Clinical Decision Support in Pediatric Care

KLINISCHE BESLISSINGSONDERSTEUNING IN DE ACUTE KINDERGENEESKUNDE

Jolt Roukema

Cover: 'Het zieke kind' Antonio Mancini (1875). Reprinted with permission of 'Museum Mesdag', The Hague, The Netherlands.

The work presented in this thesis was performed at the Erasmus Medical Center, Sophia Children's Hospital, Department of General Pediatrics, Rotterdam, the Netherlands. The studies described in this study were supported by a grant from the ErasmusMC healthcare efficiency research program (VAZ-doelmatigheids-onderzoek).

ISBN: 90-9021084-9
ISBN-13: 978-90-9021084-1

Graphic design by Mark Winkel, DeMediaWinkel
Printed by Printpartners Ipskamp BV Enschede

Clinical Decision Support in Pediatric Care

KLINISCHE BESLISSINGSONDERSTEUNING IN DE ACUTE KINDERGENEESKUNDE

PROEFSCHRIFT

Ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
woensdag 15 november 2006 om 13.45 uur

door
Jolt Roukema
geboren te Emmen

Promotiecommissie

Promotoren:	Prof.dr. H.A. Büller
	Prof.dr. J. van der Lei
Overige leden:	Dr. E.W. Steyerberg
	Prof.dr. H.M. Oudesluys-Murphy
	Prof.dr. P. Patka
Copromotor:	Dr. H.A. Moll

Contents

Chapter 1

General Introduction

Introduction

Evidence based medicine has become the standard for modern clinical practice, i.e. physicians are expected to deliver care according to the latest insights. Keeping up to date with the latest research findings and clinical evidence is however hardly feasible in daily practice. Therefore literature reviews and guidelines are published, and clinical prediction and decision rules are developed to support the utilization of the latest evidence in clinical decision making.

Several aspects of supporting clinical decision making are discussed in this thesis. First the use of an electronic medical record (EMR) in general pediatrics is evaluated. Electronic medical records are increasingly being adopted and may play an important role in the provision of clinical decision support. Second, the development of clinical prediction rules for the diagnostic management of febrile children is described. Finally the integration of an EMR and a clinical decision rule in a clinical decision support system, used in a pediatric hospital emergency department, is presented.

The overall **aim** of the studies described in this thesis was to investigate the diagnostic process of (febrile) children presenting to the pediatric emergency department, and to describe aspects of this process as a base for clinical decision support systems.

The febrile child

Fever is among the most common presenting signs of illness in children attending hospital emergency departments (EDs). Between 10 and 20 percent of all pediatric visits to hospital EDs are due to febrile illness (1-3). Fever in young children causes concern among parents and caregivers, and is a diagnostic dilemma for general practitioners and pediatricians (4). Although most children will have a benign, self-limiting viral disease, a small proportion will have a serious bacterial infection (SBI). Especially young children with fever without apparent source (FWS), seem to be at risk for SBI (5, 6). FWS is defined as a body temperature ≥ 38°C and no apparent source of the fever after history taking by a pediatrician (6). At the ED of the Sophia Children's Hospital, 12% of the children with FWS was diagnosed with a SBI, of whom 50% had pneumonia, 25% urinary tract infection, and 25% other SBIs (e.g. meningitis, sepsis, bacterial gastroenteritis). The majority of 88%, however, was diagnosed with benign viral disease, mostly (60%) a viral upper

respiratory tract infection, or a viral gastroenteritis (15%). The differentiation of children with SBI from the vast majority of children with benign illness is a daily challenge for pediatricians working in hospital EDs as clinical presentation is often a-specific (7, 8).

Over the past decades the management of febrile children has been subject of discussion, and various management strategies were proposed against a background of the introduction of new vaccines targeted at pathogens responsible for SBI (7, 9-12). For example, after the introduction of the conjugate vaccine against *Haemophilus influenza* type B (Hib) in the 1990's, the rate of (occult) bacteremia and SBI due to Hib in young children decreased significantly (12-14). A further decrease in incidence of SBI is to be expected with the introduction of a conjugate vaccine against *Streptococcus pneumoniae*; an important causative pathogen of pneumonia and meningitis (15). This knowledge warrants a continuing appraisal of the diagnostic process for febrile children attending the hospital ED. A more selective diagnostic policy in febrile children seems to be justified. The decreasing risk of SBI but the possible serious consequences of missing an SBI complicates the diagnostic process even more.

Electronic Medical Record

To date, paper-based medical records are still the usual medium for recording and collecting patient data by health professionals. Several drawbacks of the paper record have however been identified, such as illegible handwriting, incomplete records, data fragmentation, and poor availability (16). In addition, paper records often lack overview, which may threaten the continuity and quality of care.

Implementation of an Electronic Medical Record (EMR) has the potential to improve the quality of patient care as it may enhance availability, readability, and data quality (17). As most potential benefits of an EMR rely on structured, coded data, the use of structured data entry is preferable to the use of free-text. Structured data entry (SDE) applications may offer additional advantages as they can: prompt for completeness, provide a better ordering for lookup and retrieval of patient information, and perform validity checks on recorded data. Furthermore, structured patient data may facilitate clinical research and the application of computerized clinical decision support (16-18).

Despite the potential benefits, user acceptance will be the major barrier in implementing an

EMR. Documenting patient data in a structured format will put an extra burden on already busy physicians working under time constraints. Hence, the advantages of structured patient data must outweigh the disadvantages of documenting the data in a structured format (19). Functionality and user interface are therefore crucial for successful implementation (20, 21).

In this thesis the use of a structured data entry application (OpenSDE) for the documentation of patient history and physical examination is discussed. We will focus on the completeness of documented patient information as compared to the documentation of the same information in a paper record, and the feasibility of an EMR in clinical pediatric practice (**aim 1**). Furthermore, the uniformity in documentation of the same patient data by different physicians is evaluated (**aim 2**).

Evidence, guidelines and clinical decision rules

Clinical decisions regarding diagnostic and therapeutic strategy are made on a daily basis by physicians. Decision making in medicine is often a complicated process, driven by patient and population characteristics, clinical findings, physician's experience, and available knowledge or evidence regarding the problem at hand. Differences in these determinants may lead to variation and inconsistencies in decision making in comparable clinical situations.

The practice of 'evidence based medicine' (EBM) has received increasing attention during the past ten years and has become the standard in modern medicine. One of the many definitions states that EBM is "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research" (22). Evidence can be implemented in clinical practice in different ways, for example by developing a clinical practice guideline. Guidelines are defined as: "systematically developed statements to assist practitioner and patient decisions about appropriate healthcare for specific clinical circumstances" (23). More specifically, a clinical prediction rule may be developed to estimate the probability of a certain diagnostic outcome being present, for which patients need a specific diagnostic work-up. When the result of a prediction rule is linked to a management recommendation, it becomes a clinical decision rule (CDR): "a clinical tool that quantifies the individual contributions that various components of the history, physical examination, and basic laboratory results make towards the diagnosis, prognosis, or likely response to treatment in an

individual patient" (24). A CDR can thus be viewed as an individualized guideline, taking patient specific characteristics into account.

The development of a CDR consists of several steps (24, 25). First the need for a CDR for a specific clinical practice problem must be determined; is there a clinically significant problem? Are there inconsistencies in clinical practice? Does practice efficiency need to be improved? Secondly, the outcome and predictor variables must be defined. In the case of febrile children, pediatricians are interested in presence of a serious bacterial infection (SBI), which requires immediate attention and treatment. From experience and literature potential predictor variables for the presence of SBI are selected, such as age, temperature and duration of the febrile episode (1, 5, 6, 8, 26). The value of a potential predictor variable is not judged on whether the association between the characteristic and the outcome of interest (e.g. SBI) is statistically significant in a group of children, but on the predictive value of that characteristic in an individual child, i.e. if the characteristic is present, how likely is it that this individual child has SBI? (27). A clinical prediction rule should be derived based on systematic clinical observations and measurements (28). Accuracy, reliability and discriminative ability of the CDR are determined using the dataset on which the CDR was developed, called *internal validation*. Before the CDR can be widely implemented, it should prove to be both accurate and reliable during prospective validation in a broad clinical setting, called *external validation* (29, 30).

In this thesis we will discuss the development of clinical prediction rules for specific diagnostic problems in febrile children attending the hospital ED (**aim 3 and 4**), as well as methodological issues regarding modeling strategies in diagnostic research (**aim 5**).

Clinical Decision Support Systems

When a clinical decision rule (CDR) has proven to be valid and reliable in a broad clinical setting, this does not necessarily mean that the implementation will be successful. The rule will be successful if positive effects on patient or process outcome are observed, or when practice efficiency has improved (31). Guidelines and CDRs are mostly presented as paper flowcharts that still require the physician to assess whether a certain flowchart applies to a particular patient, and if so, what the implications are given the patient's characteristics and test results. Moreover, it

takes approximately five years for a guideline to be adopted in clinical practice, and when adopted, guidelines are often not followed (31, 32). For example, more than 99% of the Canadian emergency physicians reported to be familiar with the Ottawa ankle rule, but only 30% was able to correctly remember the components of the rule. Meanwhile, 90% reported using the rule most of the time, but only 40% based their decision to order radiography on the rule (33).

Clinical decision support systems (CDSS) are developed to bring clinical evidence to the point of care, i.e. the consultation room or the patient's bedside.

Many definitions of CDSS have been proposed (34, 35). In general a CDSS can be viewed upon as a computer system that integrates patient characteristics and available knowledge to provide a patient specific advice on diagnostic or therapeutic management at the time of physician decision making. Patient information should therefore be instantaneously available in an electronic, computer interpretable format, and the knowledge should be integrated as readily executable algorithms.

A CDSS can have many different functions, from flagging abnormal laboratory values to a complicated, algorithm driven patient specific advice on diagnostic or therapeutic management. A few examples of CDSS-functionality in pediatric care are given below:

- *Providing alerts or reminders*: A well known example of an alert is a laboratory system flagging values outside a (age-specific) normal range. An example of a reminder-system is the computerized generation of a reminder for scheduled immunizations. Parents who received an automatically generated reminder for scheduled immunizations (telephone message or letter) were 20% more likely to immunize their child than parents who did not receive a reminder (36).
- *Interpretation of diagnostic tests*: Automated ECG interpretation is a common feature of ECG-acquisition devices nowadays in adult and pediatric care. Computer interpretation may reduce the incidence of ECG interpretation errors and decrease the time necessary to interpret an ECG (37).
- *Supporting differential diagnosis*: An example is ISABEL; a web based differential diagnostic aid for pediatrics. ISABEL showed good clinical accuracy, by displaying the final diagnosis in 83 out of 87 real cases (95%) (38). The system however, does not so much generate an advice or alarm, but rather combines clinical characteristics with possible diagnoses.
- *Generation of management suggestions or advice*: In this thesis we will provide an example of a

CDSS that provides a patient-specific diagnostic management advice based on predicted diagnostic outcome in children with fever without apparent source.

Key features of a CDSS to become successful were recently described: A CDSS should provide decision support as part of routine workflow, use a computer system to generate decision support, provide actionable recommendations rather than risk indications, and deliver decision support at time and location of decision making (39). Directive decision rules seem to be more effective than assistive, i.e. an explicit patient specific recommendation should be given rather than a probability (40). As a CDSS uses clinical evidence (e.g. a CDR), to generate these patient specific recommendations, the effectiveness of a CDSS will be limited by the validity of the evidence base that is used (41).

In this thesis, two clinical decision support systems are evaluated: the first CDSS supports ED-nurses in determining the clinical urgency of patients attending the ED, and to subsequently manage the patient flow based on clinical urgency. This CDSS, the Manchester Triage System, was not specifically developed for determining clinical urgency in a pediatric population. We will discuss the validity of this CDSS in pediatric patients (**aim 6**). The second CDSS estimates the risk of SBI in young children with fever without apparent source attending the ED, and then provides a patient-specific diagnostic management advice. We will focus on compliance with the CDSS and effects of CDSS application on time spent in the ED by patients and the frequency of diagnostic testing (**aim 7**).

Aims of the studies presented in this thesis were:

1. To compare the completeness and uniformity of patient data in an electronic medical record and in a paper record, and to study the feasibility of an electronic medical record in general pediatrics.
2. To analyze the uniformity in structured documentation of the same patient data, by different physicians using a structured data entry application.
3. To develop a clinical prediction rule for pneumonia in children presenting with fever and cough.
4. To assess the diagnostic value of C-reactive protein in the diagnosis of febrile children.
5. To predict diagnostic outcome in children with fever without apparent source in the pediatric emergency department, and to compare the use of polytomous regression analyses with sequentially applied multivariate modeling strategies to predict diagnostic outcome.
6. To validate the Manchester Triage System in a pediatric population.
7. To assess the effects of application of a clinical decision support system for the diagnostic management of children with fever without apparent source.

Outline of this thesis

In the first part of this thesis (**Chapter 2 and 3**) the use of an electronic patient record with structured data entry, for documentation of patient history and physical examination is described. In **Chapter 2** a qualitative and quantitative comparison is made between documentation of patient data in an electronic and in a paper record. In **Chapter 3** the uniformity in documentation of the same patient data by different physicians using structured data entry are described.

In the second part of this thesis, the development and (internal) validation of clinical prediction rules is described. In **Chapter 4** a specific clinical prediction rule is developed for children presenting with fever and cough, to predict the presence of pneumonia. In **Chapter 5** the (added) diagnostic value of C-reactive protein for diagnosing febrile children is assessed. In **Chapter 6** the use of polytomous regression modeling to improve the diagnostic ability of a prediction rule is discussed.

Prediction rules for the presence of serious bacterial infection in children with fever without source are developed using different regression modeling strategies.

In **Chapter 7 and 8**, two clinical decision support systems, used in the pediatric emergency department are evaluated. In **Chapter 7** the Manchester Triage System, used in the ED to support the determination of a patient's urgency level, is validated. In **Chapter 8**, the compliance with and effects of utilization of a clinical decision support system for the diagnostic management of young children with fever without apparent source are described.

In **Chapter 9** the results of the studies presented in this thesis are summarized and future research perspectives are discussed. In **Chapter 10** the study results presented in this thesis are summarized in Dutch.

References

1. Kuppermann N, Fleisher GR, Jaffe DM. Predictors of occult pneumococcal bacteremia in young febrile children. Annals of Emergency Medicine. 1998 Jun;31(6):679-87.
2. Nelson DS, Walsh K, Fleisher GR. Spectrum and frequency of pediatric illness presenting to a general community hospital emergency department. Pediatrics. 1992 July 1, 1992;90(1):5-10.
3. Slater M, Krug SE. Evaluation of the infant with fever without source: An evidence based approach. Emergency Medicine Clinics of North America. 1999 Feb;17(1):97-+.
4. Betz MG, Grunfeld AF. 'Fever phobia' in the emergency department: a survey of children's caregivers. Eur J Emerg Med. 2006 Jun;13(3):129-33.
5. Hsiao AL, Chen L, Baker MD. Incidence and Predictors of Serious Bacterial Infections Among 57- to 180-Day-Old Infants. Pediatrics. 2006 May 1, 2006;117(5):1695-701.
6. Bleeker SE, Moons KGM, Derksen-Lubsen G, Grobbee DE, Moll HA. Predicting serious bacterial infection in young children with fever without apparent source. Acta Paediatrica. 2001 Nov;90(11):1226-32.
7. Baker MD, Bell LM, Avner JR. Outpatient Management without Antibiotics of Fever in Selected Infants. N Engl J Med. 1993 November 11, 1993;329(20):1437-41.
8. Lee GM, Harper MB. Risk of bacteremia for febrile young children in the post-Haemophilus influenzae type b era. Archives of Pediatrics & Adolescent Medicine. 1998 Jul;152(7):624-8.
9. Baraff LJ. Outpatient Management of Fever in Selected Infants. New England Journal of Medicine. 1994 Mar 31;330(13):938-9.
10. Baraff LJ. Management of fever without source in infants and children. Annals of Emergency Medicine. 2000 Dec;36(6):602-14.
11. Isaacman DJ, Rogers KD. Practice Guidelines for Management of Infants and Children with Fever without Source (Fws). Pediatrics. 1994 Feb;93(2):346-7.
12. Lee GM, Fleisher GR, Harper MB. Management of febrile children in the age of the conjugate pneumococcal vaccine: A cost-effectiveness analysis. Pediatrics. 2001 Oct;108(4):835-44.
13. Alpern ER, Alessandrini EA, Bell LM, Shaw KN, McGowan KL. Occult bacteremia from a pediatric emergency department: Current prevalence, time to detection, and outcome. Pediatrics. 2000 Sep;106(3):505-11.
14. Conyn-van Spaendonck MA, Veldhuijzen IK, Suijkerbuijk AW, Hirasing RA. [Significant decline of the number of invasive Haemophilus influenzae infections in the first 4 years after introduction of vaccination against H. influenzae type B in children]. Ned Tijdschr Geneeskd. 2000 May 27;144(22):1069-73.
15. McCarthy P. Fever without apparent source on clinical examination. Curr Opin Pediatr. 2005 Feb;17(1):93-110.
16. Dick RS, Steen EB. The Computer-Based Patient Record: An Essential Technology for Health Care. Revised edition. National Academy Press, Washington DC. 1997.
17. Adams WG, Mann AM, Bauchner H. Use of an electronic medical record improves the quality of urban pediatric primary care. Pediatrics. 2003 Mar;111(3):626-32.
18. van Ginneken AM. The computerized patient record: balancing effort and benefit. International Journal of Medical Informatics. 2002 2002/6;65(2):97-119.
19. Powsner SM, Wyatt JC, Wright P. Opportunities for and challenges of computerisation. The Lancet. 1998 1998/11/14;352(9140):1617-22.
20. Dansky KH GL, Vasey JJ, Barsukiewicz CK. Electronic medical records: are physicians ready? J Healthc Manag.

1999;44(6):440-54.
21. Sittig DF, Kuperman GJ, Fiskio J. Evaluating physician satisfaction regarding user interactions with an electronic medical record system. Proc AMIA Symp. 1999:400-4.
22. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996 January 13, 1996;312(7023):71-2.
23. Field MJ, Lohr KN. Clinical Practice Guidelines: Directions for a New Program. National Academy Press Washington DC. 1990.
24. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. Jama. 1997 Feb 12;277(6):488-94.
25. McGinn TG, Guyatt GH, Wyer PC, Naylor CD, Stiell IG, Richardson WS, et al. Users' Guides to the Medical Literature: XXII: How to Use Articles About Clinical Decision Rules. JAMA. 2000 July 5, 2000;284(1):79-84.
26. Isaacman DJ, Shults J, Gross TK, Davis PH, Harper M. Predictors of bacteremia in febrile children 3 to 36 months of age. Pediatrics. 2000 Nov;106(5):977-82.
27. Kramer MS. Management of the Young Febrile Child: A Commentary on Recent Practice Guidelines. Pediatrics. 1997 July 1, 1997;100(1):128-.
28. Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. N Engl J Med. 1985 September 26, 1985;313(13):793-9.
29. Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research:: A clinical example. Journal of Clinical Epidemiology. 2003;56(9):826-32.
30. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KGM. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. Journal of Clinical Epidemiology. 2003;56(5):441-7.
31. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud P-AC, et al. Why Don't Physicians Follow Clinical Practice Guidelines?: A Framework for Improvement. JAMA. 1999 October 20, 1999;282(15):1458-65.
32. Lomas J, Sisk JE, Stocking B. From evidence to practice in the United States, the United Kingdom, and Canada. Milbank Q. 1993;71(3):405-10.
33. Brehaut JC, Stiell IG, Visentin L, Graham ID. Clinical decision rules "in the real world": how a widely disseminated rule is used in everyday practice. Acad Emerg Med. 2005 Oct;12(10):948-56.
34. Perreault L, Metzger J. A pragmatic framework for understanding clinical decision support. Journal of Healthcare Information Management. 1999;13(2):5-21.
35. Wyatt J, Spiegelhalter D. Field trials of medical decision-aids: potential problems and solutions. Proc Annu Symp Comput Appl Med Care. 1991:3-7.
36. Dini EF, Linkins RW, Sigafoos J. The impact of computer-generated messages on childhood immunization coverage. American Journal of Preventive Medicine. 2000;19(1):68-70.
37. Salerno SM, Alguire PC, Waxman HS. Competency in Interpretation of 12-Lead Electrocardiograms: A Summary and Appraisal of Published Evidence. Ann Intern Med. 2003 May 6, 2003;138(9):751-60.
38. Ramnarayan P, Tomlinson A, Rao A, Coren M, Winrow A, Britto J. ISABEL: a web-based differential diagnostic aid for paediatrics: results from an initial performance evaluation. Arch Dis Child. 2003 May 1, 2003;88(5):408-13.
39. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. 2005 April 2, 2005;330(7494):765-.
40. Reilly BM, Evans AT. Translating Clinical Research into Clinical Practice: Impact of Using Prediction Rules To Make Deci-

sions. Ann Intern Med. 2006 February 7, 2006;144(3):201-9.

41. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al. Clinical Decision Support Systems for the Practice of Evidence-based Medicine. J Am Med Inform Assoc. 2001 November 1, 2001;8(6):527-34.

Chapter 2

Paper versus computer: feasibility of an electronic medical record in general pediatrics

Jolt Roukema, Renske K. Los, Sacha E. Bleeker, Astrid M. van Ginneken, Johan van der Lei, Henriette A. Moll

Pediatrics 2006;117:15-21

Abstract

Background. Implementation of Electronic Medical Record (EMR) systems promises significant advances in patient care as it enhances readability, availability, and data quality.

Structured data entry (SDE)-applications can prompt for completeness, provide a higher level of accuracy and better ordering for lookup and retrieval, and permit validity checks for data quality, research and especially decision support.

A generic structured data entry application (OpenSDE), to support documentation of patient history and physical examination was developed and tailored for the domain of general pediatrics.

Objective. To evaluate OpenSDE for its completeness, uniformity of reporting, and its usability in general pediatrics.

Methods. Four (trainee) pediatricians documented data of 8 first visit patients in the traditional paper-based medical record, and immediately thereafter in OpenSDE (electronic record). The 32 obtained paper records served as common data source for data entry in OpenSDE by the other three physicians (transcribed record). Data entry by two experienced users, containing all patient information present in the paper record, served as control record. Data entry time was recorded, and a questionnaire was used for users' experiences with OpenSDE.

Results. Clinicians documented 44% of all available patient information identically in the paper and electronic record. 25% of all patient information was documented only in the paper record, and 31% was present only in the electronic record. Differences were found in documentation of patient history and physical examination in the electronic record: more information was missing in patient history (38%) compared to physical examination (15%). Furthermore, physical examination contained more additional information (39%) than patient history (21%).

The inter-observer agreement of documentation of patient information from the same data source was fair to moderate with kappa values of 0.39 for patient history and 0.40 for physical examination.

Data entry time in OpenSDE decreased from 25 to less than 15 minutes, indicating a learning effect. The questionnaire revealed a positive attitude towards the use of OpenSDE in daily practice.

Conclusion. OpenSDE seems to be a promising application for the support of physician data entry in general pediatrics.

Introduction

Several weaknesses of the paper-based medical record have been identified, such as illegible handwriting, ambiguous and incomplete data, data fragmentation, and poor availability (1). In addition, the paper record often becomes bulky with time, which leads to lack of overview. As the paper record is still the usual medium for collecting and recording patient data, these weaknesses could impede the continuity and quality of care.

Implementation of Electronic Medical Record (EMR) systems promises significant advances in quality of patient care as it may enhance readability, availability, and data quality (2).

In an EMR the use of structured data is preferable to free-text, as most benefits of EMRs rely on structured, coded data. Structured data entry (SDE) applications can prompt for completeness, provide better ordering for lookup and retrieval, and permit validity checks for data quality, research, and especially decision support (1-3).

Despite potential benefits, user acceptance will be the major barrier in implementing EMR systems, as clinicians will face a change in their practicing habits. The advantages of coded data must outweigh the disadvantages of capturing such data for SDE to become successful in clinical practice (4). Functionality and user interface will therefore be crucial for successful implementation (5, 6).

A generic SDE-application for documentation of narrative data was developed (OpenSDE), and then tailored for the recording of patient history and physical examination in the domain of general pediatrics, a broad specialty (7). The purpose of this study is to analyze the pediatric OpenSDE for its completeness, uniformity of reporting, and applicability in pediatric outpatient care.

Materials.

OpenSDE has been developed at the Erasmus Medical Center, department of Medical Informatics, to support the structured recording of patient data in any medical domain (8).

To record patient history and physical examination, OpenSDE uses a tree of medical concepts that represent the available descriptive options. Each node in the tree is described in further detail by the nodes of its branches. A sequence of nodes represents a clinical "finding" (e.g. a cardiac murmur is described by the sequence physical examination/chest/heart/auscultation/ murmur, Figure 1). The tree of general pediatrics consists of 8,648 nodes with a maximum depth of nine ramifications. Patient history contained 6,312 nodes, and physical examination 2,336. The principle of OpenSDE is that a generic application is used for different domain specific tree structures (e.g. the

pediatric cardiologist uses a tree which is more detailed for cardiac history and physical examination than a general pediatrician).

