkomdagen. Vooral het uitwisselen van ervaringen was een geweldige manier om te reflecteren op het dagelijkse werk.

Verder wil ik van deze bijzondere mogelijkheid (om belangrijke mensen in mijn leven via deze papieren weg te kunnen bedanken) nog gebruik maken om een aantal voor mij speciale mensen te bedanken. Geneeskunde vriendinnen en ondertussen collega's (Veerle, Karien, Olga, Marieke, Linda, Indra, Liesbeth, Martine en uiteraard ook jullie almaar uitbreidende aanhang); heerlijk om met jullie te kunnen spuien over het werk maar ook niet-werk gerelateerde zaken, AEGEE vriendinnen (Gwenn, Charlotte, Heleen, Sandra, Beate, Annemarie, Marijtje, Henriette & Marieke), mijn trouwe (middelbare school)vriendinnen (Suzanne, Eveline, Nicole & Sanne), vrienden-op-afstand (Olivia & Matthijs) en natuurlijk Barbara, Jan-Willem, Tijmen, Floris & Pieter; het doorbrengen van mijn vrije tijd met jullie is fantastisch en een zeer goede afleiding van het werk! Familie en schoonfamilie, in het bijzonder mijn vader en moeder, Anne, Eline & Jan-Willem; bedankt dat jullie er allemaal voor mij zijn!


CV

Curriculum Vitae


CV

Curriculum Vitae

Margaretha Catharina (Marloes) Minnaard was born on March 18th 1979 in Leiden. In 1998 she started studying Medicine and in 2001 combined this with studying Psychology at Utrecht University. In 2003 she obtained her propaedeuse in Psychology and in 2006 obtained her degree as medical doctor. After working as a resident in different clinical areas (internal medicine/pulmonology and psychiatry) and preventive health care, she started her vocational training for general practice at the department of general practice of Utrecht University in 2009. In 2011 she combined the general practice vocational training with research activities as a PhD student at the Julius Center for Health Sciences and Primary Care in the University Medical Center Utrecht, and was supervised by Prof. dr. N.J. de Wit, Prof. dr. T.J.M. Verheij, Dr. A.C. van de Pol and Dr. B.D.L. Broekhuizen. In 2015 she graduated as a general practitioner (GP) and is currently working as locum general practitioner.