When recording patient data, OpenSDE displays the tree of clinical findings on the left and the descriptors (e.g. the branching nodes) of the selected clinical finding on an entry form on the right side of the screen (Figure 1). A comment can be added to any finding or description, thereby facilitating the use of free text. The user can also define custom entry forms that may contain a user-defined selection of nodes in the tree for a specific medical problem or task.

Prior to this study, data of 100 existing paper-based pediatric medical records were entered in OpenSDE by experienced users to evaluate ordering and coverage of a tree that was previously developed for the domain of general pediatrics (7). Adjustments were made to improve the usability of the application in practice, consisting of re-arranging and completing the tree-structure, and adding predefined, visible (versus on demand) entry fields for free-text comments.

Methods.

Two pediatricians and two senior residents in pediatrics voluntarily followed a short standardized education-course on the use of OpenSDE. They had never worked with OpenSDE before. In the outpatient department of the Sophia Children's Hospital, they all documented data of 8 first-contact patients (e.g. patients referred to the pediatrician for the first time by their general practitioner) in the regular pediatric paper-based medical record, the "paper record". Immediately after departure of the patient, the same patient information was entered in OpenSDE, the "electronic record". The electronic record was made after departure of the patient to avoid a 'checklist-effect', resulting in incomparable data (9). The participants were aware that the electronic records were not to be used for patient-care, as the paper record was still the standard.

Methods for comparing paper records with electronic records are not readily available (10). We did not endeavour to compare the paper and electronic record with a patient-based gold standard. Therefore, the paper record was transcribed into a control record containing all written information. Two experienced OpenSDE-users (JR, AvG) independently determined which clinical findings, documented in the paper record, could have been entered as structured data in OpenSDE by the participants. Findings in the paper record that could not be structured were entered as free-text. These two entries were compared to reach consensus on an optimal electronic record, containing all information present in the paper record, serving as "control record".

In addition, each (anonymous) handwritten paper record was entered in OpenSDE by the other three physicians, the "transcribed record". Of each record (paper, electronic, and transcribed) we assessed the number of entered details and free-text annotations, and data entry time. We corrected data entry time for number of entered details. IRB review was not required.

The participants received a questionnaire, based on the validated Questionnaire on User Interface Satisfaction (QUIS), a general user evaluation instrument for interactive computer-systems, to obtain their subjective opinion on the usage of and experiences with OpenSDE. Four major topics (content, lay-out, system, and general response) were rated on a six-point scale (11).

The collected data were used to analyze

- Similarities and differences regarding which patient data were documented in the primary paper record versus those entered in OpenSDE, comparing the electronic and control record in each case.
- Inter-observer agreement in describing patient information from a paper record in an EMR, comparing the transcribed and control records in each case for presence of information.
- Users' experiences with OpenSDE, and data entry time during the research period (learning-effect)

Results

32 paper records and a total number of 155 OpenSDE records were obtained: 32 electronic records by primary physicians, 91 transcribed records (5 missing) and 32 control records.

Similarities and differences between the electronic and control record

On average, a total number of 212 findings were entered for each case. Table I shows the average proportion of all entered patient information present in the electronic record, the control record, or both.

In the electronic record, compared to the control record, patient history lacked more information (38%) than physical examination (15%). In addition, physical examination in the electronic record contained more additional information (39%) compared to the control record than patient history (21%).

Table 1. Average number of documented findings in each case (n=32).

Patient information	**Average number of documented findings** electronic and control record	**Electronic and control record** i.e. information present in both	**Only electronic record** i.e. extra information, absent in control record	**Only control record** i.e. information missing in electronic record, present in control record
Patient history*	101	41(41)	21(21)	39(38)
Physical examination*	111	51(46)	43(39)	17(15)
Total*	212	92(44)	64(31)	56(25)

* values represent number of findings (percentages)

Uniformity of data entry from a common data source

On average three transcribed records were made from each paper record by different physicians. On average 135 (95%CI 56-214) findings were recorded in the transcribed records. Table II shows to what extent physicians have entered the information available in the paper record, have entered additional information, and how much information was missing in patient history and physical examination respectively. In the transcribed records, 67% (patient history) and 84% (physical examination) of the information was documented by at least 2 participants uniformly.

Eighty-one percent of the additional information (i.e. information present in the transcribed record, but absent in the control record) in patient history and seventy-two percent in physical examination was entered by only one participant. All transcribed records missed information and contained extra information, the amount of missing and extra information differed.

On average 39% of the missing information in patient history was missing in one of the three transcribed records whereas in physical examination 66% was missing in one of the three transcribed records. On the other hand, 29% of the missing information in patient history and 14% of the missing information in physical examination was missing in all three transcribed records. The average inter-observer agreement statistic (kappa) was 0.39 (95%CI 0.25-0.53) for patient history and 0.40 (95%CI 0.28-0.51) for physical examination.

Table 2. Uniformity of reporting (transcribed records)

	Number of physicians	Patient history	Physical examination
Same information	**2**	30 (24-36)	28 (22-34)
	3	37 (31-42)	56 (48-65)
Additional information	**2**	18 (12-24)	23 (16-30)
	3	1 (0-3)	5 (2-7)
Missing information	**2**	32 (27-37)	30 (25-36)
	3	29 (22-35)	14 (8-19)
Kappa measure of agreement (95% CI)		0.39 (0.25-0.53)	0.40 (0.28-0.51)

* values represent mean percentage (95% CI)

Use of free text

The use of free text is shown in Table III. Free text is divided into free text that could have been entered in a structured way in OpenSDE (incorrect use of free text), and free text that could not have been entered in OpenSDE, because the tree did not contain the nodes to express the findings involved (correct use of free text). Each transcribed record contained on average 18 free text entries, while a mean of 212 findings were recorded. 27.7% of the free text entries were incorrect.

Table 3. Use of free text*

Free text that could not have been entered in a structured manner in OpenSDE (i.e. *correct* use)	13 (72,3)
Free text that could have been entered in a structured manner in OpenSDE (i.e. *incorrect* use)	5 (27,7)
Total*	18 (100)

* numbers represent average number of free text entries per case (percentages)

Users' experiences with OpenSDE and data entry time

Four major topics were questioned: content, layout, system and general response. All participants returned the questionnaire. The content of patient history as well as physical examination was found to be complete or very complete. Arrangement, legibility of characters and navigation through the tree were all judged as good or very good. The system used during the research period

was found to be reliable, easy to learn and use, and the participating physicians shared the opinion that data entry could be performed fast enough.

Data entry time was recorded for both the electronic and transcribed record, and was corrected for number of entered findings. The mean data entry time of the first four electronic records was 24:22 minutes (range 11:17–50:54). The second four records took on average 21:50 minutes (range 08:57-54:34).

The mean data entry time for the first eight transcribed records was 26:42 minutes (range 10:32-50:50min). The last eight transcribed records took on average 17:46 minutes per case (range 09:43-30:29min). The average duration of documenting patient data in the paper-based record was approximately 9 minutes.

Discussion

The first objective of this study was to compare current documentation in the control record with electronic documentation of the same information in OpenSDE. The electronic record contained more patient information than the control record. In the electronic record, physical examination contained more additional information than patient history. In patient history, however, more information was missing than in physical examination. This could be explained to a great extent by the nature of information that is recorded in patient history and physical examination. Patient history contains the patient's narrative whereas physical examination generally consists of objective measures and observations, registered in a systematic way. Structured data entry might therefore be more suitable for documentation of physical examination (i.e. objective data) than for patient history (i.e. subjective data). Walsh states that, as narratives are essential to a patients' episode of illness, computers should enable clinicians to capture these narratives easily (12). For efficient capture of physician gathered information, one could consider using a less detailed tree-structure for recording patient history than for physical examination in OpenSDE (13).

The presence of 31% additional information (in the electronic versus the control record) is probably due to fact that primary physicians made the electronic record immediately after having seen the patient. The nature of structured data entry might be inviting to record more detail (9). The tree structure seems to serve as a reminder about available descriptive options and the entry forms provide further detail. For example, when the remark "abdomen no abnormalities" or "heart sounds normal" was made in the paper record, this was often further specified in the electronic

record (e.g. instead of "heart sounds normal", "normal rhythm, 1st heart sound normal, 2nd heart sound normal, cardiac murmur absent" was recorded). These findings may be of higher 'quality', but also of less importance. Furthermore, the experimental setting itself, and the fact that physicians were not distracted during documentation in the EMR, could have promoted more detailed data entry, leading to information bias.

The 25% of the patient data that was found only in the control record mostly involved concepts requiring relatively much navigation to enter in OpenSDE, but were easy to document on paper. For example, medical history was often present in detail in the paper record, whereas in the electronic record less detailed free-text was often used, although structured options for description of medical history were readily available.

The second objective was to assess the uniformity of reporting in the transcribed records. The inter-rater agreement statistic kappa was 0.39 (CI 0.25-0.53) for patient history and 0.40 (CI 0.28-0.51) for physical examination respectively. Although this indicates a fair to moderate strength of agreement, it also points out that recording data using structured data entry as currently implemented, even in an artificial environment, does not necessarily improve uniformity in coverage.

In patient history more information was missing in two or three transcribed records of one case than in physical examination. Most of the missing information concerned detailed, clinically less relevant information (e.g. number of days attending nursery, occupation of the parents). This detailed information required relatively much navigation and was therefore difficult to find, and document in OpenSDE. The information absent in the transcribed records could have been forgotten or neglected. An explanation could also be that free text in the paper records was found illegible by the participants, resulting in less complete documentation in the transcribed records. Fifty-three out of the ninety-one transcribed records contained on average two findings (1.5% of 135 findings) conflicting with the control record. It is unclear whether these discrepancies were due to misinterpretation, or erroneous transcription.

Twenty-eight percent of the free text entries could have been entered in a structured way. Mostly, these entries concerned the use of free text on a level, high in the tree, where participants were able to document their findings quick and easy, at the cost of structured entry of the same information. For example, one physician documented the presence, duration and setting of wheezing as free text, whereas the same information could have been entered in a structured manner (Figure 2). Seventy-two percent of the free text entries were correct, i.e. no structured options

were available. Mostly, these entries concerned narrative information as, for example, a description of eating habits in a child with constipation.

Structured data entry applications have already proven to be valuable in concise subject areas like sonography, radiology and endoscopy (14-18). Rosenbloom reports that documents generated with SDE contained 64% more concepts compared to documents generated with a standard dictation/transcription model (18). Kuhn et al. reported that for the description of a technical examination (i.e. upper abdominal sonography) there is evidence of superiority of a structured approach over free-text dictation (17). In broad specialties, like internal medicine or general pediatrics, the drawbacks of capturing a patients' narrative in coded data may not readily be outweighed by eventual advantages. As structured data entry by nature limits individual expression, free text should not be fully replaced by SDE (3, 13, 19).

A learning effect could be identified in data entry time, using OpenSDE. Physician data entry has proven to be a major barrier in implementation of EMRs (13, 19, 20). To be successful in clinical practice, the time spent on documenting patient data in an EMR must therefore be minimized. However, structured data entry may require extra effort (4). Initially, data entry-time was increased due to unacquaintedness with OpenSDE. As the average duration of documenting patient data in the paper record was 9 minutes, it was still faster than OpenSDE. However, the latter contained on average more information, and not all functions in OpenSDE to facilitate data entry were used in this study (e.g. templates, customized entry forms).

For EMRs to be accepted in clinical practice they must meet the demands of their user: the physician (4-6). The results of the Questionnaire on User Interface Satisfaction revealed a positive attitude towards the use of OpenSDE. All participating physicians shared the opinion that data entry could be performed fast enough. The general response was that OpenSDE could be very useful in clinical practice.

Currently OpenSDE functionality is being implemented in our Hospital Information System, following a pilot in the pediatric emergency department. Explicit information was extracted from OpenSDE to generate reports.

Conclusion

The electronic records contained approximately 65% of all patient information that was present in the control record. Physical examination was found to be more complete and contained

more additional information than patient history, indicating that structured data entry is more suitable for documentation of objective data. The participating physicians had a positive attitude towards the use of an SDE-application, a conditio sine qua non for successful implementation.

Acknowledgement

We gratefully acknowledge M. de Wilde for technical support, and C. de Bruin, MD, C.S. van het Zandt, MD, M. Twilt, MD, A.L.M. Boehmer, MD, and T. Hendriks, MD for their participation in this feasibility study.

The generic OpenSDE-application, as well as the tree structure for general pediatrics (Dutch language) is available at the OpenSDE website: http://www2.eur.nl/fgg/mi/OpenSDE/

References

1. Dick RS, Steen EB. The Computer-Based Patient Record: An Essential Technology for Health Care. Revised edition. *National Academy Press*, Washington DC 1997.
2. Adams WG, Mann AM, Bauchner H. Use of an electronic medical record improves the quality of urban pediatric primary care. *Pediatrics*. 2003;111(3):626-632.
3. van Ginneken AM. The computerized patient record: balancing effort and benefit. *Int J Med Inf.* 2002;65(2):97-119.
4. Powsner SM, Wyatt JC, Wright P. Opportunities for and challenges of computerisation. *The Lancet*. 1998;352(9140):1617-1622.
5. Dansky KH GL, Vasey JJ, Barsukiewicz CK. Electronic medical records: are physicians ready? *J Healthc Manag*. 1999;44(6):440-54.
6. Sittig DF, Kuperman GJ, Fiskio J. Evaluating physician satisfaction regarding user interactions with an electronic medical record system. *Proc AMIA Symp*. 1999:400-4.
7. Bleeker SE, Derksen-Lubsen G, van Ginneken AM, van der Lei J, Moll HA. A computer-based patient record with structured data entry for a comprehensive patient history and physical examination in pediatrics. *Unpublished data.* 2004.
8. Los RK, van Ginneken AM, de Wilde M, van der Lei J. OpenSDE: Row Modeling Applied to Generic Structured Data Entry. *J Am Med Inform Assoc*. 2004;11(2):162-165.
9. Wyatt J. Quantitative evaluation of clinical software, exemplified by decision support systems. *Int J Med Inf*. 1997;47(3):165-173.
10. Stausberg J, Koch D, Ingenerf J, Betzler M. Comparing Paper-based with Electronic Patient Records: Lessons Learned during a Study on Diagnosis and Procedure Codes. *J Am Med Inform Assoc.* 2003;10(5):470-477.
11. Chin J, Diehl V, Norman K. Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings ACM CHI'92 Conference*. 1992.
12. Walsh SH. The clinician's perspective on electronic health records and how they can affect patient care. *BMJ*. 2004;328(7449):1184-1187.
13. McDonald CJ. The Barriers to Electronic Medical Record Systems and How to Overcome Them. *J Am Med Inform Assoc.* 1997;4(3):213-221.
14. Gouveia-Oliveira A, Salgado NC, Azevedo AP, et al. A unified approach to the design of clinical reporting systems. *Methods Inf Med*. 1994;33(5):479-87.
15. Bell DS, Greenes RA. Evaluation of UltraSTAR: performance of a collaborative structured data entry system. *Proc Annu Symp Comput Appl Med Care*. 1994:216-22.
16. van der Meijden MJ, Tange HJ, Boiten J, Troost J, Hasman A. An experimental electronic patient record for stroke patients. Part 2: System description. *Int J Med Inf.* 2000;58-59:127-140.
17. Kuhn K, Gaus W, Wechsler JG, et al. Structured reporting of medical findings: evaluation of a system in gastroenterology. *Methods Inf Med.* 1992;31(4):268-74.
18. Rosenbloom ST, Kiepek W, Belletti J, et al. Generating complex clinical documents using structured entry and reporting. *Medinfo*. 2004;2004:683-7.
19. Gilbert JA. Physician data entry: providing options is essential. *Health Data Manag*. 1998;6(9):84-6, 88, 90-2.
20. Kaplan B. Reducing barriers to physician data entry for computer-based patient records. *Top Health Inf Manage*. 1994;15(1):24-34.

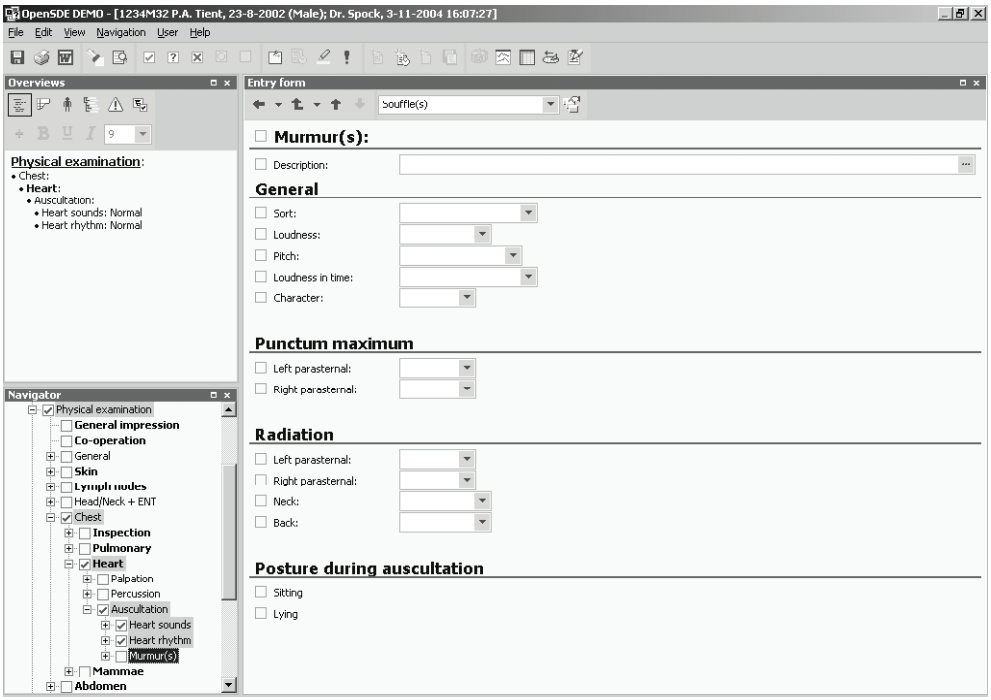


Figure 1.

Screen capture OpenSDE:

The left bottom side of the interface shows the tree of medical concepts and descriptive options. The right side of the screen displays an entry form representing the descriptive options for the selected concept in the tree ('cardiac murmur'). The left upper side presents an overview of entered data.

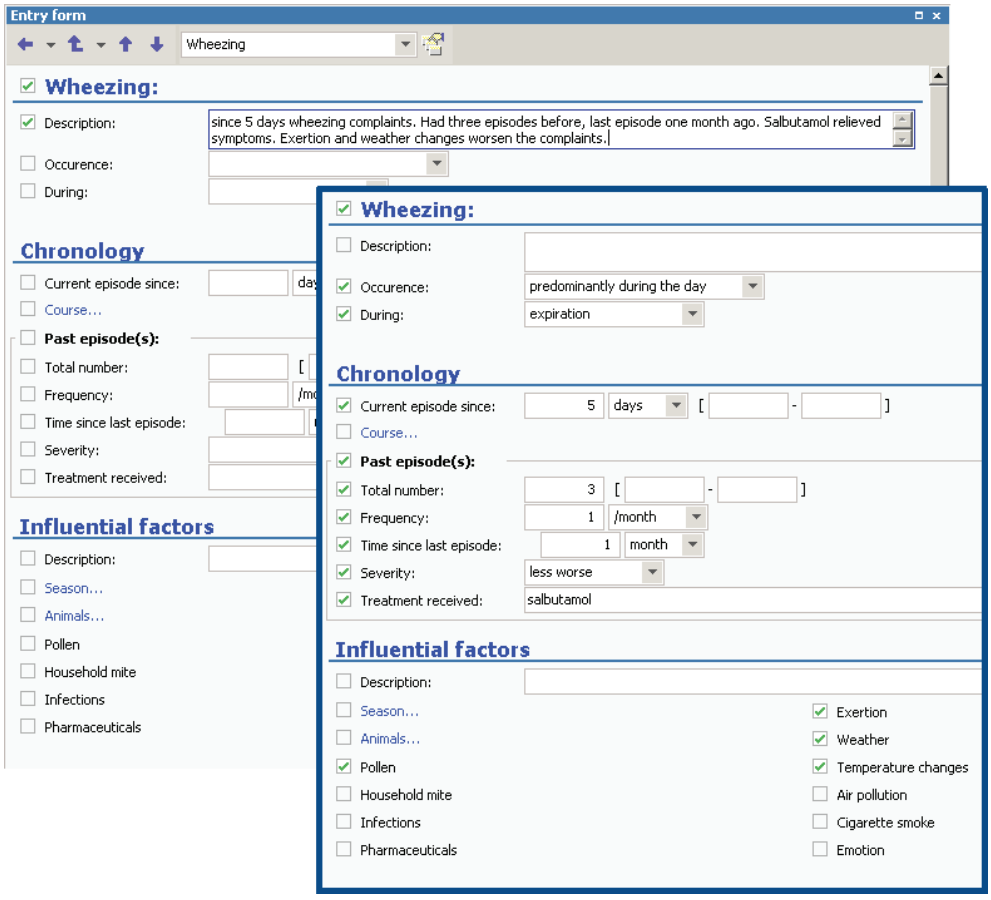


Figure 2.

Screen capture of entry form for 'wheezing' :

Example of free text documentation (upper left) and structured documentation (lower right) of the same information. Both contain overlapping and non-overlapping information. Free text that is also present in the structured information field, represents incorrect use.

Chapter 3

Are Structured Data Structured Identically? Investigating the uniformity of pediatric patient data recorded using OpenSDE

Renske K Los, Jolt Roukema, Astrid M van Ginneken, Marcel de Wilde, Johan van der Lei

Methods of Information in Medicine 2005; 44:631-638

Abstract

Objective: OpenSDE is an application that supports structured recording of narrative patient data to enable use of the data in both clinical practice and clinical research. Reliability and accuracy of collected data are essential for subsequent data use. In this study we analyze the uniformity of data entered with OpenSDE. Our objective is to obtain insight into the consensus and differences of recorded data.

Methods: Three pediatricians transcribed 20 paper patient records using OpenSDE. The transcribed records were compared and all recorded findings were classified into one of six categories of difference.

Results: Of all findings 22% were recorded identically; 17% of the findings were recorded differently (predominantly as free text); 61% was omitted, inferred, or in conflict with the paper record.

Conclusion: The results of this study show that recording patient data using structured data entry does not necessarily lead to uniformly structured data.

Introduction

Many potential advantages of electronic patient records (EPRs), such as availability of patient data for clinical research, decision support, or quality assessment(1, 2), require data to be represented in a structured manner(3, 4). Structured Data Entry (SDE) is a method by which clinicians record patient data directly in a structured format. SDE involves predefined fields for data entry. Advantages of this approach are: data are structured at the source, without requiring intervention or correction rounds; data are more uniform; predefined entry fields may predispose users to record data in more detail; and SDE offers the possibility of enhancing the quality of data(5).

SDE remains challenging to apply for medical narratives, as data vary per domain, per patient, and over time(6-8). The medical narrative comprises the medical history, physical examination, progress notes, and reports on additional tests and interventions(9). The narrative is a combination of patient narrated and clinician-observed data.

Our objective is to support structured recording of narrative data (in multiple medical domains) to enable use of the data in both clinical practice and clinical research. Therefore, we developed OpenSDE(10) as an application that offers structured data entry in a variety of settings. OpenSDE supports data entry using customizable entry forms based on domain-specific trees. OpenSDE is available in open source(11).

Although OpenSDE supports structured data entry, suggesting that data are structured uniformly, the actual concordance in data representation has not yet been explored. Reliability and accuracy of collected data are pivotal if data will be collected over long periods of time and by different users(12, 13).Therefore, in collaboration with our hospital's pediatric department, we analyzed the uniformity of recorded data when OpenSDE is used to transcribe data from the same data source. Of interest in this qualitative analysis is whether recording data using OpenSDE by definition leads to uniformly structured data. Obtaining insight into the consensus and differences in data recorded with OpenSDE is particularly important when retrieving routinely collected data for clinical research purposes (14). Uniformity in data entry facilitates data extraction and lookup: if the same data are recorded in different manners by different clinicians the chance of finding the data (in a particular place in the record) becomes smaller. If, for example, one clinician records a penicillin allergy in a structured manner and another clinician records this as free text comment in patient history, both places must be checked to see if a patient is allergic to penicillin. This problem becomes even larger when data can be recorded as free text anywhere and one does not

know in advance where to expect particular data. Data can easily be overlooked and the chance for duplicate data recording also increases. Look up may take more time and increase the workload on clinicians which can lead to a decrease in the quality of patient care and a lower success rate of the implementation of OpenSDE(15).

The purpose of this study is to provide qualitative insight into how data are recorded. It is important to understand how to format information to make data easier to find and clearer to interpret (16). We need to understand if the current format that we offer clinicians to record data leads to uniformity. If OpenSDE invites users to record the same data in exactly the same manner, retrieval and look up will be more predictable and easier to do for the user. If OpenSDE does not lead to uniform data representation we need to investigate what differences occur and how these differences can be minimized.

Materials

A. OpenSDE

OpenSDE is an application for structured recording of narrative sections of the patient record. The principle of OpenSDE is that clinicians can traverse a tree of predefined medical concepts and select those concepts that correspond with the relevant medical observations. The content of such a tree is domain specific and we refer to the tree of medical concepts as a domain model. In this tree structure, the nodes represent medical concepts and the path from the top of the tree to a particular node represents the context of a node(10).

Clinicians can select a node in the tree, and the application will display a form associated with this node alongside the tree, as shown in Figure 1. Each form presents the selected concept and the corresponding descriptors (branching nodes) of the concept(17). For the concepts presented on the entry forms, users may indicate whether or not the concept applies (present, absent, or unknown) or, when relevant, record a specific value (numerical, temporal or free text). Symptoms can be described more than once in the context of progression over time, different circumstances, or multiple occurrences. OpenSDE also supports the use of free text for particular details not covered by the content of the domain model. Users can create custom entry forms (using an integrated form editor) to suit their individual data entry preferences(10, 18).

The pediatric domain model used in this study was created by modelers with a background in pediatrics (second author) and medical informatics (third author). Prior to this study, experienced OpenSDE users recorded data of over 100 pediatric paper records in OpenSDE to evaluate the ordering and coverage of the pediatric domain model. The model was then altered to improve both ordering and coverage, as well as to facilitate data entry(19).

Methods

A. Data entry from a common data source

At our pediatric outpatient department we recruited three pediatricians. Prior to this study, the experience of these pediatricians consisted of a standardized course on the use of OpenSDE in general pediatrics and documentation of ten first-contact patients in OpenSDE.

We randomly selected 20 handwritten paper patient records created for first-contact patients at the pediatric outpatient department. These records belonged to patients that were not under care of any one of the three pediatricians involved in this study. We chose first-contact patients as intake and physical examination data for these patients are fairly standardized. Although the patient data are recorded as free text in the paper records, the data are written on semi-structured forms that contain headings such as 'family medical history', 'birth history', 'allergies', and 'neurological examination' at which the corresponding medical findings can be recorded. Due to this 'structure' the paper records were comparable in format, degree of detail, and in amount of data content. We expected that data entry by three clinicians would provide good insight into the nature of differences in data representation.

The three pediatricians transcribed the 20 paper records in OpenSDE, creating a data set of 60 transcribed records. The pediatricians were informed about the goal of the study, and knew that the transcribed records would be analyzed.

B. Consensus and differences

Our main interest in this study is the consensus and differences in the representation of structured patient data. Therefore, we conducted detailed analyses of the transcribed records to identify the types of differences in the transcribed records. We identified six categories of consensus and differences. To classify the findings into one of the six categories we developed the algorithm described below and presented in Figure 2.

Per patient we created a list of all findings recorded in OpenSDE. For every finding we analyzed how it was represented by each of the three pediatricians. If the finding was represented in exactly the same manner in all three transcribed records, the finding was classified as *identical*. If there was a difference in at least one of the transcribed records, we searched through that entire transcribed record to establish whether the finding was recorded elsewhere. If the finding was represented in a structured manner elsewhere, the finding was classified as *structured differently*.

If the finding was recorded as free text, at the same place or elsewhere in the tree, the finding was classified as *free text*.

For those cases where the finding was not represented in all three transcribed records (either identically, as free text, or structured differently) we consulted the paper record. If the finding was present in the paper record, we classified the finding as *omitted*. If, however, the description of the finding in the paper record conflicted with the description of the finding in the transcribed records, we classified the finding as *conflicting*. A last possibility is that (one of) the transcribed records contained a finding not present in the paper record. In this scenario the finding was classified as *inferred*. The classification process was repeated for all findings.

The findings were subsequently classified as normal findings (e.g. 'no cardiac murmur') or abnormal findings (e.g. 'constipation'), and were split into patient history and physical examination findings.

Results

The findings recorded in the transcribed records were divided into patient history and physical examination findings and then subdivided into normal or abnormal findings. These (sub) divisions are shown in Table 1: the rows hold the type of finding (normal or abnormal) and the columns represent the part of the record (patient history or physical examination). A total of 1764 findings were recorded for all 20 patients. Of these findings, 495 are normal patient history findings and 867 normal physical examination findings (totaling 1362 normal findings). In total, 77.2 % of all findings are normal findings, which corresponds to a mean of 68 normal findings per patient (range 23-117).

Table 1: Normal and abnormal findings in the patient record

The results presented in this table represent the number of findings per part of the patient record (patient history or physical examination) and per type or finding (normal or abnormal). The percentage corresponds to the percentage of the total number of findings.

Part of patient record	**Patient**	**Physical**	Entire Record
Type of finding	**History**	**Examination**	
Normal Findings	**495** (28.1%)	**867** (49.1%)	1362 **(77.2%)**
Mean per transcribed record	24.8	43.4	**68.1**
Range	3-59	20-61	**23-117**
Abnormal Findings	**351** (19.9%)	**51** (2.9%)	402 **(22.8%)**
Mean per transcribed record	17.6	2.6	**20.1**
Range	5-38	0-6	**5-43**
All Findings	**846** (48%)	**918** (52%)	1764 **(100%)**
Mean per transcribed record	42.3	45.9	**88.2**
Range	**8-88**	**20-63**	**28-151**

In Table 2 we present the findings as classified per category of consensus/difference. All findings could all be classified into one of the six categories of consensus/difference according to the algorithm. The first row of the table shows that we encountered 90 normal and 79 abnormal patient history findings which were recorded identically for all three patients. For the physical examination, we counted 198 normal and 21 abnormal findings that were transcribed identically by all three pediatricians. Of all findings 22% (or 388 findings) were recorded identically by all three pediatricians. In total, 4.9% of all findings were structured differently, and in 12.2% of all findings

one or two pediatricians recorded the findings as free text. Almost one third (31.1%) of all findings was inferred, and over one quarter (26.7%) of the findings was omitted by one or two pediatricians. A total of 55 findings (3.1%) were conflicting with the paper record.

Table 2: Consensus and differences in representation of findings

In this table the findings have been ordered per category of consensus/difference. Per category the findings are subdivided into patient history and physical examination, and normal and abnormal findings. The percentage behind the numbers corresponds to the percentage of the total number of findings.

Finding	**Patient History**		**Physical Examination**		***ALL FINDINGS***
Category	Normal	Abnormal	Normal	Abnormal	
Identical	90 (5.1%)	79 (4.5%)	198 (11.2%)	21 (1.2%)	**388** (22%)
Structured differently	28 (1.6%)	23 (1.3%)	32 (1.8%)	4 (0.2%)	**87** (4.9%)
Free text	68 (3.9%)	117 (6.6%)	26 (1.5%)	4 (0.2%)	**215** (12.2%)
Inferred	98 (5.6%)	30 (1.7%)	411 (23.3%)	9 (0.5%)	**548** (31.1%)
Omitted	188 (10.7%)	90 (5.1%)	184 (10.4%)	9 (0.5%)	**471** (26.7%)
Conflicting	23 (1.3%)	12 (0.7%)	16 (0.9%)	4 (0.2%)	**55** (3.1%)
ALL FINDINGS	***495*** *(28.1%)*	***351*** *(19.9%)*	***867*** *(49.1%)*	***51*** *(2.9%)*	***1764*** *(100%)*

Discussion

Structured data entry offers the possibility to improve the quality of data(5) and standardize data collection(20). In this study we investigated the uniformity of recorded data when OpenSDE is used to transcribe data from a common data source. We analyzed 60 transcribed records, which in total covered 1764 findings. Our results show that only 22% of all findings were recorded identically by all three clinicians and in more than three quarters of the findings there was difference in data representation or data content.

Evaluation of data quality in medical records is a topic of ongoing interest in medical informatics. Evaluation methods and measurement means are not standardized and different studies focus on different aspects as different stakeholders pose different requirements on data quality (21) (14).

Evaluating the quality of the data recorded in OpenSDE is associated with one particular difficulty. Data quality, especially completeness and accuracy, can only be measured as a function of the question that the data set should answer (22). Winthereik concludes that the goal should not be to produce data, which are accurate in and by themselves, but to produce data, which are pertinent to specific questions (23). However, the idea behind OpenSDE is that data are recorded during routine care to be available for patient care and clinical research. There is thus no clear question that can be used to evaluate whether a routine data set meets the desired quality. We, therefore, chose to investigate whether the users at least record the same data in the same manner. Although it is difficult to identify criteria against which quality should be judged (24), we feel that uniformity is an important aspect of data quality as it has effects on the ease of data look up and retrieval, which are important incentives for recording data in a structured manner.

The results in Table 2 show that only 22% of the findings were recorded identically by all three clinicians. This number will become progressively lower as the number of clinicians increases. Obviously, when looking at identical recording on a pair wise basis, the percentage will be much higher. However, as long as one or more clinicians have recorded the same piece of medical information differently, a researcher extracting data (or a clinician treating a patient) will have to be aware of it. Whether a patient's penicillin allergy is recorded in a structured manner as allergy or recorded as free text somewhere in the record, it is important that this information is not overlooked. Hence, to obtain insight in the challenges of data extraction, emphasis is on the nature of differences when people record the same information. We will discuss these differences by first focusing on the findings that are represented differently, but where data content is the same

(categories: "structured differently" and "free text"). We will subsequently analyze those findings where there is discordance in data content (categories: "inferred", "omitted", and "conflict").

A. Different representation

In OpenSDE a medical finding can be represented differently either in a structured manner or as free text. In 4.9% of all findings, clinicians recorded the same finding differently in a structured manner. This can occur, for example, if a patient's mother has a heart condition. The heart condition can be recorded as part of the family history or as reason for giving birth to the patient in the hospital instead of at home (which is usual in the Netherlands). Nevertheless, it is still the same heart condition which, depending on the context, can be entered at different places in the tree. The heart condition is relevant both for the family history and the birth conditions of the child. Hence, the same information can be relevant in more than one context. In such scenarios the finding is ideally represented at one place in the domain model with only a reference to this description at the other nodes where the information is relevant.

A finding can also be represented as free text at various nodes in the tree, as was the case with 12.2% of all findings. We encountered data for a patient's father who suffered from hay fever *and* who was allergic to particular types of food. Instead of recording that the father has food intolerance, the clinician chose to record all of the father's allergies at the hay fever node as free text. Such use of free text makes the search for data more complex and less reliable, as data can be recorded anywhere as free text.

Although findings categorized as "structured differently" and "free text" are transcribed in OpenSDE, i.e., the findings are recorded in the patient's electronic record, they are not transcribed by all clinicians in the same manner. The findings are thus not recorded uniformly. Hence, in patient care clinicians may overlook data when they only look at one particular data item in the tree and do not consult all data about the patient. For bulk retrieval, the lack of uniformity can also have consequences. When data are not recorded uniformly, searching for a particular finding requires searching the entire tree of recorded data to ensure that the finding is not overlooked. As it is unpredictable where and how (e.g. abbreviations, codes, spelling or typing errors) findings can be recorded, look up and retrieval can hardly be automated. The potential benefits of SDE are then barely achieved and one should question whether the benefits of structuring the data still outweigh the efforts.

In their study, Pringle et al.(25) conclude that subjective data are less consistently recorded than objective data. Peat et al.(26) investigated how reliable structured clinical history-taking is and conclude that subjective information leads to higher inter- and intraobserver variability. Our results show that physical examination findings are recorded identically more often than patient history findings. Furthermore, our results show that clinicians use the free text possibility more often for patient history than for physical examination findings. Although data are suitable for more purposes than free text, patient history often requires more use of free text to cover patient-specific detail. Also, a few lines with the essence of the patient's story provide more overview at a glance than when this story is 'scattered' under the various nodes of the tree. Therefore, free text should not be fully replaced with SDE, but rather be combined with SDE(2, 27).

B. Discordance in data content

The largest of the three categories that cover discordance in data content, is the inferred findings category which constitutes 31.1% of all findings. The majority of these inferred findings constitutes normal physical examination findings; of the 548 inferred findings, 39 (7.1%) were abnormal, meaning that such abnormalities were not recorded in the patient's paper record. Inferred abnormal findings include pain during defecation and a father with hay fever. Regarding the latter example, the paper source revealed that the clinician misread its content: the mother's and father's histories were written directly below each other, where the mother suffered from hay fever and the father from other allergies. Hence, inferred findings, as found in step six of the algorithm, may include such misreadings of the paper record.

The inferred findings are predominantly interpretations of routine expressions and interpretations based on data that are written in the paper record. For a patient suffering from constipation accompanied by blood loss, one clinician recorded that the blood was clear red, whereas the paper record made no reference to the color of the blood. Such inferred findings, of which this is just one example, lead to believe that if there are no data recorded for an observation, then it is probably normal.

Transcribing data includes interpreting routine expressions such as 'vesicular breath sounds.' When routine expressions are not identically represented in the domain model, some clinicians opt to record such expressions as free text, whereas others opt to "translate" the routine expression into those concepts that approach the meaning of the routine expression (e.g. normal breath

sounds). Such translations involve interpreting the routine expression, which can lead to different representations and derivations. This form of discordance can be reduced by adhering to terms frequently used in a particular setting when constructing the domain model.

The inferred findings, especially the inferred normal findings, which constitute the majority of the inferred findings, indicate that OpenSDE induces an effect similar to the checklist effect(28): clinicians are triggered by the available entry fields to record more data.

Even though OpenSDE triggers clinicians to record particular data, our results also show that 26.7% of all findings were omitted by one or more clinicians. Although the majority of omitted findings (79%) were normal findings, an omission of 26.7% of all findings is dissatisfying when the purpose of SDE is to improve quality, completeness, and consistency of data. Findings are omitted when clinicians overlook findings in the paper record, ignore "irrelevant" findings, or when recording the finding in OpenSDE is not straightforward.

In their study assessing the completeness and accuracy of computer medical records, Pringle et al.(25) suggest that "it was clear that practices were selecting areas that they considered important to record on their computer systems". In line with these results, our results suggest that clinicians are more inclined to omit normal findings than abnormal findings, as normal findings are generally of lesser importance when examining the patient than abnormal findings. Nevertheless, for the purpose of improving data quality, recording observed normal findings is also important.

The last category that we analyzed involved the conflicting findings. Just over 3% of all recorded findings were in direct conflict with the data in the paper record. Conflicting findings include recording previously used medication as current medication, recording a cardiac murmur when the patient does not have a cardiac murmur, and recording incorrect numerical values. In real life, when clinicians directly record findings using OpenSDE instead of transcribing findings from paper records, the percentage of conflicting data will probably be lower; errors in judgment or typographic errors will still be made but transcription errors due to, for example, misreading of the paper record, will no longer apply. SDE may, however, introduce a new category of errors, such as erroneous selection from checklists.

C. Limitations of the study

Studies such as ours have three particular limitations. The first limitation is that we only analyzed the uniformity in the *transcribed* records. We did not analyze whether the transcription

of findings in the paper record using OpenSDE was complete. Such a study would provide insight into the effect of OpenSDE on promoting completeness(29). However, this was not our goal. Our interest in this particular study is *how* people represent patient data using OpenSDE when they have the same source of data.

The source of data is a second limitation. Transcribing findings from a paper record involves interpreting the recorded interpretations of a colleague. In an ideal situation all clinicians are confronted with the actual patient instead of only with a paper record. However, this is undesirable for patient care, as one cannot ask of patients (especially children) to tell their story to several clinicians. Besides that, the clinicians will approach the patients differently and ask different questions to which the patient may give answers that vary, for example, in level of detail. This introduces bias into the study as the source of data is not identical for all clinicians. In addition, the patient may also become biased by the questions of a previous clinician when elaborating about his complaints to the next clinician. We should also keep in mind our research question: does OpenSDE invite users to record data from the same source in a uniform structure? To answer this question it is essential that all involved clinicians consult the same data source. The results of our analysis give us insight into *how* people can and will represent data and what the differences in representation are.

The third limitation is the number of clinicians included in the study. Even though there was not one clinician that outperformed the others in terms of the number of structured findings per record, an increased number of clinicians may reveal even more differences in data representation. However, we feel that three clinicians do provide enough insight into how data can be recorded, especially as we encountered situations which we had not considered. There was one clinician, for example, that had four different manners of mapping the same free text term from paper to OpenSDE. Nevertheless, these results do give us insight into how data are represented in a routine clinical situation and which potential pitfalls this creates for data extraction.

Conclusion

Structured data entry is intended to improve the quality and consistency of data by obtaining the data directly from clinicians in a structured format. To analyze the uniformity of data recorded with OpenSDE, we performed a study in which three pediatricians used OpenSDE to transcribe 20 handwritten paper patient records. Our results show that data recorded using SDE are not necessarily represented in the same manner and nearly two-thirds of the recorded data are discordant (i.e., inferred, omitted, or conflicting with data in the paper record).

In line with other studies our results indicate that even though information is more accessible it is not necessarily creditable(30), directly usable(31), or structured uniformly. As a result, data collected with OpenSDE cannot unconditionally be used for subsequent purposes such as clinical research. Mikkelsen and Aasly take the claim even further and say that inconsistencies in information elements used to characterize clinical information represents a potential threat to the safety of using EPRs as source of clinical information (14). We did not go as far as to evaluate the actual clinical consequences of the differences in data representation but based on our differences and the conclusions of other work, this also requires attention. Studies such as ours increase insight into retrieval pitfalls, independent of the system being used. Standardizing data entry and multiple search strategies are certainly necessary before aggregated data can be relied upon(25).

Based on the results of this study, we are currently addressing the following two aspects. Firstly, we are focusing on increasing the uniformity in data entry by limiting the number of ways in which the same information can be recorded, without limiting the level of detail in which data can be recorded. Secondly, we are investigating multiple search strategies for data retrieval, to increase the probability that all relevant data are actually retrieved.

This study has pointed out those aspects in the design of OpenSDE where problems arise both during data entry as well as during use of the data. Insight into the difference in data recording is useful because it helps us improve the design of OpenSDE for data entry, with the aim of improving the data set and enhancing the potential clinical use of the database(31). The results of this study show that recording data using structured data entry does not necessarily lead to uniformly structured data.

In general what can be learned from this study is that for data lookup and retrieval one must be aware of all possible ways in which an item of information may have been recorded.

Acknowledgements

We would like to thank the pediatricians involved in this study for the time and effort spent transcribing data.

The work presented in this paper is funded by a grant from the Netherlands Organization for Health Research and Development (ZonMW).

References

1. Dick RS, Steen EB, Detmer DE, eds. The computer-based patient record: an essential technology for health care. Revised Edition ed. Washington: National Academy Press; 1997.
2. van Ginneken AM. The computerized patient record: balancing effort and benefit. Int J Med Inf. 2002 Jun;65(2):97-119.
3. Powsner SM, Wyatt JC, Wright P. Opportunities for and challenges of computerisation. Lancet. 1998 Nov 14;352(9140):1617-22.
4. Brown PJ, Sönksen P. Evaluation of the quality of information retrieval of clinical findings from a computerized patient database using a semantic terminological model. J Am Med Inform Assoc. 2000 Jul-Aug;7(4):392-403.
5. Moorman PW, van Ginneken AM, van der Lei J, van Bemmel JH. A model for structured data entry based on explicit descriptional knowledge. Methods Inf Med. 1994 Dec;33(5):454-63.
6. Tange H. How to approach the structuring of the medical record? Towards a model for flexible access to free text medical data. Int J Biomed Comput. 1996 Jul;42(1-2):27-34.
7. Walsh SH. The clinician's perspective on electronic health records and how they can affect patient care. BMJ. 2004 May 15;328(7449):1184-7.
8. Klar R. Selected impressions on the beginning of the electronic medical record and patient information. Methods Inf Med. 2004;43(5):537-42.

9. Tange HJ, Hasman A, de Vries Robbe PF, Schouten HC. Medical narratives in electronic medical records. Int J Med Inf. 1997 Aug;46(1):7-29.
10. Los RK, van Ginneken AM, de Wilde M, van der Lei J. OpenSDE: row modeling applied to generic structured data entry. J Am Med Inform Assoc. 2004 Dec 7;11(2):162-5.
11. OpenSDE. OpenSDE (OSS). [cited March 31, 2005]; Available from: http://www2.eur.nl/fgg/mi/OpenSDE/
12. Beard CM, Yunginger JW, Reed CE, O'Connell EJ, Silverstein MD. Interobserver variability in medical record review: an epidemiological study of asthma. J Clin Epidemiol. 1992 Sep;45(9):1013-20.
13. Williams JG. Measuring the completeness and currency of codified clinical information. Methods Inf Med. 2003;42(4):482-8.
14. Mikkelsen G, Aasly J. Consequences of impaired data quality on information retrieval in electronic patient records. Int J Med Inform. 2005 Jun;74(5):387-94.
15. Ammenwerth E, Shaw NT. Bad health informatics can kill--is evaluation the answer? Methods Inf Med. 2005;44(1):1-3.
16. Wyatt J. Same information, different decisions: format counts. Bmj. 1999 5 june 1999;318:1501-2.
17. van Ginneken AM, Verkoijen MJ. A multi-disciplinary approach to a user interface for structured data entry. Medinfo. 2001;10(Pt 1):693-7.
18. Los RK, van Ginneken AM, van der Lei J. OpenSDE: A strategy for expressive and flexible structured data entry. Int J Med Inform. 2005 Jul;74(6):481-90.
19. Roukema J, van Ginneken AM, Moll HA. The use of structured data entry in the outpatient's clinic for paediatrics. Health Information Developments in the Netherlands. 2003;6:27-30.
20. Kahn CE, Jr., Huynh PN. Knowledge representation for platform-independent structured reporting. Proc AMIA Annu Fall Symp. 1996:478-82.
21. Stoop AP, Berg M. Integrating quantitative and qualitative methods in patient care information system evaluation: guidance for the organizational decision maker. Methods Inf Med. 2003;42(4):458-62.
22. Rubenfeld GD. Using computerized medical databases to measure and to improve the quality of intensive care. J Crit Care. 2004 Dec;19(4):248-56.
23. Winthereik BR. "We fill in our working understanding": on codes, classifications and the production of accurate data. Methods Inf Med. 2003;42(4):489-96.
24. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. Bmj. 2003 May 17;326(7398):1070.
25. Pringle M, Ward P, Chilvers C. Assessment of the completeness and accuracy of computer medical records in four practices committed to recording data on computer. Br J Gen Pract. 1995 Oct;45(399):537-41.
26. Peat G, Wood L, Wilkie R, Thomas E. How reliable is structured clinical history-taking in older adults with knee problems? Inter- and intraobserver variability of the KNE-SCI. J Clin Epidemiol. 2003 Nov;56(11):1030-7.
27. McDonald CJ. The barriers to electronic medical record systems and how to overcome them. J Am Med Inform Assoc. 1997 May-Jun;4(3):213-21.
28. Wyatt J. Quantitative evaluation of clinical software, exemplified by decision support systems. Int J Med Inf. 1997 Dec;47(3):165-73.
29. Roukema J, Los RK, Bleeker SE, van Ginneken AM, van der Lei J, Moll HA. Paper versus computer: feasibility of an electronic medical record in general pediatrics. Pediatrics. 2005;Accepted for publication.
30. Burnum JF. The misinformation era: the fall of the medical record. Ann Intern Med. 1989 Mar 15;110(6):482-4.
31. Wyatt JC, Wright P. Design should help use of patients' data. Lancet. 1998 Oct 24;352(9137):1375-8.

Figure 1. Screen capture of the OpenSDE data entry application. The top left of the screen shows an overview of the data recorded for the patient in the current session. The bottom left shows the domain model tree with medical concepts. On the right is the form on which data are entered. The form is associated with the selected node, in this case 'defecation'.

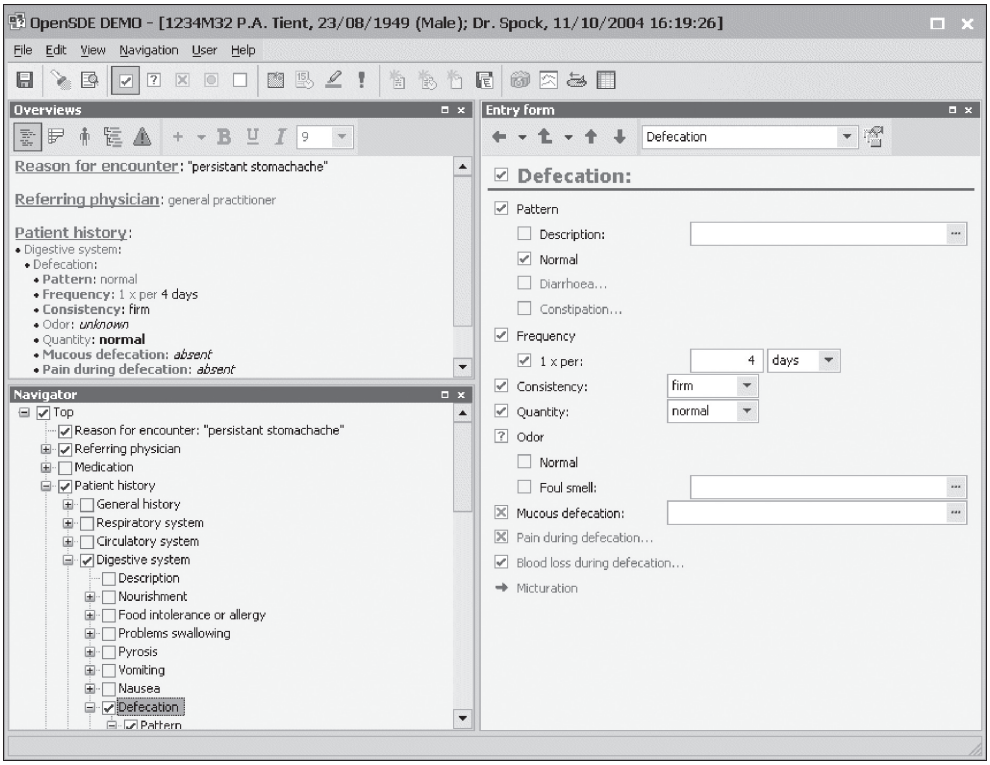
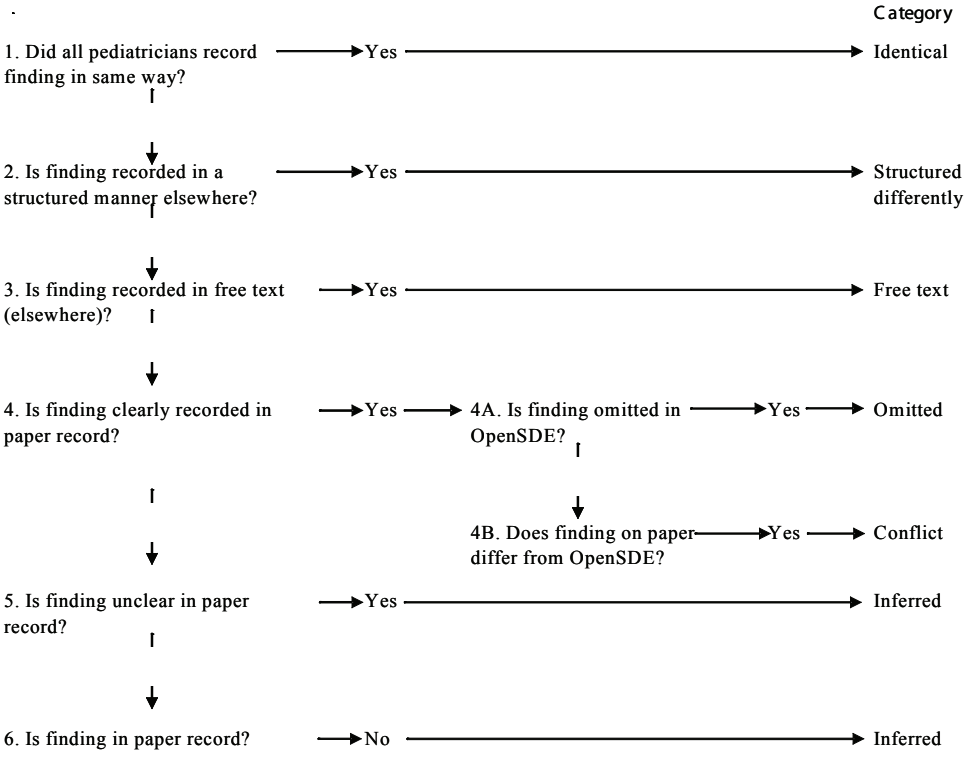


Figure 2. Algorithm used to categorize findings.



Chapter 4

Predicting pneumonia in children: a multivariable analysis

Jolt Roukema, Ewout W. Steyerberg, M.H. Lequin, Johan van der Lei, Henriette A. Moll

Abstract

Background Respiratory tract infections in children occur frequently. Early identification of children at high or low risk of pneumonia, using readily available characteristics from patient history and physical examination, could support the diagnostic and therapeutic strategy in children presenting with cough and fever. The aim of this study was: to develop different prediction models to identify children at risk for pneumonia based on clinical characteristics; to assess the added diagnostic value of laboratory test results (white blood cell count (WBC) and C-reactive protein (CRP)); and to illustrate the differences that occur due to patient selection and multiple imputation of missing predictor variables.

Methods From July 2003 until December 2005, data were collected of all consecutive patients, aged 1 month to 16 years, presenting with fever and cough at the emergency department of the Sophia Children's Hospital in Rotterdam, The Netherlands. Missing values for WBC (n=248) and CRP (n=259) were statistically imputed. First, a multivariable logistic regression model was developed using clinical patient characteristics. Subsequently, laboratory test results were added to the model. The discriminative ability of the models was quantified by the area under the receiver operating characteristic curve (AUC), and corrected for optimism by bootstrapping.

Results 504 patients with fever and cough were included. The mean age was 2.3 years, and 78 (15%) were diagnosed with pneumonia. Laboratory test results were available in 242 patients. Clinical characteristics predictive for pneumonia included age, duration of the fever, and tachypnea. A model based on clinical characteristics of the 242 patients with laboratory tests (model I) had an AUC of 0.65, which was significantly lower than the AUC of the model based on clinical characteristics of all 504 patients (model II, AUC 0.74). Adding WBC and CRP had significant additional diagnostic value in both models; the AUCs increased from 0.65 to 0.75 and from 0.74 to 0.80 in model I and II respectively.

Conclusion The risk of pneumonia can reasonably be predicted with a limited number of patient characteristics. Laboratory tests (WBC and CRP) add significantly to the diagnosis of radiographic pneumonia. Clinical prediction models may be used to guide the diagnostic process. The prediction model should then be presented as a clinical decision rule, and should be further validated.

Introduction

Respiratory tract infections in children occur frequently and cause significant morbidity. The annual incidence of pneumonia in North America and Europe per 1,000 children ranges from 30-45 in children less than 5 years old, 16-20 among children 5 to 9 years old, and 6-12 among older children and adolescents (1-5). In children suspected of pneumonia, the diagnostic work-up mostly includes laboratory tests (e.g. white blood cell count, c-reactive protein) and a chest radiograph (6). Early identification of children at high or low risk of pneumonia, using readily available characteristics from patient history and physical examination, can support the diagnostic strategy. Several attempts have been undertaken to diagnose or exclude pneumonia based on clinical characteristics (7-14).

A complicating factor in the development of clinical prediction models is missing values for covariates/predictor variables. It has been shown that the use of multiple imputation of missing values can reduce bias and increase efficiency (15, 16).

In this study, our aim was to assess predictors of pneumonia in children presenting with cough and fever in the hospital emergency department, considering readily available patient characteristics. Further, we aimed to evaluate the added diagnostic value of white blood cell count (WBC) and C-reactive protein (CRP) in the diagnosis of pneumonia. Since the indication for laboratory tests were left to the discretion of the attending physician, laboratory results were only available for a subgroup of patients. Therefore we developed different multivariable logistic regression models to illustrate the differences that occur due to patient selection and handling of missing laboratory test values.

Patients and methods

This study was part of a larger study on diagnostic management of febrile children, approved by the local Institutional Review Board. The Sophia Children's Hospital is an inner city university hospital, and the ED delivers general pediatric care to 4,500 children annually.

We prospectively enrolled all patients (1 month – 16 years old) presenting from July 2003 until December 2005 at the emergency department (ED), with cough and fever (i.e. body temperature ≥ 38.0ºC). Patients with an immune deficiency, multiple handicaps, or a preexisting pulmonary disease (e.g. broncho-pulmonary dysplasia, cystic fibrosis) were excluded.

Determinants

Potential diagnostic determinants were: age (years), temperature (degrees Celsius), duration of the fever episode (days), sick impression, presence of chest-wall retractions, and tachypnea (8-10, 14, 17-19). Respiratory rate was measured by one minute observation of chest wall movements. Tachypnea was defined as a respiratory rate of more than 60/minute (age < 2 months), more than 50/minute (age 2-12 months) and more than 40/minute (age > 12 months) according to WHO recommendations (13). White blood cell count (WBC) and C-reactive protein (CRP) were performed at the discretion of the attending physician and considered as additional diagnostic determinants.

Final diagnoses were based on information in the standardized medical records, and the results of all available additional diagnostic tests, and established by consensus (JR, HM). The final diagnosis pneumonia was based on a reference standard (pulmonary infiltrate on chest radiograph). Pulmonary infiltrate was defined as the presence of micronodular or macronodular infiltrations or consolidation in the lung on chest radiograph. Each chest radiograph was interpreted by two board certified radiologists (R1 and R2), blinded for clinical information. The inter-observer agreement for identification of a pulmonary infiltrate was good (kappa= 0.64, 95% confidence interval 0.52-0.75). All chest radiographs with disagreement on the presence of a pulmonary infiltrate (30/170) were reviewed by a pediatric pulmonologist (R3) to reach final judgment on the presence of a pulmonary infiltrate. A follow-up period of one week was the standard for ruling out the possibility of a missed pneumonia and to avoid verification bias. Data on recovery were collected at a control visit or by telephone call by one of the pediatricians or pediatric residents.

Statistical analysis

Associations of potential diagnostic determinants from patient history and physical examination with the presence of pneumonia were assessed using the Chi-square test for categorical variables, and the Mann-Whitney test for continuous variables. Candidate predictor variables with a univariable p-value <0.20 were subsequently entered into multivariable logistic regression analyses to develop a 'clinical model' (20). Stepwise backward selection, using the likelihood ratio test (LR-test), was performed to assess the contribution of each candidate predictor variable to the predictive performance of the model. At every step the least significant predictor variable was removed until the model only contained variables with a p-value <0.20. These variables were considered important predictors of pneumonia. Two clinical models were fitted: clinical model I was derived using data of all patients and clinical model II was based only on data of the patients in whom laboratory tests were performed.

WBC and CRP were added to the clinical model. The first model was only derived from data of patients who had laboratory tests performed. In the second model, multiple imputation methods were used to complete missing laboratory test values in patients who had no tests performed. Multiple imputation allows logistic regression on completed data sets. Multiple imputation methods may reduce bias and increase efficiency in regression analyses (16). We first assessed the association between clinical variables and missingness of laboratory test results, and the association of pneumonia as diagnosis with missingness of laboratory tests. The imputation model included 3 types of variables, as suggested by Van Buuren et al. (21): the patient characteristics and laboratory tests considered in the prediction model; auxilliary variables (associated with missingness, but not included in the prediction model); and the diagnosis. We created 5 imputed data sets to incorporate the uncertainty of the imputation process in the final logistic regression coefficients of the prediction model.

For each of the four models (clinical model I and II, and clinical +lab model I and II), the ability to discriminate between patients with and without pneumonia was quantified with the area under the receiver operating characteristic curve (AUC). The AUC can range from 0.5 (no discrimination) to 1.0 (perfect discrimination). The Hosmer & Lemeshow goodness of fit test was used to assess the reliability of the models. Bootstrapping techniques were used to adjust the AUC for statistical optimism. Bootstrap samples were drawn with replacement as an internal validation technique

(22). Models were developed in each bootstrap sample, including the backward stepwise selection strategy, and tested in the original sample. The average difference in AUC over 200 iterations indicated the optimism, which was used to correct the performance estimates of the original models (20, 23).

We used SPSS software (version 12.0, SPSS Inc, Chicago, Ill) for data management and initial logistic regression analyses. We used the *Design* library (23) for R (version 2.3.1) for imputation of missing values (aregImpute and fit.mult.impute functions) and bootstrap validation (validate.lrm function).

Results

A total of 2,821 patients presented with fever during the study period. 831 patients were excluded because of an underlying chronic condition or because their ED-visit was a scheduled follow-up visit. Of the 1,990 eligible patients, 504 presented with fever and cough. Their mean age was 2.3 years, and a primary respiratory tract infection was diagnosed in 434 patients (86%), including pneumonia in 78 patients (15%, Table 1). Three radiographs were taken at follow-up within one week of the first ED-visit; two were positive for a pulmonary infiltrate. Seventy patients (14%) had other final diagnoses (among others urinary tract infection, gastro-enteritis). Laboratory tests (WBC and CRP) were performed in 242 patients.

Table 1. General characteristics of the patients in whom laboratory tests were performed, patients in whom laboratory tests were not performed, and of the total population.

Characteristics	Lab tests performed (n=242)	Lab tests not performed (n=262)	Total population (n=504)
Male gender *	128 (53)	113 (43)	241 (48)
Age (per year) †	2.2 (2.2)	2.5 (2.4)	2.3 (2.3)
Febrile episode (per day) †	3.4 (2.5)	2.5 (2.1)	2.9 (2.3)
Temperature †	39.3 (0.8)	39.0 (0.8)	39.2 (0.8)
Tachypnea *	85 (35)	35 (13)	120 (24)
Chest-wall retractions *	56 (23)	45 (17)	101 (20)
Poor peripheral circulation *	36 (15)	19 (7)	55 (11)
Sick impression *	181 (75)	137 (52)	318 (63)
White blood cell count (per 10^9/L) †	14.3 (7.7)	-	-
Serum C-reactive protein (per 10 mg/L) †	59 (86)	-	-
Pneumonia present	69 (29)	9 (3)	78 (15)

* Absolute number (percentage)
† Mean ± standard deviation

Table 1 shows the general characteristics of patients who had laboratory tests performed (n=242), patients who had no laboratory tests performed (n=262) and for the total population (n=504). Several variables were associated with the decision to perform laboratory tests at p<0.20 levels: duration of the febrile episode, temperature, tachypnea, clinical impression, peripheral circulation, and gender. Patients without lab only had pneumonia in 9 of 262 cases, reflecting selective missingness of lab values.

Table 2. Univariable analysis: association of clinical characteristics with the presence of pneumonia in patients who had laboratory tests performed and in the total population.

	Lab tests performed (n=242)			**Total population (n=504)**		
	Pneumonia			**Pneumonia**		
Characteristics	***present*** **(n=69)**	***absent*** **(n=173)**	**OR (95% CI)**	***present*** **(n=78)**	***absent*** **(n=426)**	**OR (95% CI)**
Male gender *	37 (54)	91 (53)	1.0 (0.5-1.7)	40 (51)	201 (47)	0.8 (0.5-1.4)
Age (per year) †	2.8 (2.9)	2.0 (1.9)	1.1 (1.0-1.3)**	2.9 (3.0)	2.2 (2.1)	1.1 (1.0-1.2)**
Febrile episode (per day) †	3.9 (2.6)	3.2 (2.4)	1.1 (1.0-1.2)**	3.9 (2.5)	2.8 (2.3)	1.2 (1.1-1.3)**
Temperature †	39.3 (0.8)	39.3 (0.8)	1.1 (0.8-1.6)	39.3 (0.8)	39.1 (0.7)	1.3 (1.0-1.8)**
Tachypnea *	32 (46)	53 (31)	2.0 (1.1-3.5)**	36 (46)	84 (20)	3.5 (2.1-5.8)**
Chest-wall retractions *	19 (28)	37 (21)	1.4 (0.7-2.7)	22 (28)	79 (19)	1.7 (1.0-3.0)**
Poor peripheral circulation *	10 (15)	26 (15)	1.0 (0.4-2.1)	10 (13)	45 (11)	1.2 (0.6-2.6)
Sick impression *	56 (81)	125 (72)	1.7 (0.8-3.3)**	61 (78)	257 (60)	2.4 (1.34.2)**

* Absolute number (percentage)
† Mean ± standard deviation
** p-value < 0.20

Table 2 shows the results of the univariable analysis of the association of clinical characteristics with the presence of pneumonia in the patients who had laboratory tests performed and in the total population. In the selected group of patients, i.e. with laboratory tests performed, the univariable odds ratios were slightly lower than in the total population. Predictor variables with a univariable p-value < 0.20 were subsequently included in the multivariable analysis. Temperature and chest-wall retractions had a p-value <0.20 in the total population but >0.20 in the group of patients with laboratory tests.

Clinical models

As shown in Table 3, age, duration of the febrile episode and tachypnea were associated with the presence of radiographic pneumonia in the multivariable analyses in both the total population and the patients with laboratory tests.

The multivariable odds ratios of the predictor variables were quite similar in the selection of 242 patients with lab and the total group of 504 patients, except for the odds ratio of tachypnea which was 2.3 and 4.2 respectively (Table 3). The AUC of clinical model I was lower than model II: 0.65 (95%CI 0.57 – 0.72) versus 0.74 (95%CI 0.67-0.79). When corrected for optimism the AUCs of model I and II were 0.62 and 0.72 respectively.

Clinical + lab models

WBC and CRP had additional diagnostic value, i.e. their contribution was significant when added to clinical model I and II (table 3, clinical + lab model I and II). Clinical + lab model I, derived from data of patients in whom laboratory tests were performed (n=242) had an AUC of 0.75 (95% CI 0.67 – 0.82), and 0.72 when adjusted for optimism. Clinical + lab model II was derived from data on 504 patients of which 262 had no laboratory tests performed, and thus had CRP and WBC values imputed. The model had an AUC of 0.80 (95%CI 0.74-0.85). The Hosmer & Lemeshow goodness of fit test showed no lack of fit in any of the models.

Table 3. Results of the multivariable logistic regression analyses: predictors for the presence of radiographic pneumonia ($p < 0.20$).

	Clinical model I (n=242)	**Clinical model II (n=504)**	**Clinical + lab model I (n=242)**	**Clinical + lab model II (n=504) *imputed***
Characteristics	Odds ratio (95%CI)*	Odds ratio (95%CI)*	Odds ratio (95%CI)*	Odds ratio (95%CI)*
Age (per year)	1.1 (1.0-1.3)	1.1 (1.0-1.2)	1.1 (1.0-1.3)	1.1 (1.0-1.2)
Febrile episode (per day)	1.1 (1.0-1.3)	1.2 (1.1-1.3)	1.1 (1.0-1.2)	1.2 (1.1-1.3)
Tachypnea	2.3 (1.3-4.2)	4.2 (2.4-7.1)	2.0 (1.0-4.2)	3.8 (2.1-6.9)
White blood cell count (per 10^9/L)			1.06 (1.02-1.11)	1.06 (1.01-1.10)
Serum C-reactive protein (per 10 mg/L)			1.07 (1.03-1.11)	1.07 (1.03-1.11)
AUC (95% CI)	0.65 (0.57 – 0.72)	0.74 (0.57 – 0.72)	0.75 (0.67 – 0.82)	0.80 (0.74 – 0.85)
AUC (95% CI) adjusted for optimism	0.62 (0.55 – 0.67)	0.72 (0.55 – 0.67)	0.72 (0.65 – 0.79)	0.78 (0.72-0.82)

* Multivariable odds ratio

Discussion

We found that a combination of clinical characteristics and laboratory test results could reasonably predict radiographic pneumonia in children presenting with fever and cough. Important predictors were: Age, duration of fever, and tachypnea, white blood cell count (WBC) and C-reactive protein (CRP).

In this study we mimicked the clinical decision making process in which the first decision is made after patient history taking and physical examination. Therefore a stepwise modeling approach was used, with laboratory tests added to clinical characteristics. The inclusion criteria for this study (children presenting with cough and fever) were chosen to define a clinically recognizable group of patients. The 'a priori risk' of radiographic pneumonia (i.e. prevalence) was 16%. In the group of patients who had laboratory tests performed, the prevalence of radiographic pneumonia was 29% (69/242), in contrast to only 3% in the group of patients in whom the laboratory tests were not performed (9/262).

Clinical models

Predicting pneumonia in a selected group of patients (all patients who had laboratory tests performed, clinical model I) based on clinical characteristics, resulted in a moderate discriminative ability of the model (AUC 0.65). This can be explained by the fact that the group of children who had laboratory tests performed was more homogeneous than the total study population. In clinical model II, all patients presenting with fever and cough were included, as information on clinical characteristics was available for all patients. The discriminative ability of this model (AUC 0.75) was significantly better than that of clinical model I, which can be partly explained by a larger heterogeneity in the total group of febrile children (n=504) than in the selected group of children with laboratory tests (n=242).

The WHO promoted an algorithm in which the presence of tachypnea indicated pneumonia (13). Palafox found that the sensitivity of tachypnea was 74%, and that this symptom was the most useful clinical sign to identify pneumonia (10). Furthermore, several studies found that the absence of tachypnea reliably excluded pneumonia (8, 9, 14). In our study, tachypnea was indeed a strong independent predictor of pneumonia (OR 4.2, 95%CI 1.7-5.9), when all children with fever and cough

were included. The absence of tachypnea, however, did not rule out pneumonia.

We did not consider auscultatory findings (i.e. crackles, wheeze) as potential predictor variables. Margolis found that inter-observer agreement was good for most observations on physical examination, but only fair for the results of chest auscultation (9). Moreover, physicians more often agreed on the presence of a finding than on the absence, whereas most guidelines rely on the absence of certain signs (9).

Clinical + lab models

Both models including additional laboratory tests (i.e. WBC, CRP, the 'clinical + lab models') had a better diagnostic ability than the clinical models. Therefore, in predicting radiographic pneumonia, WBC and CRP add significantly to the diagnostic value of clinical characteristics. This is indicated by the significant higher AUCs of the clinical+lab models compared to the clinical models and also by the minimal decrease in odds ratios of the clinical variables in the clinical+lab models. The differences between clinical+lab model I and II can, as shown in the clinical models, partly be explained by the contrast between patients with and without pneumonia, which was larger in model II (n=504). Imputation of the missing laboratory values was theoretically motivated by the association between missingness of the lab values and the diagnosis. The increase in AUC from clinical model I to clinical+lab model I (n=242) was however larger than in model II (n=504); 0.65 to 0.74 (model I) versus 0.74 to 0.80 (model II). Therefore, it seems that imputation of missing values did not result in an overestimation of the diagnostic value of laboratory test results, which is also indicated by comparable odds ratios for WBC and CRP in both clinical+lab models. A prediction model for radiographic pneumonia based on data completed by multiple imputation of missing values was reliable and had a better discriminative ability than a model that was only based on available data.

To appreciate the results of this study, some aspects need to be addressed. First, pneumonia was defined as the presence of a pulmonary infiltrate on the chest radiograph (8-10, 14, 24). However, the chest radiograph findings may be negative in patients with early bacterial pneumonia (25). Therefore, duration of the fever episode might be associated with the reference standard, resulting in overestimation of the predictive value of fever duration. The finding of consolidation for the diagnosis of pneumonia in children was found to be highly reliable, in contrast to the moderate inter-observer agreement among radiologists concerning chest radiographs in general (26-28). In

our study we found good inter-observer agreement for the finding of consolidation (kappa 0.64). Second, for practical and ethical considerations, a chest radiograph was not taken in all patients. Chest radiographs were obtained at the discretion of the attending ED-physician in 33% of all patients (n=170). The rate of positive chest radiographs (i.e. pulmonary infiltrate present) was 46%, which is substantially higher than previously reported rates, ranging from 9% to 36% (8, 10, 29). Furthermore, to prevent verification bias, all discharged patients received follow-up, either by a control visit or telephone call, to detect clinical detoriation or a missed pneumonia. Two pneumonias were identified in 82 follow-up visits. If still some cases of pneumonia were missed, this would have resulted in an underestimation of the models' ability to predict pneumonia. Finally, we used a dataset of 504 patients, with 78 cases of pneumonia. By allowing more than 1 potential predictor variable per 10 cases initially (7 clinical and 2 laboratory variables), the prediction model may be slightly overfitted (20, 23). The results of the bootstrapping analyses however, showed that the AUC decreased only moderately for all four models. Hence, we expect good discriminative ability for patients in a similar setting, but further external validation in new patients is required before clinical application can be advocated (30). Clinical prediction models, as presented in this study, may be used by emergency department nurses immediately after nurse assessment to initiate the diagnostic process. The prediction model should then be presented as a clinical decision rule, and should be further validated.

In conclusion, diagnosing pneumonia in children remains difficult, but clinicians seem to be able to discriminate children at high risk for pneumonia from children who are at low risk, based on clinical characteristics. Laboratory test results (i.e. WBC and CRP) have additional diagnostic value for radiographic pneumonia. More generally, patient selection has an important effect on the discriminative ability of a prediction model; imputation of missing values for diagnostic test results is a sensible approach (21, 31). Laboratory test results are important to guide the decision whether or not to take a chest radiograph.

Acknowledgement

We gratefully acknowledge the ED-personnel of the Sophia Children's Hospital for their participation and careful collection of all required data. We are grateful to Dr. R. Failo for interpretation of chest radiographs.

References

1. Alexander ER, Foy HM, Kenny GE, Kronmal RA, McMahan R, Clarke ER, et al. Pneumonia due to Mycoplasma pneumoniae. Its incidence in the membership of a co-operative medical group. N Engl J Med 1966;275(3):131-6.
2. Foy HM, Cooney MK, Maletzky AJ, Grayston JT. Incidence and etiology of pneumonia, croup and bronchiolitis in preschool children belonging to a prepaid medical care group over a four-year period. Am J Epidemiol 1973;97(2):80-92.
3. Murphy TF, Henderson FW, Clyde WA, Jr., Collier AM, Denny FW. Pneumonia: an eleven-year study in a pediatric practice. Am J Epidemiol 1981;113(1):12-21.
4. Wright AL, Taussig LM, Ray CG, Harrison HR, Holberg CJ. The Tucson Children's Respiratory Study. II. Lower respiratory tract illness in the first year of life. Am J Epidemiol 1989;129(6):1232-46.
5. Jokinen C, Heiskanen L, Juvonen H, Kallinen S, Karkola K, Korppi M, et al. Incidence of community-acquired pneumonia in the population of four municipalities in eastern Finland. Am. J. Epidemiol. 1993;137(9):977-988.
6. Sinaniotis CA. Community-acquired pneumonia: diagnosis and treatment. Pediatr Pulmonol Suppl 1999;18:144-5.
7. Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, et al. A Prediction Rule to Identify Low-Risk Patients with Community-Acquired Pneumonia. N Engl J Med 1997;336(4):243-250.
8. Lynch T, Platt R, Gouin S, Larson C, Patenaude Y. Can We Predict Which Children With Clinically Suspected Pneumonia Will Have the Presence of Focal Infiltrates on Chest Radiographs? Pediatrics 2004;113(3):e186-189.
9. Margolis P, Gadomski A. Does This Infant Have Pneumonia? JAMA 1998;279(4):308-313.
10. Palafox M, Guiscafre H, Reyes H, Munoz O, Martinez H. Diagnostic value of tachypnoea in pneumonia defined radiologically. Arch. Dis. Child. 2000;82(1):41-45.
11. Dai Y, Foy HM, Zhu Z, Chen B, Tong F. Respiratory rate and signs in roentgenographically confirmed pneumonia among children in China. Pediatr Infect Dis J 1995;14(1):48-50.
12. Tanen DA, Trocinski DR. The use of pulse oximetry to exclude pneumonia in children. The American Journal of Emergency Medicine 2002;20(6):521-523.
13. WHO. The management of acute respiratory infections in children. *Practical guidelines for outpatient care*. Geneva: WHO, 1995.
14. Leventhal JM. Clinical predictors of pneumonia as a guide to ordering chest roentgenograms. Clin Pediatr (Phila) 1982;21(12):730-4.
15. Arnold AM, Kronmal RA. Multiple Imputation of Baseline Data in the Cardiovascular Health Study. Am. J. Epidemiol. 2003;157(1):74-84.
16. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am J Epidemiol 1995;142(12):1255-64.
17. British Thoracic Society of Standards of Care Committee. BTS Guidelines for the Management of Community Acquired Pneumonia in Childhood. Thorax 2002;57(90001):1i-24.
18. Campbell H, Byass P, Lamont AC, Forgie IM, O'Neill KP, Lloyd-Evans N, et al. Assessment of clinical criteria for identification of severe acute lower respiratory tract infections in children. Lancet 1989;1(8633):297-9.
19. Taylor JA, Del Beccaro M, Done S, Winters W. Establishing clinically relevant standards for tachypnea in febrile children younger than 2 years. Arch Pediatr Adolesc Med 1995;149(3):283-7.
20. Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med 2000;19(8):1059-79.
21. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis.

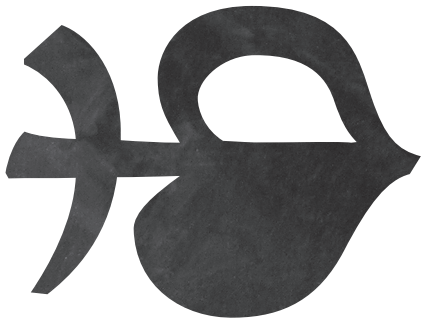
Stat Med 1999;18(6):681-94.
22. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001;54(8):774-81.
23. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15(4):361-87.
24. Rothrock SG, Green SM, Fanelli JM, Cruzen E, Costanzo KA, Pagane J. Do published guidelines predict pneumonia in children presenting to an urban ED? Pediatr Emerg Care 2001;17(4):240-3.
25. Mulholland EK, Simoes EA, Costales MO, McGrath EJ, Manalac EM, Gove S. Standardized diagnosis of pneumonia in developing countries. Pediatr Infect Dis J 1992;11(2):77-81.
26. Davies HD, Wang EE, Manson D, Babyn P, Shuckett B. Reliability of the chest radiograph in the diagnosis of lower respiratory infections in young children. Pediatr Infect Dis J 1996;15(7):600-4.
27. Albaum MN, Hill LC, Murphy M, Li YH, Fuhrman CR, Britton CA, et al. Interobserver reliability of the chest radiograph in community-acquired pneumonia. PORT Investigators. Chest 1996;110(2):343-50.
28. Hopstaken RM, Witbraad T, van Engelshoven JM, Dinant GJ. Inter-observer variation in the interpretation of chest radiographs for pneumonia in community-acquired lower respiratory tract infections. Clin Radiol 2004;59(8):743-52.
29. Mahabee-Gittens EM, Grupp-Phelan J, Brody AS, Donnelly LF, Bracey SE, Duma EM, et al. Identifying children with pneumonia in the emergency department. Clin Pediatr (Phila) 2005;44(5):427-35.
30. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. J Clin Epidemiol 2003;56(9):826-32.
31. Oostenbrink R, Moons KG, Bleeker SE, Moll HA, Grobbee DE. Diagnostic research on routine care data: prospects and problems. J Clin Epidemiol 2003;56(6):501-6.

Chapter 5

Diagnostic value of C-reactive protein in febrile children

Jolt Roukema, Ewout W. Steyerberg, Johan van der Lei, Henriette A. Moll

Submitted



Abstract

Background Most children presenting with fever at a hospital emergency department (ED) have a benign viral disease. A minority however suffers from serious bacterial infection (SBI). It is a daily challenge for paediatricians to discriminate between those with and without SBI. Frequently, C-reactive protein (CRP) and white blood cell count (WBC) are used to support diagnostic and therapeutic decisions.

Objective To determine the (added) value of CRP in the diagnosis of SBI in febrile children.

Methods Data were collected of all febrile children presenting to the ED of the Sophia Children's Hospital in Rotterdam, the Netherlands from July 2003 until December 2005, who underwent laboratory evaluation. Patients with chronic co-morbidity and patients who used antibiotics within one week prior to the ED visit were excluded. SBI was defined as culture or radiographically proven bacterial infection. Sensitivity and specificity of CRP and WBC for SBI were calculated. The added diagnostic value of CRP and WBC for SBI was assessed in multivariable logistic regression analyses.

Results A total of 768 children with a median age of 1.7 years and 56% male were included. A SBI was diagnosed in 174 children (23%). At a cutoff of 10 and 40 mg/L, CRP had a sensitivity of 0.90 and 0.65 and a specificity of 0.39 and 0.78 for detecting SBI respectively. WBC had, at a cut-off of 10 and $15*10^9$/L, a sensitivity of 0.81 and 0.51 and a specificity of 0.37 and 0.69 respectively. In univariable logistic regression analyses, CRP had an area under the curve of 0.79 (95% confidence interval [CI] 0.75-0.83) and WBC of 0.65 (95% CI 0.60-0.70). A multivariable model including clinical characteristics (age, temperature, duration of the fever), WBC and CRP had a similar AUC as CRP alone (0.79, 95%CI 0.75-0.83).

Conclusion C-reactive protein is a reliable marker for SBI in febrile children. A rapid CRP-test may be useful for the diagnostic management of febrile children presenting in a hospital emergency department.

Introduction

Fever is among the most common presenting signs of illness in children. Between 10 and 20 percent of all pediatric visits to hospital emergency departments are due to febrile illness [1-3]. Most children will have a benign, mostly self-limiting viral disease, but a small proportion will have a serious bacterial infection (SBI). Especially young children with no identifiable source of the fever are at risk for SBI [4, 5]. Several guidelines and algorithms have been developed to support

the diagnostic management (e.g. antibiotic therapy or hospital admission) of febrile children [6-9]. In daily practice, patient characteristics (i.e. general characteristics, patient history, and physical examination), clinical experience and guidelines are used to identify febrile children at high or low risk for SBI.

White blood cell count (WBC) and C-reactive protein (CRP) are among the most commonly applied laboratory tests to support the diagnosis in febrile children [10, 11]. Although WBC is traditionally used as an indicator of SBI, CRP appeared to be a more valuable laboratory test in the evaluation of young children with fever without apparent source [12-17]. Studies published to date however, mostly assessed the diagnostic value of WBC and CRP in detecting (occult) bacteraemia rather than SBI [1, 11, 17, 18].

A decreasing incidence of SBI in children and the invasive character of further laboratory testing (including urine and blood culture, and lumbar punctures), justifies a selective diagnostic approach to febrile children. The availability of rapid, minimally invasive CRP-tests may serve clinicians in deciding on diagnostic and therapeutic management. Rapid tests for CRP have been demonstrated to be reliable when routinely used in GP practices, and may be superior to WBC in predicting SBI [14, 19]. The aim of this study was to determine the (added) value of C-reactive protein in the diagnosis of SBI in febrile children.

Patients and methods

Data were prospectively collected at the emergency department (ED) of the ErasmusMC, Sophia Children's Hospital in Rotterdam, the Netherlands. The Sophia Children's Hospital is an inner city university hospital, and the ED annually delivers general pediatric care to approximately 4,500 children up to 16 years of age. We prospectively enrolled all patients (1 month–16 years) presenting from July 2003 until December 2005 at the ED with fever (i.e. body temperature ≥ 38.0°C), and had laboratory evaluation (complete blood count and quantitative CRP concentration). Laboratory tests were obtained at the discretion of the attending ED physician. Patients with chronic co-morbidity (e.g. malignancies, cystic fibrosis, severe psychomotor retardation) and patients with a history of using antibiotics within 1 week of their presentation to the ED were excluded. This study on diagnostic management of febrile children was approved by the Institutional Review Board.

ED-nurses routinely registered all febrile children using a structured data entry application (OpenSDE) that, for the purpose of this study, was tailored for data collection on febrile children [20]. Patient characteristics (gender, age, reason for encounter), duration of the febrile episode, and height of the fever were collected during or immediately after ED-nurse evaluation.

All final diagnoses were classified as either serious bacterial infection (SBI) or non-SBI. SBI was defined as culture or radiographically proven bacterial infection (e.g. meningitis, sepsis, bacteraemia, pneumonia, urinary tract infection, bacterial gastroenteritis, osteomyelitis, cellulitis or ethmoiditis) [13, 21]. A follow-up period of one week was the standard for ruling out the possibility of a missed diagnosis of SBI.

Statistical analysis

General characteristics and laboratory test results were compared between children with and without SBI, using a Chi-square test for categorical and the Mann-Whitney U test for continuous variables. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of WBC and CRP were assessed using different cut-off points, based on previous literature (i.e. WBC: 10, 15, and $20*10^9$/L, and CRP: 10, 40, and 70 mg/L) [12, 14, 18, 22].

In order to assess the added value of CRP to the diagnosis of SBI, different logistic regression analyses were performed. First, two univariable logistic regression models were fitted to quantify the unadjusted diagnostic value of C-reactive protein and of WBC for the presence of an SBI:

1) log odds (SBI) = $\alpha + \beta$ * CRP

2) log odds (SBI) = $\alpha + \beta$ * WBC

Here, α represents the intercept, β the regression coefficient.

Subsequently, we aimed to assess the added effect of CRP and WBC over a set of readily available clinical characteristics, known to be correlated to the presence of SBI, namely age, temperature, and duration of the febrile episode [1, 11, 13, 23, 24]. The following multivariable logistic regression models were considered:

3) log odds (SBI) = $\alpha + \beta_i$ * Clinical characteristic i

4) log odds (SBI) = $\alpha + \beta_i$ * Clinical characteristic i + β_2 * CRP

5) log odds (SBI) = $\alpha + \beta_i$ * Clinical characteristic i + β_2 * WBC

6) log odds (SBI) = $\alpha + \beta_i$ * Clinical characteristic i + β_2 * CRP + β_3 * WBC

The ability to discriminate between patients with and without SBI was quantified with the area under the receiver operating characteristic curve (AUC). The AUC can range from 0.5 (no discrimination) to 1.0 (perfect discrimination). We also compared the percentage variance, explained by the models (Nagelkerke's R-square) [25, 26].

We used SPSS software (version 12.0, SPSS Inc, Chicago, Ill) for logistic regression analyses.

Results

From July 2003 until December 2005, 2,821 patients attended the ED of the Sophia Children's hospital with fever. A total of 1,034 children were excluded: 722 because of chronic co-morbidity that may have affected the diagnostic approach, 259 patients because of antibiotic use within one week prior to the ED visit, and 53 because their visit was a scheduled follow-up visit. Of the remaining 1,787 patients, 768 had laboratory tests (i.e. both complete blood count and C-reactive protein) performed and were thus enrolled in the present study. The prevalence of SBI was 12% in our total febrile population (n=2,821) and 23% in the enrolled patients (n=768). The SBI diagnoses among the enrolled patients included: pneumonia (47%), urinary tract infection (25%), bacterial meningitis (7%), bacterial gastroenteritis (4%), cellulitis (3%), sepsis (2%), other (12%).

Table 1 shows general characteristics of the enrolled patients, stratified for presence or absence of SBI. Children with SBI were older than children without SBI (median 2.8 versus 1.6 years), and had a higher body temperature and longer duration of the fever. Furthermore, both white blood cell count and C-reactive protein values were significantly higher in children with SBI than in children without.

Table 1. General characteristics, stratified for presence or absence of SBI

Characteristics	SBI present (n=174)	SBI absent (n=594)	p-value
General			
Male gender*	87 (50.0)	343 (58)	0.07
Age†	2.8 (1.0-4.8)	1.6 (0.8-3.2)	<0.001
Temperature (°C)†	39.4 (38.7-40.1)	39.2 (38.5-39.9)	0.03
Duration of the febrile episode (days)†	3.0 (1-4)	2.0 (1-3)	<0.0001
Laboratory tests			
White blood cell count (per 10^9/L)†	15.2 (11.2-21)	11.8 (8.5-16.4)	< 0.0001
Serum C-reactive protein (per mg/L)†	75 (25-165)	16 (5-36)	< 0.0001

* Absolute number (percentage)
† Median (25th-75th percentile).

In Table 2 and 3 the sensitivity, specificity, positive predictive value and negative predictive value of CRP and WBC are shown. At different cut-off points CRP had both higher sensitivity and specificity than WBC.

Table 2. Sensitivity, specificity, positive predictive value, and negative predictive value of C-reactive protein.

Cut-off point CRP (per mg/L)	Sensitivity	Specificity	PPV	NPV
> 10	0.90	0.39	0.30	0.93
> 40	0.65	0.78	0.47	0.88
> 70	0.52	0.89	0.59	0.86

CRP = C-reactive protein
PPV = positive predictive value
NPV = negative predictive value

Table 3. Sensitivity, specificity, positive predictive value, and negative predictive value of white blood cell count.

Cut-off point WBC (per 10^9/L)	Sensitivity	Specificity	PPV	NPV
> 10	0.81	0.37	0.27	0.87
> 15	0.51	0.69	0.32	0.83
> 20	0.25	0.88	0.38	0.80

WBC = white blood cell count
PPV = positive predictive value
NPV = negative predictive value

In the univariable logistic regression analyses, CRP had an area under the curve of 0.79 and WBC of 0.65, indicating that the discriminative value of CRP alone was better than WBC alone (see table 4). The model including clinical characteristics (age, temperature and duration of the fever) had an AUC of 0.66. Furthermore, the AUC of the univariable model 1, including CRP, was similar to model 6, the multivariable model including clinical characteristics, CRP and WBC. CRP adjusted for clinical variables only (model 4) had also the same AUC of 0.79 as model 1 and 6. This indicates that CRP has additional diagnostic value over the clinical characteristics age, temperature and duration of the fever, and over WBC.

Table 4. Logistic regression analyses

	Odds ratio (95% CI)	AUC (95% CI)	R^2
Model 1		0.79 (0.75-0.83)	0.24
CRP (per 10 mg/L)	1.14 (1.11-1.17)		
Model 2		0.65 (0.60-0.70)	0.07
WBC (per 10^9/L)	1.05 (1.06-1.10)		
Model 3		0.66 (0.62-0.70)	0.07
Age (per year)	1.11 (1.06-1.16)		
Temperature (per degree Celsius)	1.34 (1.12-1.61)		
Duration of the fever (per day)	1.13 (1.08-1.19)		
Model 4		0.79 (0.75-0.83)	0.28
Age (per year)	1.07 (1.01-1.14)		
Temperature (per degree Celsius)	1.09 (0.86-1.39)		
Duration of the fever (per day)	1.07 (0.99-1.16)		
CRP (per 10 mg/L)	1.13 (1.10-1.17)		
Model 5		0.72 (0.67-0.76)	0.15
Age (per year)	1.12 (1.06-1.18)		
Temperature (per degree Celsius)	1.27 (1.01-1.59)		
Duration of the fever (per day)	1.15 (1.08-1.24)		
WBC (per 10^9/L)	1.08 (1.05-1.11)		
Model 6		0.79 (0.75-0.83)	0.28
Age (per year)	1.08 (1.02-1.15)		
Temperature (per degree Celsius)	1.10 (0.86-1.39)		
Duration of the fever (per day)	1.08 (1.00-1.17)		
CRP (per 10 mg/L)	1.12 (1.08-1.15)		
WBC (per 10^9/L)	1.04 (1.00-1.07)		

CRP = C-reactive protein
WBC = White blood cell count
AUC = Area under the curve

In figure 1 and 2 the ROC curves of the models including CRP, and in figure 2 of the models including WBC, are shown. Figure 1 shows that clinical characteristics and WBC are of limited diagnostic value once CRP is known. Furthermore, the R-square value of the models including CRP was higher than the models without CRP (see table 4). Only small differences in R-square occurred when CRP was adjusted for clinical characteristics and WBC, indicating that CRP is an important independent predictor of SBI.

Figure 1. ROC-curves of the models including CRP

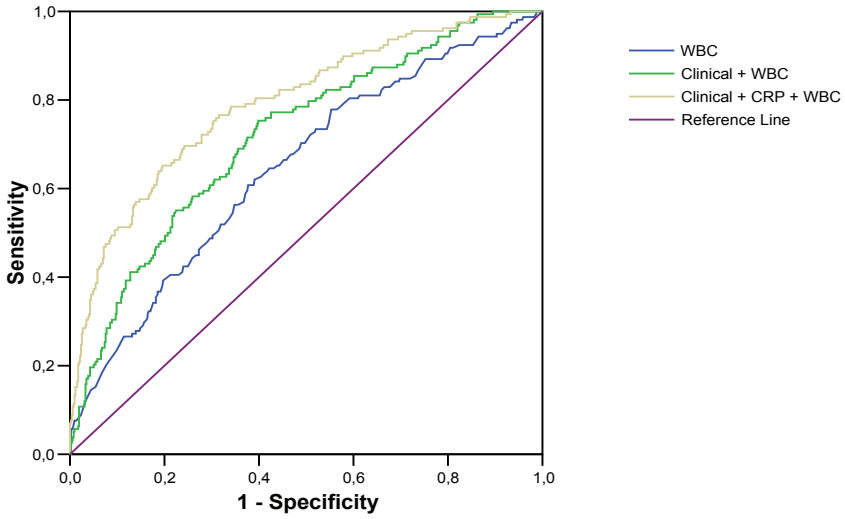
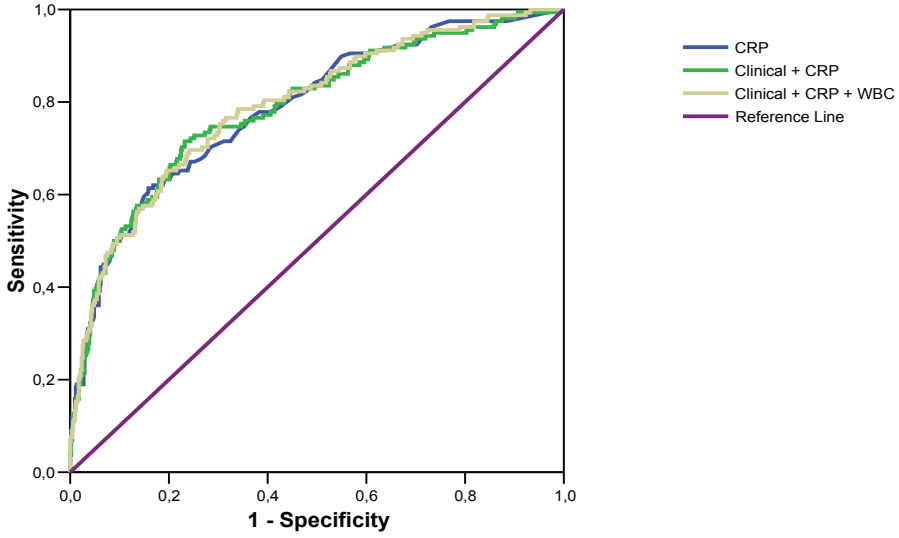


Figure 2. ROC-curves of the models including WBC



Discussion

C-reactive protein was found to be an important marker of SBI in children attending the emergency department with fever. CRP had better sensitivity and specificity for SBI than WBC, and the discriminative ability of CRP for SBI was better than that of WBC. CRP has additional diagnostic value over the clinical characteristics age, temperature and duration of the fever, and over WBC.

In the current literature, a broad range of cut-off points for CRP were proposed. Gendrel *et al* found a sensitivity of 98% and a specificity of 50%, when using a cut-off point of 10 mg/L. At a cut-off point of 40 mg/L, Galetto-Lacour *et al* found a sensitivity of 79% and specificity of 79% for SBI in young children with fever without source, who all underwent laboratory evaluation [14]. In a study by Pulliam *et al*, a sensitivity of 79% and specificity of 91% were reported at a cut-off point of 70 mg/L [12]. In this study children with a clinically undetectable source of the fever were included and all had laboratory tests performed.

In the current study the sensitivity of CRP for SBI was slightly lower than previously reported, at the same cut-off points. The specificity was comparable to previously reported values. Most studies published to date however, differed regarding patient inclusion and outcome measures. For example the studies by Galetto-Lacour *et al* and Pulliam *et al* restricted to young children (<36 months) with fever without source [12, 14]. Other studies had (occult) bacteraemia as outcome measure [1, 11, 17, 18].

In clinical practice sensitivity and specificity of diagnostic tests are less important than the diagnostic value of the test-result in an individual patient. The diagnostic value of a test should then be expressed as the change in the odds of having SBI, given the prior odds and the test result. Therefore we calculated univariable and multivariable (i.e. adjusted for other predictor variables) odds ratios (OR's) for CRP and WBC.

The discriminative ability of CRP was already good in the univariable analysis (AUC 0.79, 95% confidence interval [CI] 0.75-0.83) and was comparable to the recent findings by Hsiao *et al.* [23]. A multivariable model including important clinical characteristics, CRP and WBC had the same AUC. Moreover, the multivariable OR of CRP, adjusted for clinical characteristics and WBC was 1.12 per 10 mg/L, which was similar to the univariable OR (1.14 per 10 mg/L). This indicates that CRP is an important independent predictor for SBI.

To appreciate the results of this study, some aspects need to be addressed. First, in the present study we did not interfere with the clinical procedure; laboratory tests were requested at the discretion of the attending physician. Therefore we studied a more seriously ill group of patients, as indicated by the difference in prevalence of SBI between our total febrile population (12%) and the enrolled patients (23%). However, we do not expect that the relation between CRP and SBI would have been different if CRP was determined in all subjects, as the diagnostic value of CRP (OR and AUC) did not change after adjustment for clinical characteristics. Second, our aim was to assess the added diagnostic value of CRP over a set of clinical characteristics known to be associated with the presence of SBI. The set of clinical variables was not a full representation of all known predictors of SBI, but only the most important and readily obtainable objective parameters. Third, we assessed CRP in the broad range of clinical diversity in febrile children attending a hospital ED, and we did therefore not restrict ourselves to young children or children with fever without source. In subanalyses we found no differences in the sensitivity, specificity or diagnostic value of CRP and WBC for different age strata (i.e. 1-12 and 1-36 months of age, data not shown). Finally, we did not study the diagnostic value of procalcitonin (PCT), as this test was not routinely available at the hospital's ED [15].

In conclusion, CRP is a valuable diagnostic test for SBI in febrile children. The discriminative ability of CRP alone, as compared to a combination of important clinical characteristics, CRP and WBC, was similar. Implementation of a rapid CRP-test for febrile children presenting in the ED may increase the efficiency of the diagnostic process.

Acknowledgement

We gratefully acknowledge the ED-nursing staff of the Sophia Children's Hospital for their participation and careful collection of all required data.

References

1. Kuppermann, N., G.R. Fleisher, and D.M. Jaffe, *Predictors of occult pneumococcal bacteremia in young febrile children.* Annals of Emergency Medicine, 1998. 31(6): p. 679-687.
2. Nelson, D.S., K. Walsh, and G.R. Fleisher, *Spectrum and frequency of pediatric illness presenting to a general community hospital emergency department.* Pediatrics, 1992. 90(1): p. 5-10.
3. Slater, M. and S.E. Krug, *Evaluation of the infant with fever without source: An evidence based approach.* Emergency Medicine Clinics of North America, 1999. 17(1): p. 97-+.
4. Baker, M.D., L.M. Bell, and J.R. Avner, *Outpatient Management without Antibiotics of Fever in Selected Infants.* N Engl J Med, 1993. 329(20): p. 1437-1441.
5. Baraff, L.J., et al., *Probability of Bacterial-Infections in Febrile Infants Less Than 3 Months of Age - a Metaanalysis .1.* Pediatric Infectious Disease Journal, 1992. 11(4): p. 257-265.
6. Baraff, L.J., et al., *Practice Guidelines for Management of Infants and Children with Fever without Source (Fws) - Reply.* Pediatrics, 1994. 93(2): p. 349-351.
7. Baraff, L.J. and S.I. Lee, *Fever without Source - Management of Children 3 Months to 36 Months of Age.* Pediatric Infectious Disease Journal, 1992. 11(2): p. 146-151.
8. Harper, M.B., *Update on the management of the febrile infant.* Clinical Pediatric Emergency Medicine, 2004. 5(1): p. 5-12.
9. Isaacman, D.J. and K.D. Rogers, *Practice Guidelines for Management of Infants and Children with Fever without Source (Fws).* Pediatrics, 1994. 93(2): p. 346-347.
10. McCarthy, P.L., *Fever without apparent source on clinical examination.* Current Opinion in Pediatrics, 2003. 15(1): p. 112-120.
11. Isaacman, D.J., et al., *Predictors of bacteremia in febrile children 3 to 36 months of age.* Pediatrics, 2000. 106(5): p. 977-982.
12. Pulliam, P.N., M.W. Attia, and K.M. Cronan, *C-reactive protein in febrile children 1 to 36 months of age with clinically undetectable serious bacterial infection.* Pediatrics, 2001. 108(6): p. 1275-1279.
13. Bleeker, S.E., et al., *Predicting serious bacterial infection in young children with fever without apparent source.* Acta Paediatrica, 2001. 90(11): p. 1226-1232.
14. Galetto-Lacour, A., S.A. Zamora, and A. Gervaix, *Bedside procalcitonin and C-reactive protein tests in children with fever without localizing signs of infection seen in a referral center.* Pediatrics, 2003. 112(5): p. 1054-1060.
15. Hsiao, A.L. and M.D. Baker, *Fever in the new millennium: a review of recent studies of markers of serious bacterial infection in febrile children.* Curr Opin Pediatr, 2005. 17(1): p. 56-61.
16. Lacour AG, et al., *Procalcitonin, IL-6, IL-8, IL-1 receptor antagonist and C-reactive protein as identificators of serious bacterial infections in children with fever without localising signs.* Eur J Pediatr, 2001. 160(2): p. 95-100.
17. Isaacman, D.J. and B.L. Burke, *Utility of the serum C-reactive protein for detection of occult bacterial infection in children.* Arch Pediatr Adolesc Med, 2002. 156(9): p. 905-9.
18. Bonsu, B.K. and M.B. Harper, *Identifying febrile young infants with bacteremia: Is the peripheral white blood cell count an accurate screen?* Annals of Emergency Medicine, 2003. 42(2): p. 216-225.
19. Dahler-Eriksen BS, 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): p. 2064-75.
20. Roukema, J., et al., *Paper Versus Computer: Feasibility of an Electronic Medical Record in General Pediatrics.* Pediatrics, 2006. 117(1): p. 15-21.
21. Oostenbrink, R.M., et al., *Signs of meningeal irritation at the emergency department: How often bacterial meningitis?* Pedi-

atric Emergency Care June 2001;17(3):161-164, 2001.

22. Gendrel, D., et al., *Comparison of procalcitonin with C-reactive protein, interleukin 6 and interferon-alpha for differentiation of bacterial vs. viral infections.* Pediatr Infect Dis J, 1999. 18(10): p. 875-81.
23. Hsiao, A.L., L. Chen, and M.D. Baker, *Incidence and Predictors of Serious Bacterial Infections Among 57- to 180-Day-Old Infants.* Pediatrics, 2006. 117(5): p. 1695-1701.
24. Lee, G.M. and M.B. Harper, *Risk of bacteremia for febrile young children in the post-Haemophilus influenzae type b era.* Archives of Pediatrics & Adolescent Medicine, 1998. 152(7): p. 624-628.
25. Harrell, F.E., *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis.* Springer, New York, 2001.
26. Nagelkerke, N.J.D., *A note on a general definition of the coefficient of determination.* Biometrika, 1991. 78(No. 3): p. 691-692.

Chapter 6

Polytomous regression did not outperform dichotomous logistic regression in diagnosing children with fever without apparent source

Jolt Roukema, Rhiannon B van Loenhout, Ewout W Steyerberg
Karel G M Moons, Sacha E Bleeker, Henriëtte A Moll

Submitted

Abstract

Background Polytomous logistic models may be used to estimate the probabilities of the presence of multiple diagnoses simultaneously. The objective of this study was to compare results from polytomous and dichotomous logistic regression modelling in a diagnostic study with respect to the predictive ability and internal validity of the derived models.

Patients and Methods We analyzed data of a diagnostic study among 595 children aged 1-36 months, who attended the emergency department with fever without apparent source suspected of having a serious bacterial infection (SBI). Eleven potential diagnostic predictors were selected based on previous studies and literature. Outcome categories were SBI - subdivided in pneumonia and other-SBI (OSBI) - and non-SBI. Four models were developed; model 1 was a polytomous model, estimating probabilities for the three diagnostic outcome categories simultaneously. Model 2 and 3 were sequential dichotomous models, discriminating SBI and non-SBI in step 1, and pneumonia and OSBI in step 2, and differed in variable selection method. In model 4 each outcome category was opposed to the other two using separate (non-sequential) dichotomous models. All four models were compared with respect to the area under the receiver operating characteristic curve (AUC) and calibration (Hosmer and Lemeshow test) for each of the three outcome categories, and the selected predictors. Internal validation was performed using bootstrap techniques.

Results The AUCs of the three outcome categories were similar for each modelling strategy. The ability to discriminate pneumonia was good for each derived model, with an AUC of 0.81 in each of the models. Small differences were found in the predictors that were selected in the polytomous and dichotomous models. The developed models had comparable calibration and internal validity.

Conclusion A polytomous logistic regression analysis had results comparable to sequential and separate application of dichotomous logistic regression analyses. In diagnostic research, the definition of homogeneous and clinically relevant outcome categories may be more important than the choice of modelling strategy.

Introduction

In diagnostic practice, diagnoses are commonly based on multiple test results including symptoms, signs, and results of additional tests. In diagnostic research of multiple diagnostic tests, multivariable regression analyses are often used to calculate the independent contribution of different test results to include or exclude a particular diagnosis. The important test results and their relative contributions can then be use to develop a so-called diagnostic prediction model or scoring rule, which in turn can be used to support physicians in their decision making for further diagnostic work-up and treatment (1-6). The outcome in diagnostic research is generally dichotomous, i.e. a particular diagnosis - often referred to as the target disease - being present or absent. Accordingly, the resulting prediction model obtained from multivariable dichotomous logistic regression can be used to estimate the probability of presence of this target disease. One minus this probability is the probability that the target disease is absent, i.e. another diagnosis is present.

Dichotomization of the outcome in diagnostic research may however not always be preferable. Test results can point to the presence of different diseases compromising the discrimination between presence or absence of the target disease. A clinical example is the prediction of the presence of a serious bacterial infection (SBI) in young febrile children (7-9). But heterogeneity among the possible SBI-diagnoses in these children – e.g. meningitis, pneumonia, urinary tract infection - may adversely affect the ability of test results to discriminate between presence and absence of SBI-diagnoses in a dichotomous logistic regression model (10). In a polytomous logistic regression model, however, it is possible to estimate probabilities for more than two outcome categories at once (11, 12). A polytomous model allows that the test results or predictors have different effects on the different outcome categories. Polytomous modelling may thus be of value when the (diagnostic) outcome is rather heterogeneous. To date, polytomous logistic regression analysis has only rarely been used in diagnostic research (11, 12).

In this study we reanalysed the data of a recently conducted study among young children presenting at the hospital emergency department (ED) with fever without apparent source who were suspected of having SBI (8, 10). The present aim was to compare the results obtained from polytomous logistic regression analysis to the results obtained from dichotomous logistic regression analysis, with respect to selected predictors, diagnostic ability (discrimination and calibration) and internal validity of the derived models.

Patients and methods

Patients

The previously described study was conducted study among children presenting at the ED of the Sophia Children's Hospital in Rotterdam and Juliana Children's Hospital in The Hague The Netherlands, for the evaluation of acute fever without apparent source (FWS).(8, 10) In brief, between 1996 and 2002 595 children, aged 1-36 months, attended the ED's with acute FWS. FWS was defined as fever (body temperature ≥ 38.0°C) without source after evaluation by the general practitioner or after history taking by the paediatrician. Data on patient characteristics, test results, and final diagnoses were prospectively collected, and (if needed) completed by review of the standardised medical records and the computer-documented hospital information system.

Diagnostic outcome and diagnostic predictors (index tests)

As described before(8, 13), the final diagnosis was established in every patient with FWS either by laboratory testing, radiography testing, and/or consensus diagnosis. In a previous study based on a part of these data, a diagnostic prediction model using dichotomous logistic regression modelling was developed to discriminate between SBI versus non-SBI (8). SBI presence was based on the presence of laboratory or radiographically proven bacterial meningitis, sepsis or bacteraemia, pneumonia, urinary tract infection, bacterial gastroenteritis, osteomyelitis or ethmoiditis. Detailed descriptions of the applied reference tests and outcome diagnoses have been published (8, 13). The derived logistic regression model had an area under the receiver operating characteristic curve (AUC) of 0.75. External validation of the prediction rule showed a very disappointing discriminative ability, with an AUC of 0.60 (10, 14). As mentioned above, a possible explanation for this limited discriminative ability of the (dichotomous logistic regression) model includes heterogeneity among the SBI-diagnoses.

Therefore, in the current study, the outcome category SBI was subdivided in two groups: pneumonia and other-SBI (OSBI). Pneumonia was chosen as one of the three outcome categories because 40% of all SBI-diagnoses were pneumonias and because pneumonia has a specific diagnostic work-up and treatment strategy, making it clinically relevant to discriminate pneumonia from the other SBI diagnoses (15). For the present analysis, the three diagnostic outcome categories were thus pneumonia, OSBI, and non-SBI.

Based on earlier studies, we preselected for the present analysis eleven potential diagnostic predictors – obtained from patient history and physical examination - of the diagnostic outcome: gender, age, duration of fever, changed crying pattern, vomiting, sick impression, pale skin, impaired peripheral circulation, dyspnoea, temperature, and weeks of gestation (7, 8, 16, 17).

Statistical analyses

In the present study we compared the results of four modelling strategies (see Figure 1 and text below): 1. A polytomous logistic regression analysis (model 1), 2. Consecutive or sequential dichotomous logistic regression analyses (models 2 and 3), and 3. Separate (non-consecutive) dichotomous logistic regression analyses (model 4a-4c). For all modelling approaches except for modelling approach 2 in which a fixed set of variables was used, we first estimated the association of each diagnostic predictor and the presence of each outcome category separately using the Chi-square test for categorical, ANOVA-test for continuous variables and logistic regression to obtain odds ratios (univariable analysis).

Variables with a univariable p-value ≤ 0.20 were selected for the multivariable model development which always started with the overall model. Then the Likelihood Ratio (LR) test was used to assess the contribution of each predictor variable to the outcome prediction, by excluding one by one the least significant predictor variable from the model until the model only contained variables with a multivariable p-value ≤ 0.20.

For each developed model, the ability to discriminate between the considered diagnostic outcome categories was quantified using the AUC. To do this for the polytomous model, each outcome of the three categories was contrasted as presence versus its absence, e.g. presence of OSBI versus absence of OSBI. The AUC can range from 0.5 (no discrimination) to 1.0 (perfect discrimination). Subsequently, the Hosmer-Lemeshow goodness-of-fit test was used to assess the calibration of the models by comparing the predicted probabilities and observed outcome frequencies across deciles of risk for the dichotomized outcomes. Finally, internal validation of the developed models was performed using a bootstrap technique in which samples are drawn with replacement (18). In each bootstrap sample (n=200) the modelling approach was repeated - including the stepwise selection strategy (except for approach 2) - to develop a final model, which was then tested in the original sample. The apparent model performance after each bootstrap sample was compared to the original model performance. The average difference in AUC over the 200 comparisons was then used to cor-

rect the AUC's of the original models for statistical optimism (2, 3). The Statistical Package for the Social Sciences (SPSS) version 12.01 was used for dichotomous and polytomous logistic regression analyses. The bootstrap analyses were performed in S-plus (version 6.0, Insightful Inc, Seattle, WA, using the *Design* and *nnet* library) (2).

Description of the four modelling approaches

Model 1: Polytomous logistic regression model.

A polytomous model was fitted in which the probability of each outcome (pneumonia, OSBI and non-SBI) was estimated in one maximum likelihood procedure.

Model 2: Consecutive (sequential) dichotomous logistic regression models with fixed variables.

First, a dichotomous model was fitted to distinguish SBI from non-SBI (model 2a), using the same (fixed) variables as retained in the final polytomous model (model 1). A second model, using the same variables again, was then fitted to further discriminate the SBI diagnoses in pneumonia and OSBI (model 2b). Hence, the estimated probability of pneumonia was obtained by multiplying the probability of SBI, in model 2a by the probability of pneumonia in model 2b: P (pneumonia) = P (SBI) * P (pneumonia | SBI)

Model 3: Consecutive (sequential) dichotomous logistic regression models with selected variables based on statistical significance. The strategy of model 3 was similar to model 2, except for variable selection. In model 3 all candidate predictor variables were used instead of the fixed variables from the polytomous analysis as in model 2. First, a model was fitted to distinguish SBI from non-SBI (model 3a). This approach was methodologically identical to the previous model by Bleeker et al. (8). A second model was then fitted to further divide the SBI diagnoses in pneumonia and OSBI (model 3b). The estimated probability of pneumonia was again obtained by multiplying the probability of SBI in model 3a by the probability of pneumonia in model 3b, as described for approach 2.

Model 4: Separate dichotomous models

Finally, three separate dichotomous models were fitted, in which each outcome category (pneumonia, OSBI and non-SBI), was opposed to the other two. In model 4A, the probability of pneumonia was modelled, in model 4B the probability of OSBI, and in model 4C of non-SBI .

Results

Table 1 shows the patient characteristics and diagnostic predictors (index tests) across the diagnostic outcomes (pneumonia, OSBI, and non-SBI). An SBI was present in 23% (n=138) of the patients, subdivided in 9% having pneumonia (n=54) and 14% having OSBI (n=84). SBI was absent in 77% (n=457) of the patients.

Table 1. Patient characteristics stratified for diagnostic outcome; SBI (pneumonia and OSBI) and non-SBI.

	SBI (n=138)		Non-SBI (n=457)	
	Pneumonia	*OSBI*		
Characteristics	54 (9.1)	84 (14.1)	457 (76.8)	*P-value*
Gender§	27 (50.0)	45 (53.6)	262 (57.3)	0.518
Age (years)†	1.3 (0.8)	0.7 (0.7)	0.9 (0.8)	<0.001
Duration of fever (days)†	3.9 (3.6)	2.8 (2.3)	2.6 (2.3)	0.001
Changed crying pattern§	18 (33.3)	42 (50.0)	193 (42.2)	0.034
Vomiting§	32 (59.3)	31 (36.9)	155 (33.9)	0.006
Sick impression§	31 (57.4)	51 (60.7)	193 (42.2)	<0.001
Pale skin§	5 (9.3)	25 (29.8)	47 (10.3)	<0.001
Impaired peripheral circulation§	4 (7.4)	29 (34.5)	55 (12.0)	<0.001
Dyspnea§	19 (35.2)	11 (13.1)	34 (7.4)	<0.001
Temperature (°C)†	39.7 (1.0)	39.4 (1.1)	39.2 (1.0)	0.006
Weeks of gestation†	38.8 (2.5)	39.0 (4.2)	39.0 (2.8)	0.941

§ absolute number (percentage)
† mean (standard deviation)
SBI = serious bacterial infection
OSBI = other serious bacterial infection
Non-SBI = no serious bacterial infection

For each of the eleven diagnostic predictors four odds ratios (ORs) were calculated (SBI vs. non-SBI, pneumonia vs. non-SBI, OSBI vs. non-SBI, and pneumonia vs. OSBI). Table 2 gives an example of the calculation of the four ORs for 'dyspnea'. The univariable ORs of all diagnostic predictors are presented in table 3. Table 2 and 3 show that some diagnostic predictors had different accuracy for the different outcome categories. Dyspnea for example, had an OR of 3.5 between SBI and non-SBI and an OR of 6.8 between pneumonia and non-SBI. Gender and weeks of gestation, with univariable p-values >0.20 (table 1), were excluded from further model development in all modelling strategies, leaving 9 diagnostic determinants.

Table 2. Example of the calculation of the four univariable odds ratio's.

Dyspnea	Pneumonia	OSBI	Non-SBI	*total*
Present	19	11	34	64
Absent	35	73	423	531
total	54	84	457	595

OR SBI(Pneumonia & OSBI) / non-SBI = [(19+11)/ (35+73)] / (34/423) = (30/108) / (34/423) = 3.5
OR OSBI / non-SBI = (11/73) / (34/423) = 1.9
OR Pneumonia / non-SBI = (19/35) / (34/423) = 6.8
OR Pneumonia / OSBI = (19/35) / (11/73) = 3.6
SBI = serious bacterial infection
OSBI = other serious bacterial infection
Non-SBI = no serious bacterial infection
OR = odds ratio

Table 3. Results of univariable analyses.

	SBI/non-SBI	OSBI/non-SBI	Pneumonia/non-SBI	Pneumonia/OSBI
Potential diagnostic determinants	OR (95%CI)	OR (95%CI)	OR (95%CI)	OR (95%CI)
Gender‡	0.8 (0.6-1.2)	0.9 (0.5-1.4)	0.7 (0.4-1.3)	0.9 (0.4-1.7)
Age°	1.0 (0.8-1.3)	0.7 (0.5-0.9)	1.8 (1.3-2.5)	2.8 (1.7-4.5)
Duration of fever°	1.1 (1.0-1.2)	1.0 (0.9-1.1)	1.2 (1.1-1.3)	1.2 (1.0-1.3)
Changed crying pattern‡	1.0 (0.7-1.5)	1.5 (0.9-2.5)	0.6 (0.3-1.0)	0.4 (0.2-0.8)
Vomiting‡	1.6 (1.1-2.4)	1.1 (0.7-1.9)	2.5 (1.4-4.5)	2.2 (1.1-4.5)
Sick impression‡	2.3 (1.5-3.5)	2.4 (1.4-4.0)	2.2 (1.2-4.2)	0.9 (0.4-2.0)
Pale skin‡	2.4 (1.4-4.0)	3.8 (2.2-6.8)	0.9 (0.3-2.3)	0.2 (0.1-0.6)
Impaired peripheral circulation‡	2.2 (1.4-3.6)	3.6 (2.1-6.3)	0.6 (0.2-1.7)	0.2 (0.1-0.5)
Dyspnea‡	3.5 (2.0-5.9)	1.9 (0.9-3.9)	6.8 (3.5-13.1)	3.6 (1.5-8.4)
Temperature°	1.3 (1.1-1.6)	1.1 (0.9-1.4)	1.7 (1.2-2.3)	1.3 (1.0-1.9)
Weeks of gestation°	1.0 (0.9-1.1)	1.0 (0.9-1.1)	1.0 (0.9-1.1)	1.0 (0.9-1.1)

‡ categorical variable
° continuous variable
SBI = serious bacterial infection
OSBI = other serious bacterial infection
Non-SBI = no serious bacterial infection
OR = odds ratio
CI = confidence interval

Table 4 shows the results of the four modelling strategies. The final polytomous model contained seven variables. The AUC for discrimination of pneumonia was 0.82, 0.69 for OSBI, and 0.73 for non-SBI. The consecutive (sequential) dichotomous logistic regression models fitted with the seven predictors from the polytomous model (models 2A and 2B) showed an AUC for pneumonia of 0.81, 0.70 for OSBI, and 0.71 for non-SBI. Of the consecutive dichotomous models based

Table 4. Results of the multivariable logistic regression analyses. Variables included in the final models (multivariable p-value < 0.20).

	Polytomous	Sequential Dichotomous		Sequential Dichotomous		Separate Dichotomous		
Variables	Pneumonia/OSBI /non-SBI	2A SBI/non-SBI	2B Pneumonia/OSBI	3A SBI/non-SBI	3B Pneumonia/OSBI	4A Pneumonia	4B OSBI	4C non-SBI
Age	x	x	x		x	x	x	
Duration of fever	x	x	x	x	x	x		x
Changed crying pattern							x	
Vomiting	x	x	x	x		x	x	x
Sick impression	x	x	x	x			x	x
Pale skin	x	x	x	x			x	x
Impaired peripheral circulation	x	x	x	x	x	x	x	x
Dyspnoea	x	x	x	x	x	x		x
Temperature				x		x		x
Total # of variables	7	7	7	7	4	6	6	7
AUC*								
Pneumonia	0.82 (.76-.88)	0.81 (.75-.86)		0.81 (.75-.87)		0.82 (.75-.88)		
OSBI	0.69 (.62-.76)	0.70 (.63-.77)		0.70 (.63-.77)		0.66 (.59-.74)		
non-SBI	0.73 (.68-.79)	0.71 (.66-.77)		0.72 (.66-.77)		0.72 (.66-.77)		
Bootstrap corrected AUC								
Pneumonia	0.78	0.77		0.77		0.78		
OSBI	0.64	0.64		0.64		0.61		
non-SBI	0.69	0.67		0.68		0.68		

* AUC (95% CI)
SBI = serious bacterial infection
OSBI = other serious bacterial infection
Non-SBI = no serious bacterial infection
OR = odds ratio
CI = confidence interval

on predictor selection, model 3A (distinguishing SBI from non-SBI) contained seven variables and model 3B (dividing SBI in pneumonia and OSBI) contained only 4 predictors. In the latter, 5 predictors (changed crying pattern, vomiting, sick impression, pale skin, and temperature) were thus removed because of limited predictive contribution beyond the 4 selected predictors. The AUC for discrimination of pneumonia was 0.81, 0.70 for OSBI and 0.72 for non-SBI. Of the three separate (non-consecutive) dichotomous models (approach 4), model 4A predicting the presence of pneumonia versus its absence, contained six variables with an AUC of 0.82. Model 4B predicting presence of OSB contained six variables (AUC=0.66). Model 4C was essentially the same as model 3A with an AUC of 0.72.

The Hosmer-Lemeshow tests of all (dichotomous or dichotomized) models was non-significant, indicating no clear lack of fit. The bootstrap corrected AUCs in all models were .04 to .06 lower than the (above presented) apparent AUCs, reflecting moderate optimism.

Discussion

In this study we used different multivariable logistic regression strategies to predict the diagnostic outcome in children with fever without apparent source. Polytomous, sequential dichotomous and single dichotomous logistic regression analyses were performed. The ability to discriminate between the diagnostic outcomes in these patients (pneumonia, OSBI, and non-SBI) was similar across the different modelling strategies (table 4). The ability to predict the presence of pneumonia was good in each of the modelling strategies (AUC=0.81). Furthermore, all modelling strategies had comparable internal validity. Variable selection (model 3) did not affect the discriminative ability of the models as compared to the fixed variables dichotomous model (model 2).

In 1983, Wijesinha et al discussed the use of polytomous logistic regression to facilitate simultaneous prediction of more than two unordered outcome categories. Widespread implementation of the polytomous logistic model in diagnostic research did, however, not take place, although polytomous regression analysis is available in many statistical packages nowadays. The management of young children with fever without apparent source is an example of a diagnostic problem in which the use of polytomous modelling may have additional value. The major goal in the evaluation of these children is the identification of an SBI to support further diagnostic work-up or treatment decisions. Using dichotomous logistic regression analysis, a diagnostic prediction model was previously developed to distinguish SBI and non-SBI, with an AUC of 0.75 after bootstrapping (8). At external validation, the discriminative ability of this model was disappointing (AUC 0.60) (14). We hypothesized that this could be explained by heterogeneity among the SBI-diagnoses. We therefore subdivided for this study, the SBI outcome category in pneumonia and other-SBI (OSBI), to define more homogeneous outcome categories, possibly with more specific signs and symptoms. The result was indeed that all models had good discriminative ability for pneumonia, with an AUC of 0.81. However, discriminative ability for the other two outcome categories was only moderate, with an AUC of approximately 0.69 for OSBI, and 0.72 for non-SBI.

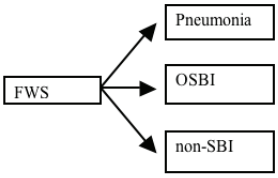
Differences – though relatively small - were found in the predictors that were selected in the

polytomous and dichotomous models. In each model, age, vomiting, and impaired peripheral circulation were selected as important predictors. The determinant 'changed crying pattern' was removed in all models, except for dichotomous model 4B, predicting OSBI versus pneumonia and non-SBI. Temperature was removed in five of the eight models as it did not contribute significantly to those models. Temperature only remained in the models discriminating SBI from non-SBI (model 3A and 4C) and in the model discriminating pneumonia from OSBI and non-SBI (model 4A). This is not surprising as higher body temperature is associated with the presence of an SBI (9, 19-22). The differences in predictor selection also reflect that the predictors had different effects for the three outcome categories. These differences in effects can be noted both in univariable and multivariable analyses.

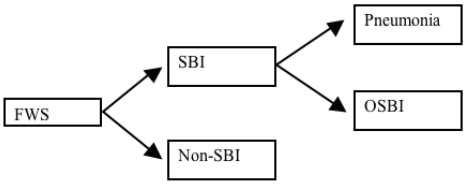
To appreciate the results of this study, some remarks have to be made. The number of predictor variables used in this study may have resulted in overfitted models. Eleven potential diagnostic predictors were analysed, although the smallest outcome category (pneumonia) contained only 54 cases. A moderate drop in AUC was noted in predicting each outcome category. This observation stresses the importance of internal and external validation of prediction models derived from small samples (10, 18, 23). For more reliable polytomous logistic regression analyses, larger datasets may be necessary.

In this study, a polytomous logistic regression analysis yielded results comparable to single and sequential application of dichotomous logistic regression analyses. We conclude therefore that the definition of homogeneous and clinically relevant outcome categories may be more important than the specific modelling strategy that is chosen.

Figure 1. Graphical representation of modelling strategies

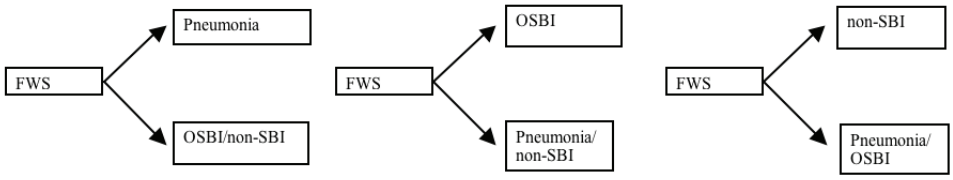


Model 1. Polytomous model. The probability of each outcome diagnosis was estimated in a single maximum likelihood procedure.



2A / 3A **2B / 3B**

Model 2 and 3. Consecutive or sequential dichotomous logistic regression models: modelling approach 2 with fixed variables, and modelling approach 3 with selected variables based on statistical significance. The 'A' model distinguished SBI and non-SBI, and the subsequent 'B' model between pneumonia and OSBI within the SBI-group



4A **4B** **4C**

Model 4. Three separate (non-consecutive) dichotomous logistic regression models.

FWS = fever without apparent source
SBI = serious bacterial infection
OSBI = other serious bacterial infection
Non-SBI = no serious bacterial infection

References

1. Moons KGM, Harrell FE. Sensitivity and Specificity should be De-emphasized in Diagnostic Accuracy Studies. Academic Radiology. 2003;10(6):670-2.
2. 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 Feb 28;15(4):361-87.
3. Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med. 2000 Apr 30;19(8):1059-79.
4. Knottnerus AJ, Leffers P. The influence of referral patterns on the characteristics of diagnostic tests. Journal of Clinical Epidemiology. 1992;45(10):1143-54.
5. Moons KGM, Grobbee DE. Diagnostic studies as multivariable, prediction research. J Epidemiol Community Health. 2002 May 1, 2002;56(5):337-8.
6. Oostenbrink R, Moons KGM, Bleeker SE, Moll HA, Grobbee DE. Diagnostic research on routine care data: Prospects and problems. Journal of Clinical Epidemiology. 2003;56(6):501-6.
7. Berger RM, Berger MY, van Steensel-Moll HA, Dzoljic-Danilovic G, Derksen-Lubsen G. A predictive model to estimate the risk of serious bacterial infections in febrile infants. Eur J Pediatr. 1996 Jun;155(6):468-73.
8. Bleeker SE, Moons KGM, Derksen-Lubsen G, Grobbee DE, Moll HA. Predicting serious bacterial infection in young children with fever without apparent source. Acta Paediatrica. 2001 Nov;90(11):1226-32.
9. Isaacman DJ, Shults J, Gross TK, Davis PH, Harper M. Predictors of bacteremia in febrile children 3 to 36 months of age. Pediatrics. 2000 Nov;106(5):977-82.
10. Bleeker SE, Moll HA, Steyerberg EW, Donders ART, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research:: A clinical example. Journal of Clinical Epidemiology. 2003;56(9):826-32.
11. Wijesinha A, Begg CB, Funkenstein HH, McNeil BJ. Methodology for the differential diagnosis of a complex data set. A case study using data from routine CT scan examinations. Med Decis Making. 1983;3(2):133-54.
12. Begg CB, Gray R. Calculation of Polychotomous Logistic Regression Parameters Using Individualized Regressions. Biometrika. 1984 Apr.;71(1):11-8.
13. Oostenbrink R, Moons KGM, Theunissen CCW, Derksen-Lubsen G, Grobbee DE, Moll HA. Signs of meningeal irritation at the emergency department: How often bacterial meningitis? [Article]. Pediatric Emergency Care June 2001;17(3):161-164. 2001.
14. Bleeker S, Derksen-Lubsen, G, Grobbee, DE, Donders, ART, Moons KGM, Moll, HA. External validation and update of a prediction rule for serious bacterial infection in young children referred with fever without apparent source. (Abstracts from the European Society for Medical Decision Making Meeting Rotterdam, the Netherlands, June 7-8, 2004). Med Decis Making. 2004;24 (4):423.
15. Sinaniotis CA. Community-acquired pneumonia: diagnosis and treatment. Pediatr Pulmonol Suppl. 1999;18:144-5.
16. Baraff LJ, Oslund SA, Schriger DL, Stephen ML. Probability of Bacterial-Infections in Febrile Infants Less Than 3 Months of Age - a Metaanalysis .1. Pediatric Infectious Disease Journal. 1992 Apr;11(4):257-65.
17. McCarthy PL, Lembo RM, Baron MA, Fink HD, Cicchetti DV. Predictive value of abnormal physical examination findings in ill- appearing and well-appearing febrile children. Pediatrics. 1985 August 1, 1985;76(2):167-71.
18. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001 Aug;54(8):774-81.

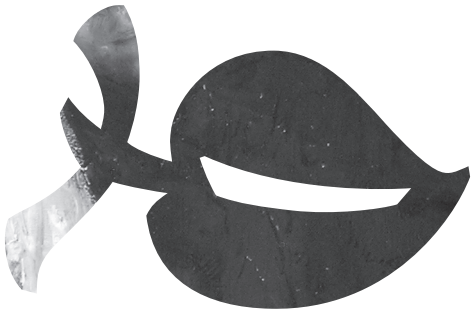
19. Kuppermann N, Fleisher GR, Jaffe DM. Predictors of occult pneumococcal bacteremia in young febrile children. Annals of Emergency Medicine. 1998 Jun;31(6):679-87.
20. Baraff LJ, Lee SI. Fever without Source - Management of Children 3 Months to 36 Months of Age. Pediatric Infectious Disease Journal. 1992 Feb;11(2):146-51.
21. Bonsu BK, Harper MB. Identifying febrile young infants with bacteremia: Is the peripheral white blood cell count an accurate screen? Annals of Emergency Medicine. 2003 Aug;42(2):216-25.
22. Jaffe DM, Fleisher GR. Temperature and Total White Blood-Cell Count as Indicators of Bacteremia. Pediatrics. 1991 May;87(5):670-4.
23. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KGM. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. Journal of Clinical Epidemiology. 2003;56(5):441-7.

Chapter 7

Validity of the Manchester Triage System in Pediatric Emergency Care

Jolt Roukema, Ewout W. Steyerberg, Alfred van Meurs, Madelon Ruige,
Johan van der Lei, Henriette A. Moll

Accepted for publication, Emergency Medicine Journal 2006



Abstract

Objective To assess the validity of the Manchester Triage System (MTS) in paediatric emergency care, using information on vital signs, resource utilization and hospitalization.

Methods Patients were eligible if they attended the emergency department of a large inner-city hospital in The Netherlands, from August 2003 to November 2004 and were under 16 years of age. A representative sample of 1,065 patients was drawn from 18,469 eligible patients. The originally assigned MTS urgency level was compared with resource utilization, hospitalization and a predefined reference classification for true urgency, based on vital signs, resource utilization, and follow-up. Sensitivity, specificity and percentage over- and undertriage of the MTS were calculated.

Results The number of patients who used more than 2 resources increased with a higher level of MTS urgency. The percentage of hospital admissions increased with the increase in level of urgency, from 1% in the non-urgent patients to 54% in emergent patients. According to the reference classification, the sensitivity of the MTS to detect emergent/very-urgent cases was 63%, the specificity was 78%. Undertriage occurred in 15%, of which 96% by one urgency category lower than the reference classification. Overtriage occurred in 40%, mostly in lower MTS categories. In 36% of these cases, the MTS classified two or more urgency categories higher than the reference classification.

Conclusions The MTS has moderate sensitivity and specificity in paediatric emergency care. Specific modifications of the MTS should be considered for paediatric emergency care to reduce overtriage, whilst maintaining sensitivity in the highest urgency categories.

Introduction

Hospital Emergency Departments (EDs) are increasingly visited by patients with non-urgent problems (1-3). This leads to overcrowded waiting rooms and long waiting times. As a consequence, patients urgently needing care may not be treated in time, whereas patients with non-urgent problems may unnecessarily receive expensive emergency care. Therefore, a reliable, valid triage system is required for patient safety. Triage of paediatric patients is difficult as presenting signs and symptoms and final diagnoses differ from adult patients (4). Several triage systems have been developed to categorize patients by 'urgency of care' (5-12). The Manchester Triage System (MTS), used by emergency-department nurses, is a triage system that supports the determination of a patient's urgency level on the basis of discriminators embedded in problem-specific flow-charts (9, 13). The triage-nurse selects the most suitable flow-chart for each presenting problem and uses general and specific discriminators to identify the patient's acuity. The MTS provides clarity about maximum allowed waiting time for the different levels of urgency: 'emergent' (red) needs instantaneous evaluation, 'very urgent' (orange) needs evaluation within 10 minutes, 'urgent' (yellow) within 60 minutes, 'standard' (green) within 120 minutes and 'non-urgent' (blue) problems are admitted to wait for up to 240 minutes.

The MTS has predominantly been implemented throughout Europe (e.g., United Kingdom, Ireland, Portugal, and The Netherlands). It is clear that assigning an inappropriate low urgency level may lead to possibly dangerous delays in patient care. However, assigning an inappropriate high triage level may increase waiting time for the truly urgent cases (14).

The MTS was developed by the Manchester Triage Group and is based on expert opinion. The scientific validity of the MTS is based on three key-articles that are focused on high-risk, adult patients. In a literature review Zimmerman concludes that the MTS is reliable and valid (15). A MTS sensitivity of 87% was found for identifying chest pain of cardiac origin, and the MTS was found to be a sensitive tool for identifying critically ill adult patients (16, 17). However, the MTS validity for standard and non-urgent problems has not been evaluated. Furthermore, no specific evaluation of the use of the MTS for paediatric patients has been made. As a result, only anecdotal information is available on the use of the MTS in paediatric emergency care. Especially overtriage, assigning an inappropriate high triage level, seems to be a problem. For example, in the MTS, most children presenting with fever require medical assessment within 10 minutes. However, in a hospital ED, many children present with fever, and assessment of these children within 10 minutes is neither feasible

nor necessary in clinical practice (18).

The objective of this study is to assess the validity of the MTS in paediatric emergency care, using information on vital signs, resource utilization and hospitalization.

Methods

Study Design

This study is a retrospective observational study of the validity of the Manchester Triage System (MTS) in paediatric emergency care. The validity of the MTS was assessed by correlating MTS urgency categories to resource utilization, hospitalization and to a predefined reference classification for true urgency. The requirement for informed consent was waived by the Institutional Review Board.

Study Setting and Population

Children were included from the Haga hospital, Juliana Children's Hospital site, a large inner-city teaching hospital with a mixed paediatric and adult emergency department (ED) in The Hague, The Netherlands. Annually, approximately 30,000 patients attend the ED, of which nearly 15,000 are children.

Study Protocol and Measurements

The standardized medical records of 1,065 children, who visited the ED of the Haga Hospital in the period from August 2003 until November 2004, were reviewed. Selection of the records was based on the originally assigned MTS urgency category. All retrievable records of patients assigned to 'emergent' or 'non-urgent' were selected, and a random sample was drawn of the other three categories; 'very urgent', 'urgent', and 'standard'. The random sample (an approximate percentage of the original group size) was taken in each of the three categories, using the Statistical Package for the Social Sciences (SPSS) version 12.0.1. Patients whose urgency category according to the MTS was manually overruled by ED-nurses were excluded (2.4%).

To assess the validity of the MTS, a reference classification for true urgency is necessary. First we assess the correlation between MTS-assigned urgency category and resource utilization or hospitalization as overall indicators for urgency (5, 7, 19). Resource utilization was defined, according to others, as having simple laboratory, extensive laboratory or radiology tests (radiography, ultrasound, CT/MRI) performed, or having received medication or an intervention at the ED (see Appendix 1 for detailed definitions) (5).

Subsequently, in an expert meeting a reference classification as proxy for true urgency was de-

fined. The reference classification was based on a combination of objective information on vital signs, presence of a possible life-threatening condition, resource utilization (diagnostic work-up and therapy), and follow-up (hospitalization, ambulatory care or discharge). For each patient, these data were collected from the standardized ED form, medical record, and the computer-based hospital information system (see Appendix 1 for collected data). Data were collected by trained medical students using standardized data collection forms, and was blinded for originally assigned MTS urgency category. All patients were assigned an urgency category according to the predefined reference classification based on consensus between paediatricians and paediatric surgeons. The classification is hierarchical, i.e., when vital signs are outside the PRISM III – normal range, the patient is assigned to *emergent*; when vital signs are normal, the presence of a possible life-threatening condition discriminates *very urgent* patients; if a possible life-threatening condition is absent, the extensiveness of diagnostic tests, therapy and follow-up discriminates between *urgent* and *standard* patients; finally, *non-urgent* patients were defined as patients who had no diagnostic tests performed, did not have therapy (at the ED), and who were discharged without follow-up. A classification matrix and detailed definitions of all reference classifications are shown in Appendix 2.

Data Analysis

First we compared MTS level of urgency with resource utilization or hospitalization. Next, the MTS urgency classification was compared with the reference classification. The validity of the MTS is expressed as percentage agreement between the reference classification and the MTS. Undertriage is defined as an MTS classification at least one urgency level lower than the reference classification. Overtriage is defined as an MTS classification of at least one urgency level higher than the reference classification. SPSS version 12.0.1 was used for statistical analyses with weighting by the inverse of the sampling ratio.

Results

A total of 18,469 patients, 0-16 years old, visited the ED during the study period from August 2003 until November 2004. A total of 1,065 patients were included in this study. Stratified for urgency category, no significant differences were found in the distribution of gender, age, specialty, and referral profile between the original dataset and the study sample (Table 1).

Table 1. Baseline characteristics

	Original (n=18,469)	**Sample (n=1,065)**
Male gender*	10,575 (57.3)	588 (55.2)
Age (years)†	4.9 (4.3)	4.6 (4.3)
Specialty		
paediatrics	10,648 (57.7)	725 (68.1)
paediatric surgery	6,792 (36.8)	303 (28.5)
Other‡	1,029 (5.5)	37 (3.5)
*Referral**		
self-referred	11,216 (60.7)	588 (55.2)
general practitioner	3,985 (21.6)	230 (21.6)
paediatrician/paediatric surgeon	1,886 (10.2)	90 (8.4)
ambulance	977 (5.3)	60 (5.6)
other/unknown**	405 (2.2)	97 (9.1)
*MTS category**		
emergent	152 (0.8)	127
very urgent	3,638(19.7)	276
urgent	4,414 (23.9)	271
standard	7,535 (40.8)	284
non-urgent	148 (0.8)	107
missing	2142 (11.6)	n.a.
overruled	440 (2.4)	n.a.

* Absolute number (percentage).
† Mean (standard deviation).
‡ Paediatric subspecialties, e.g., neurology, gastro-enterology.
** Other hospital, other department than paediatrics or paediatric surgery, unknown.

Table 2 gives an overview of resource utilization stratified for level of urgency according to the MTS. The number of patients who used more than 2 resources increased with a higher level of (MTS) urgency. In the lowest urgency category, more patients used no resources (68%) than in the highest urgency category (13%). In the non-urgent category, 28% of the patients used 1 resource, and in the standard category 42% used at least 1 resource.

Table 2. Resource utilization, per MTS urgency category.

	Number of resources used*			
MTS classification	**None**	**1**	**2**	**>2**
Emergent (n=127)	12.6	45.7	29.1	12.6
Very urgent (n=276)	29.3	45.3	20.3	5.1
Urgent (n=271)	26.9	42.8	28.4	1.8
Standard (n=284)	41.5	41.9	16.2	0.4
Non-urgent (n=107)	68.2	28.0	3.7	0.0

* Numbers represent percentages, unless stated otherwise.

Table 3 shows the follow-up after the ED visit. Hospitalization is correlated to MTS urgency: the percentage hospital admission increased with higher level of urgency, from 1% in the non-urgent patients to 54% in emergent patients.

Table 3. Hospitalization, per MTS urgency category.

MTS classification (n)	Hospitalization*
Emergent (127)	53.5
Very urgent (276)	28.6
Urgent (271)	16.2
Standard (284)	6.0
Non-urgent (107)	0.9

* Numbers represent percentages.

In 65 patients the reference classification could not be determined, as essential information was missing. Hence, a comparison between reference and MTS classification could be made in 1,000 patients (94%). Table 4 shows the agreement between the MTS classification and the reference classification. The 1,000 cases were weighted with the inverse of the sampling ratio. In 45% of all cases,

patients had exactly the same priority in the MTS and reference classification (0.7+1.6+11.5+31.0+0.2%). In 40% of all cases the assigned MTS urgency category was higher than the reference classification (i.e., overtriage). In 15% of all cases the MTS urgency category was lower than the reference classification (i.e., undertriage). In 96% of these cases, the MTS urgency category was one urgency category lower than the reference classification.

The sensitivity of the MTS to detect emergent/very urgent patients was 63% (calculated as (0.7+1.3+0.1+1.6)/(2.0+3.9)), the specificity was 78%.

Table 4. Comparison between MTS and reference classification, based on 1,000 cases.

	Reference classification					
MTS classification	Emergent	Very urgent	Urgent	Standard	Non-urgent	*Total**
Emergent	**0.7**	0.1	0.1	0.1	0	1.0
Very urgent	1.3	**1.6**	8.0	11.8	0.9	23.6
Urgent	0	1.6	**11.5**	13.5	1.2	27.8
Standard	0	0.5	11.1	**31.0**	4.1	46.7
Non-urgent	0	0	0	0.6	**0.2**	0.9
*Total**	2.0	3.9	30.7	57.0	6.4	100

Numbers are percentages, weighted for distribution in the total ED-population.
* Differences in column and row totals may have occurred due to rounding.
□ = Undertriage.
■ = Overtriage.

Discussion

The objective of this study was to assess the validity of the Manchester Triage System (MTS) in paediatric emergency care. A high level of urgency according to the MTS was correlated with the utilization of two or more resources and with a high percentage of hospitalization. However, in the lower urgency categories, 28% of the non-urgent and 42% of the standard patients used at least one resource. Hospitalization was correlated with MTS urgency. Using a predefined reference classification including refined indicators for true urgency, we found that the MTS was neither very sensitive nor very specific in a paediatric population.

In 15% of all cases, the urgency category as assigned by the MTS was lower than the reference classification. We found a sensitivity of 63% and a specificity of 78% for detecting emergent/very urgent cases. Yet, as emergent cases according to the reference classification were always classified as emergent or very urgent in the MTS, the system seems to be reasonably safe in a paediatric population. In only 0.6% of all cases, the MTS urgency category was two categories lower than the reference classification. According to the reference classification, 14% of the very urgent patients (i.e., having a potential life-threatening condition) were classified as standard by the MTS. This applied mainly to children who had an apparently life-threatening event (ALTE). Although most of these children do not present as having an urgent condition, the patient history requires the attending physician to exclude a serious illness as cause of the ALTE (20).

Overtriage by the MTS seemed to be a problem in the paediatric population. Of all the paediatric patients, 40% were assigned an urgency category in the MTS that was too high according to the reference classification. A difference of at least two urgency categories between the MTS and the reference classification was found in 14% of all patients. These were, for example, patients who presented with fever, required no diagnostic work-up, received the advice to use antipyretic medication and were followed up by telephone. In the MTS these patients were classified as very urgent whereas in the reference classification they were rated as standard.

In the adult population, MTS performance was only assessed for urgent cases or specific conditions. An MTS sensitivity of 87% was found for identifying chest pain of cardiac origin (17). Furthermore, it was concluded that the MTS is a sensitive tool for identifying critically ill patients (16). Overtriage has not been studied, though mixing non-urgent conditions with urgent conditions can result in delayed care for true urgent cases. The validity of the MTS in a paediatric population has

not been discussed in current literature. A paediatric version of the Canadian Triage and Acuity Scale CTAS was developed (Paed-CTAS) and paediatric adjustments were made in the ESI (v.4) (8, 12, 21). These examples underline that triage of the paediatric patient differs from adult triage. Furthermore, the *reliability* of triage systems, measured as interrater agreement, is mostly assessed rather than *validity* (5, 10, 22-25). Furthermore, in triage systems the *reliability* of the system, measured as interrater agreement, rather than the *validity* is assessed.

It is difficult to evaluate the validity of a triage system. Resource utilization or hospitalization have been shown to be correlated to level of urgency in paediatric patients, and have previously been used to validate the ESI (5, 7, 19). In our study, resource utilization was only moderately correlated to level of urgency when more than 2 resources were used, or when no resources were used. In the patients who used 1 or 2 resources, no correlation was found. Hospitalization was correlated with MTS level of urgency, but appeared to be a crude, dichotomous measure for true urgency. The reference classification used in this study was therefore based on a combination of objective data on vital signs, presence of a possible life-threatening condition, resource-utilization (diagnostic tests and therapy), and follow-up (hospitalization, ambulatory care, discharge).

In the MTS, however, the decision made at triage is based on a rapid assessment rather than a diagnosis, as the urgency of a condition must be assessed at presentation (9). The MTS is advocated to be a dynamic system; therefore, it could be an advantage that, in the reference classification, improvement and worsening of the clinical condition are taken into account. However, it is a disadvantage that these tests and decisions do not necessarily reflect the child's condition at presentation. Windle argues that, in the absence of a gold standard for assessing urgency at presentation, it is not possible to prove that any triage system works (26). From a decision-analytic point of view, urgency should be defined by the degree of deterioration in outcome that is caused by delay in seeing a physician and the delay in initiating diagnostic work-up and treatment. A true emergent situation has a steep slope of deterioration regarding the patient's outcome with every minute of delay. This slope is much less steep for truly less urgent conditions. So, urgency requires judgment of the natural course of a condition, which is hard to obtain empirically. Therefore, to date most studies used either expert panels or resource utilization as proxies for true urgency (5, 19, 25, 27, 28).

In our sample, exploration of the effects of a modification of the MTS based on patient's age did not improve MTS performance. Flowchart-specific modifications seem to be necessary to improve

the MTS for paediatric emergency care.

This study has some limitations. First, we have attempted to include all patients originally assigned to the 'emergent' and 'non-urgent' MTS categories. However, not all medical records of these patients were retrievable (25 emergent cases and 41 non-urgent cases). Second, the reference classification was conservative, especially regarding the definition of emergent cases. All patients with a vital sign outside the normal range according to the PRISM III were classified as emergent (29). This may have resulted in a relatively high number of patients classified as emergent according to the reference classification, and subsequently to a low sensitivity of the MTS. Third, local variation in case-mix and medical practice regarding diagnostic and therapeutic interventions may have influenced the results.

In conclusion, the MTS has moderate sensitivity and specificity in paediatric emergency care. A substantial proportion of seriously ill patients will be recognized, but overtriage is a problem. Specific modifications of the MTS should be considered for paediatric emergency care to reduce overtriage, whilst maintaining sensitivity in the highest urgency categories.

References

1. Emergency Medical Care. Little cause for optimism. Report: Dutch Medical Inspectorate of Health; 2004 September 2004.
2. McCaig LF, Burt CW. National Hospital Ambulatory Medical Care Survey: 2002 emergency department summary. Adv Data. 2004 Mar 18(340):1-34.
3. Zimmer KP, Walker A, Minkovitz CS. Epidemiology of Pediatric Emergency Department Use at an Urban Medical Center. Pediatr Emerg Care. 2005 Feb;21(2):84-9.
4. Cain P, Waldrop RD, Jones J. Improved pediatric patient flow in a general emergency department by altering triage criteria. Acad Emerg Med. 1996 January 1, 1996;3(1):65-71.
5. Baumann MR, Strout TD. Evaluation of the Emergency Severity Index (version 3) triage algorithm in pediatric patients. Acad Emerg Med. 2005 Mar;12(3):219-24.
6. Beveridge R, Ducharme J, Janes L, Beaulieu S, Walter S. Reliability of the Canadian Emergency Department Triage and Acuity Scale: Interrater agreement. Annals of Emergency Medicine. 1999;34(2):155-9.
7. Eitel DR, Travers DA, Rosenau AM, Gilboy N, Wuerz RC. The emergency severity index triage algorithm version 2 is reliable and valid. Acad Emerg Med. 2003 Oct;10(10):1070-80.
8. Gilboy N, Tanabe P, Travers DA. The Emergency Severity Index Version 4: changes to ESI level 1 and pediatric fever criteria. J Emerg Nurs. 2005 Aug;31(4):357-62.
9. Mackway-Jones K. Emergency Triage: Manchester Triage Group: BMJ Publishing Group; 1997.
10. Travers DA, Waller AE, Bowling JM, Flowers D, Tintinalli J. Five-level triage system more effective than three-level in tertiary emergency department. J Emerg Nurs. 2002 Oct;28(5):395-400.
11. Wuerz RC, Travers D, Gilboy N, Eitel DR, Rosenau A, Yazhari R. Implementation and Refinement of the Emergency Severity Index. Acad Emerg Med. 2001 February 1, 2001;8(2):170-6.
12. Canadian Association of Emergency Physicians. Canadian Paediatric Triage and Acuity Scale: Implementation Guidelines for Emergency Departments. Can J Emerg Med. 2001;3(4):1-32.
13. Windle J. The extent to which the; environment, triage event, documentation, components of assessment and training & development affect departmental accuracy when using the Manchester Triage System [Dissertation]: University of Sheffield 2001; 2001.
14. Maldonado T, Avner JR. Triage of the Pediatric Patient in the Emergency Department: Are We All in Agreement? Pediatrics. 2004 August 1, 2004;114(2):356-60.
15. Zimmermann PG. The case for a universal, valid, reliable 5-tier triage acuity scale for US emergency departments. Journal of Emergency Nursing. 2001 2001/6;27(3):246-54.
16. Cooke MW, Jinks S. Does the Manchester triage system detect the critically ill? J Accid Emerg Med. 1999 May;16(3):179-81.
17. Speake D, Teece S, Mackway-Jones K. Detecting high-risk patients with chest pain. Emerg Nurse. 2003 Sep;11(5):19-21.
18. Scoble M. Implementing triage in a children's assessment unit. Nursing Standard. 2004.
19. Wuerz RC, Milne LW, Eitel DR, Travers D, Gilboy N. Reliability and Validity of a New Five-level Triage Instrument. Acad Emerg Med. 2000 March 1, 2000;7(3):236-42.
20. Brand DA, Altman RL, Purtill K, Edwards KS. Yield of diagnostic testing in infants who have had an apparent life-threatening event. Pediatrics. 2005 Apr;115(4):885-93.
21. Gouin S, Gravel J, Amre DK, Bergeron S. Evaluation of the Paediatric Canadian Triage and Acuity Scale in a pediatric ED.

Am J Emerg Med. 2005 May;23(3):243-7.

22. Wuerz R, Fernandes CM, Alarcon J. Inconsistency of emergency department triage. Emergency Department Operations Research Working Group. Ann Emerg Med. 1998 Oct;32(4):431-5.
23. Fernandes CM, Wuerz R, Clark S, Djurdjev O. How reliable is emergency department triage? Ann Emerg Med. 1999 Aug;34(2):141-7.
24. Brillman JC, Doezema D, Tandberg D, Sklar DP, Davis KD, Simms S, et al. Triage: limitations in predicting need for emergent care and hospital admission. Ann Emerg Med. 1996 Apr;27(4):493-500.
25. Dong SL, Bullard MJ, Meurer DP, Colman I, Blitz S, Holroyd BR, et al. Emergency triage: comparing a novel computer triage program with standard triage. Acad Emerg Med. 2005 Jun;12(6):502-7.
26. Windle J, Mackway-Jones K. Don't throw triage out with the bathwater. Emerg Med J. 2003 March 1, 2003;20(2):119-20.
27. Tanabe P, Gimbel R, Yarnold PR, Kyriacou DN, Adams JG. Reliability and Validity of Scores on the Emergency Severity Index Version 3. Acad Emerg Med. 2004 January 1, 2004;11(1):59-65.
28. Lee A, Hazlett CB, Chow S, Lau F-l, Kam C-w, Wong P, et al. How to minimize inappropriate utilization of Accident and Emergency Departments: improve the validity of classifying the general practice cases amongst the A&E attendees. Health Policy. 2003 2003/11;66(2):159-68.
29. Pollack MM, Patel KM, Ruttimann UE. The Pediatric Risk of Mortality III--Acute Physiology Score (PRISM III-APS): a method of assessing physiologic instability for pediatric intensive care unit patients. J Pediatr. 1997 Oct;131(4):575-81.

Appendix 1 Reference classification parameters

Vital signs: normal values according to PRISM III (29)

Age	Respiratory rate / min	Systolic BP (mmHg)	Heart rate / min
<1 month	15-90	55-160	80-215
1-12 months	10-70	65-160	60-215
1-12 year		75-200	45-185
>12 year		85-200	40-145

heart rhythm: arrhythmia.
respiration pattern: inspiratory stridor, respiratory insufficiency.
temperature: ≤33°C of >41°C.
oxygen saturation: absolute percentage, cut-off = < 90%.
level of consciousness: decreased, convulsive at arrival, coma.

*Presence of a possible life-threatening condition (***PLC***):*

Meningitis, sepsis, high-energy trauma, substantial external blood loss or trauma (sharp/blunt) leading to substantial blood loss, aorta dissection, ≥10% dehydration, (near)-drowning, electric trauma, apparently life-threatening event (ALTE), possible dangerous intoxication, ≥10% burns, facial burns or possible inhalation trauma, other (specified).

Diagnostic work-up :
- **Simple** laboratory tests (CBC, electrolytes, liver enzymes, renal function, urine/stool cultures, nasal swabs) **.
- **Imaging** (radiograph, ultrasound imaging) *.
- **Extensive** laboratory tests (blood culture, CSF-puncture) or CT/MRI *.

Therapy :
- **Rx:** simple advice or medication on prescription
- **Rx at the ED:** oral medication at the ED (i.e., ORS, prednisone, antibiotics) or small surgical intervention (suture, debridement, bandage) *.
- **Intervention**: Intravenous medication or intervention at the ED (including fluids, aerosoles) or surgical intervention (including casting, gastrogavage, inguinal hernia reposition, luxation reposition) *.
- Other (specified) *.

Follow-up :
- **General practitioner** / telephone contact
- **Outpatient**/ED
- **Hospital admission**

* Each item was counted as one *resource*

** One or more items were counted as one *resource* (5)

Appendix 2 Reference classification matrix and definitions of reference urgency categories.

			Diagnostics			Therapy			Follow-up		
	Vital	**PLC**	***Simple***	***Imaging***	***Extensive***	***Rx***	***Rx at ED***	***Intervention***	***Tel./GP***	***Outpatient***	***Hospitalization***
Emergent	**1**	n/a	n/a			n/a			n/a		
Very urgent	0	**1**	n/a			n/a			n/a		
Urgent	0	0	n/a			0	0	**1**	n/a		
	0	0	n/a			0	**1**	0	n/a	0	**1**
	0	0	**1**	0	0	n/a			n/a	0	**1**
	0	0	0	**1**	0	n/a			n/a	0	**1**
	0	0	0	0	**1**	n/a			n/a	0	**1**
	0	0	0	**1**	0	**1**	0	0	n/a	**1**	0
	0	0	0	0	**1**	**1**	0	0	n/a	**1**	0
	0	0	0	**1**	0	0	**1**	0	n/a		
	0	0	0	0	**1**	0	**1**	0	n/a		
	0	0	0/**1**	**1**	**1**	n/a			n/a		
Standard	**All other combinations**										
Non-urgent	0	0	0	0	0	0/**1**	0	0	0	0	0

1 = present / 0= absent; PLC = possible life-threatening condition; n.a. = not applicable;
Rx = medication on prescription.

Emergent :
Vital parameters: ab normal

Very urgent
Possible life-threatening condition: present

Urgent :
One of the following combinations:
- intervention at the ED, diagnostic work-up and follow-up not applicable.
- extended laboratory diagnostics AND X-ray/ultrasound imaging, intervention.
- extended laboratory diagnostics or X-ray/ultrasound imaging AND oral medication or small surgical intervention at the ED. Extended laboratory diagnostics or X-ray/ultrasound imaging AND medication on prescription, AND outpatient/ED follow-up within 48 hrs.
- hospital admission AND some diagnostic work-up, Rx at ED, or intervention.

Standard :
All patients that were not classified as urgent or non-urgent.

Non-urgent :
Diagnostic work-up: none.
Therapy: none/medication on prescription.
Follow-up: none.

Chapter 8

Randomized Trial of a Clinical Decision Support System: Unexpected impact on the management of children with fever without apparent source

Jolt Roukema, Ewout W. Steyerberg, Johan van der Lei, Henriette A. Moll

Submitted

Abstract

Objective To assess compliance with a clinical decision support system (CDSS) for diagnostic management of children with fever without apparent source and to study the effects of application of the CDSS on time spent in the emergency department (ED) and number of laboratory tests.

Design The CDSS was used by ED nursing-staff to register children presenting with fever. The CDSS identified children that met inclusion criteria (1-36 months and fever without apparent source (FWS)) and provided patient-specific diagnostic management advice. Children at high risk for serious bacterial infection were randomized for the 'intervention' (n=74) or the 'control' (n=90) group. In the intervention group, the CDSS provided the advice to immediately order laboratory tests and in the control group the ED-physician first assessed the children and then decided on ordering laboratory tests.

Results Compliance with registration of febrile children was 50% (683/1,399). Adherence to the advice to order laboratory tests was 82% (61/74). Children in the intervention group had a median (25^{th}-75^{th} percentile) length of stay at the ED of 138 (104-181) minutes. The median length of stay at the ED in the control group was 123 (83-179) minutes. Laboratory tests were significantly more frequently ordered in the intervention group (82%) than in the control group (44%, $p<0.001$, χ^2 test).

Conclusion Implementation of a CDSS for diagnostic management of young children with fever without apparent source was successful regarding compliance and adherence to CDSS recommendations, but had unexpected effects on patient outcome in terms of ED length of stay and number of laboratory tests.

Introduction

For several diagnostic and therapeutic problems, guidelines or clinical prediction rules were developed (1-7). Implementation of guidelines may result in reduced diagnostic testing, improved documentation, more appropriate treatment and a reduction of the time spent in the ED (2, 3). However, the translation of guidelines and prediction rules into clinical practice is still a major challenge (8). Clinical decision support systems (CDSS) may be able to integrate available knowledge (e.g., guidelines) with clinical practice (9). Key features for successful CDSS implementation have been described, but little is known about the effects of CDSS-utilization on patient outcomes (10-12).

The management of young febrile children is an everyday challenge for emergency department (ED) physicians. Distinguishing children with mild viral disease from those with serious bacterial infection (SBI) is difficult as clinical presentation is often a-specific (13, 14). Early identification of children at risk for SBI could support appropriate management in terms of diagnostic and therapeutic decisions. At the Erasmus Medical Center in Rotterdam, The Netherlands, a CDSS was developed for the diagnostic management of young children with fever without apparent source (FWS). From July 2003, the computerized CDSS was used routinely by the ED nursing-staff to register children presenting with fever. In each individual case, the CDSS automatically identified children with FWS and provided a patient-specific diagnostic management advice.

Our aims were to assess compliance with the system and to assess the effects of application of the CDSS on time spent at the ED and amount of performed diagnostic tests in children with FWS. Our hypothesis was that initiation of diagnostic work-up directly after ED-nurse evaluation increased ED efficiency by reducing both time spent in the ED and number of diagnostic tests (3).

Methods

Study population and setting

From July 1, 2003 until December 31, 2005, children presenting with fever at the emergency department (ED) of the Sophia Children's Hospital in Rotterdam were routinely registered by ED-nurses, using standardized ED-forms. Additionally, ED-nurses used OpenSDE; a structured data entry application that, for the purpose of this study, was tailored for data collection on febrile children (see figure 1) (15). Patient characteristics (gender, age, reason of the visit, visit date), referral profile, duration of the febrile episode, symptoms (earache, sore throat, runny nose, coughing, vomiting, diarrhea), and observations and measures from physical examination (e.g., vital signs, temperature, presence of chest-wall retractions, clinical appearance, and meningeal irritation) were collected during or immediately after ED-nurse evaluation, as the patient was waiting for the attending ED-physician.

Patients with chronic co-morbidity (e.g., malignancies, cystic fibrosis, severe psychomotor retardation) were excluded. Furthermore, we excluded acutely ill patients since they were treated immediately. This study was approved by the Institutional Review Board of the Erasmus Medical Center: Informed consent was obtained (94%, n=164) after discharge from the ED as an informed consent procedure prior to the intervention would have interfered with the outcome measures (see *D. Measurements*), and because the intervention had no risks for the patient.

Prediction rules

Two prediction rules were previously developed for children with fever without apparent source (16-18). The prediction rule generated patient-specific risk-scores for SBI and was developed using data of 381 patients between 1 and 36 months of age, who presented to the ED of either the Sophia Children's Hospital in Rotterdam (1996-1998, 2000-2001) or the Juliana Children's Hospital in The Hague (1998, 2000-2001) with fever without apparent source (FWS). FWS was defined as a body temperature of at least 38.0 degrees Celsius, for which no clear source was identified after evaluation by the GP or after the history was taken by the pediatrician. A 'clinical model' and a 'clinical + lab model' were derived with AUCs of 0.69 and 0.86, respectively. See Appendix 1 for predictor variables. For self-referred patients, a separate prediction rule was developed based on data of 109 self-referred patients attending the ED of either the Sophia Children's Hospital in

Rotterdam (1996-1998) or the Juliana Children's Hospital in The Hague (1998), with FWS. A 'clinical' and 'clinical + lab' model for self-referred patients were derived, with AUCs of 0.70 and 0.81, respectively (18). See Appendix 2 for predictor variables.

Methods

A clinical decision support system (CDSS) for the diagnostic management of children attending the ED with fever without apparent source (FWS) was previously developed and implemented at the ED from July 1, 2003, two months prior to the start of the study. In this period, all ED-nurses received standardized training in how to use the CDSS. Registration of the data took 1-2 minutes per patient and did not alter workflow significantly.

FWS was defined as body temperature ≥38.0 degrees Celsius, and no apparent source found after evaluation by the ED-nurse (16). The following symptom or combinations of symptoms were considered an apparent source of the fever: neck stiffness, two or more specified upper-airway symptoms (earache, sore throat, rhinitis), coughing and at least one upper-airway symptom, or the combination of vomiting and diarrhea. All others were classified as FWS, and were automatically identified based on collected data. In each individual case, the CDSS calculated a clinical risk-score for SBI (see Appendices 1 and 2 for score-charts). Based on a prior defined cut-off point, children were classified according to the likelihood of having either a low or a high risk for SBI (see Appendices 1 and 2) (16). When data items necessary to calculate the risk-score (i.e., predictors for the presence of SBI) were missing, the CDSS provided a reminder during patient data registration.

Children with a *high* risk-score were eligible for early initiation of diagnostic work-up: all children with a high risk-score were randomized to determine whether the advice 'order laboratory tests' was shown to the user (see figure 1).

Patient selection and randomization is graphically shown in figure 2. Randomization was based on a computer algorithm, sampling a number between 1 and 1000 in each case. Patients were assigned to the 'intervention group' (i.e., order laboratory tests) when an even number and to the 'control group' when an odd number was sampled. The laboratory tests to be ordered were based on prior consensus and consisted of complete blood count (CBC) and C-reactive protein (CRP) (16). All patients were evaluated by ED physicians.

Measurements

Based on the time of arrival at the ED and the time of departure from the ED as registered in the ED nursing-record, total ED-time was calculated as the difference between time of arrival and time of departure in minutes. Data on all performed laboratory tests and additional diagnostic tests (e.g., Roentgenograms) were collected for each patient from the computer-based hospital information system.

All final diagnoses were classified as either serious bacterial infection (SBI) or non-SBI. SBI was defined as culture or radiographically proven bacterial infection (e.g., meningitis, sepsis, bacteremia, pneumonia, urinary tract infection, bacterial gastroenteritis, osteomyelitis or ethmoiditis). Detailed descriptions of outcome diagnoses had been published earlier (16, 19). Assessment of the outcome measures was blinded for risk-scores and CDSS recommendations.

Statistical analysis

First, we calculated compliance with CDSS-registration of febrile patients as the percentage of all febrile patients who attended the ED with fever. Secondly, we assessed the adherence with CDSS recommendations by calculating the percentage of cases in which laboratory tests were ordered as advised.

In the children identified with FWS and having a high SBI risk-score, we compared the 'intervention group' to the 'control group' regarding total time spent at the ED and the frequency of diagnostic testing. A Mann-Whitney *U*-test was used to quantify the difference in total ED-time between the two groups, and a Chi-square test to assess the difference in frequency of diagnostic testing. Both an intention to treat analysis (all children in the intervention group analyzed) and per protocol analysis (only the children in the intervention group in whom laboratory tests were actually ordered analyzed) were performed. In a sub-analysis we only analyzed children who had laboratory tests ordered (intervention and control groups). Furthermore, the frequency of SBI was compared between the children with a low and children with a high risk-score, using the Chi-square test. The area under the receiver operating characteristic curve (AUC) of the prediction rule was calculated as measure of discriminative ability.

General characteristics, laboratory test results, risk-scores and incidence of SBI were compared between the intervention and the control group, using a Chi-square test for categorical and the Mann-Whitney *U*-test for continuous variables. A p-value lower than 0.05 was considered sta-

tistically significant. We used SPSS software (version 12.0, SPSS Inc, Chicago, IL) for the statistical analyses.

Results

From September 1, 2003 to December 31, 2005, 1,774 children, aged 1-36 months, attended the ED of the Sophia's Children's Hospital in Rotterdam with fever. 375 children were excluded because of chronic co-morbidity. Compliance with CDSS-registration of febrile children was 49%: 683 of the 1,399 eligible patients were registered. Of the 683 registered patients, 390 were identified with fever without apparent source (FWS). A total of 172 children had a high risk-score and were thus randomized (see figure 2). In 30 patients, total ED-time could not be calculated due to missing time of arrival at or departure from the ED (16 in the intervention group, 14 in the control group). Informed consent was obtained in 164 patients (95%).

Table 1. General characteristics of 172 children with fever without apparent source and a high risk-score.

Characteristics	**Randomization**	
	Intervention (n=74)	**Control (n=90)**
Patient history		
Male gender*	44 (59)	46 (51)
Age (years)†	1.0 (0.7-1.6)	0.9 (0.6-1.4)
Duration of fever (days)†	2.5 (1.0-4.0)	3.0 (1.8-6.0)
History of vomiting*	34 (46%)	46 (51%)
Physical examination		
Temperature (° Celsius)†	39.5 (39.0-40.0)	39.4 (38.9-40.0)
Ill clinical appearance*	53 (72)	61 (68)
Poor peripheral circulation*	5 (7)	13 (14)
Chest-wall retractions ± tachypnea*	10 (14)	7 (8)
Clinical risk-score		
Score†	11 (9-14)	11 (9-14)
Final diagnosis		
SBI*	10 (14)	16 (18)

* Absolute number (%); † Median (25th and 75th percentile).

Table 1 shows the characteristics of children with a high risk-score, stratified for intervention and control group. Both groups had similar characteristics.

In Table 2, the median time spent in the ED is shown for both the intervention and the control group. A difference in total ED-time was found between the intervention and the control group: children in the intervention group spent on average 20 minutes more in the ED, but the difference was not significant (p=0.16). Also in the per protocol analysis, a median difference of 20 minutes was not significant (p=0.06).

In a sub-analysis ('lab-tests ordered), we found that total ED-time for children in the intervention-group *with* laboratory tests performed (median 140 minutes) was shorter than for children in the control group *with* laboratory tests performed (160 minutes, p=0.43).

A significant difference in the number of laboratory tests was found between the intervention and the control group. In 82% (61 out of 74) of the cases in which the ED-nurse received the advice to order laboratory tests, the tests were actually ordered. In the control group, laboratory tests were ordered in 44% (n=40) of the patients, at the discretion of the ED-physician.

Table 2. Time spent in the ED for intervention and control group.

Analysis	**Total ED-time** (minutes)				**p-value**
	Intervention	**(n)**	**Control**	**(n)**	
Intention to treat	138 (104-181)	58	123 (83-179)	76	0.16
Per protocol	140 (116-184)	52	123 (83-179) †	76	0.06
Lab tests ordered	140 (116-184)	52	160 (115-213)	33	0.43

Numbers represent median (25th and 75th percentile).
† By definition identical to intention to treat value.

Table 3 shows a cross-tabulation of the presence of a SBI and the clinical risk-score. SBI was found slightly more often among the children with a high risk-score (15%) than among the children with a low risk-score (10%, p=0.10). Furthermore, the ability to predict SBI based on high or low risk-score was disappointing, as indicated by an AUC of 0.56 (95%CI 0.48-0.65).

Table 3. Presence of SBI

	Clinical score		
	Low score	*High score*	
SBI present	21 (10)	26 (15)	47
SBI absent	197 (90)	146 (85)	343
	218	172	390

Numbers represent absolute numbers (percentage within risk stratum).
p-value: χ^2 test = 0.10.

Discussion

In this study we implemented a clinical decision support system (CDSS) for the diagnostic management of young children with fever without apparent source, who are at risk for serious bacterial infection (SBI). Compliance with registration of febrile children in the CDSS was moderate, with 50% of the children being registered by ED-nursing staff. Adherence with the advice to order laboratory tests was good: in 82% of the cases in which the advice to order laboratory tests was given, the tests were actually ordered. Surprisingly, the children in whom laboratory tests were ordered immediately after nurse evaluation spent more time in the ED than the children in whom laboratory tests were ordered at the discretion of the attending physician.

In a recent review, four CDSS features were identified that were closely correlated with the ability of a CDSS to improve patient care (10). First, the decision support should be part of the routine workflow. Second, a computer system should be used to provide decision support. Third, an explicit patient-specific recommendation should be given rather than a probability, and fourth, decision support should be delivered at the time and location of decision making. In this study most of these features were included: children presenting with fever were registered in a computer system that automatically provided an actionable recommendation on diagnostic management at time and location of the nurse's decision. However, registration of the children was not yet part of the routine workflow and was performed voluntarily by ED-nursing staff. This might account for the fact that only 50% of the febrile children were registered. Furthermore, ED-crowding and time constraints may have accounted for a decrease in registration rate, as it was an extra task. Although we did not define a minimum compliance and adherence rate prior to the study, we conclude that implementation of the CDSS with regard to application of the system and adherence to the advice was successful.

In contrast to our expectations, the children in the intervention group spent more time in the ED than the children in the control group. The difference of 20 minutes in the 'per protocol analysis' was, however, statistically not significant. Explanations for the prolonged ED-time in the intervention group include the large difference in the frequency in which laboratory tests were ordered in the intervention (82%) and control group (44%). However, this difference also indicates that the amount of diagnostic tests significantly increased when the CDSS was used, which was in contrast

with the expected decrease. The sub-analysis in which only the children who had laboratory tests performed were compared, revealed that children whose tests had been ordered according to the CDSS recommendation had a median ED-time that was 20 minutes less than children whose tests had been ordered at the discretion of the physician (140 vs. 160 minutes). This indicates that the CDSS may potentially be effective, when children at high risk for SBI are more accurately identified. It has already been recognized that the effectiveness of a CDSS depends for a major part on the strength of the knowledge base that is used. Sim et al. stated that "a CDSS can only be as effective as the strength of the underlying evidencw better than that of the children who were not registered. We are, however, not able to compare these groups as information on whether the non-CDSS-registered children had FWS according to the ED-nurse, cannot be determined accurately in hindsight. Secondly, all children were evaluated by the attending physician. When laboratory tests were already ordered by the ED-nurse, the physician was automatically aware that the child had a high risk-score for SBI. This may have affected subsequent diagnostic testing.

Conclusion

Implementation of a CDSS for the diagnostic management of young children with fever without apparent source was successful regarding compliance and adherence to CDSS recommendations, but had unexpected effects on patient outcome in terms of ED length of stay and amount of performed laboratory tests. This study stresses the importance of including patient outcomes in the evaluation of clinical decision support systems.

Acknowledgements

We gratefully acknowledge the ED-nursing staff of the Sophia Children's Hospital for their participation and careful collection of all required data.

This study was supported by a grant from the Erasmus Medical Center Healthcare Efficiency research program (VAZ-doelmatigheids-onderzoek).

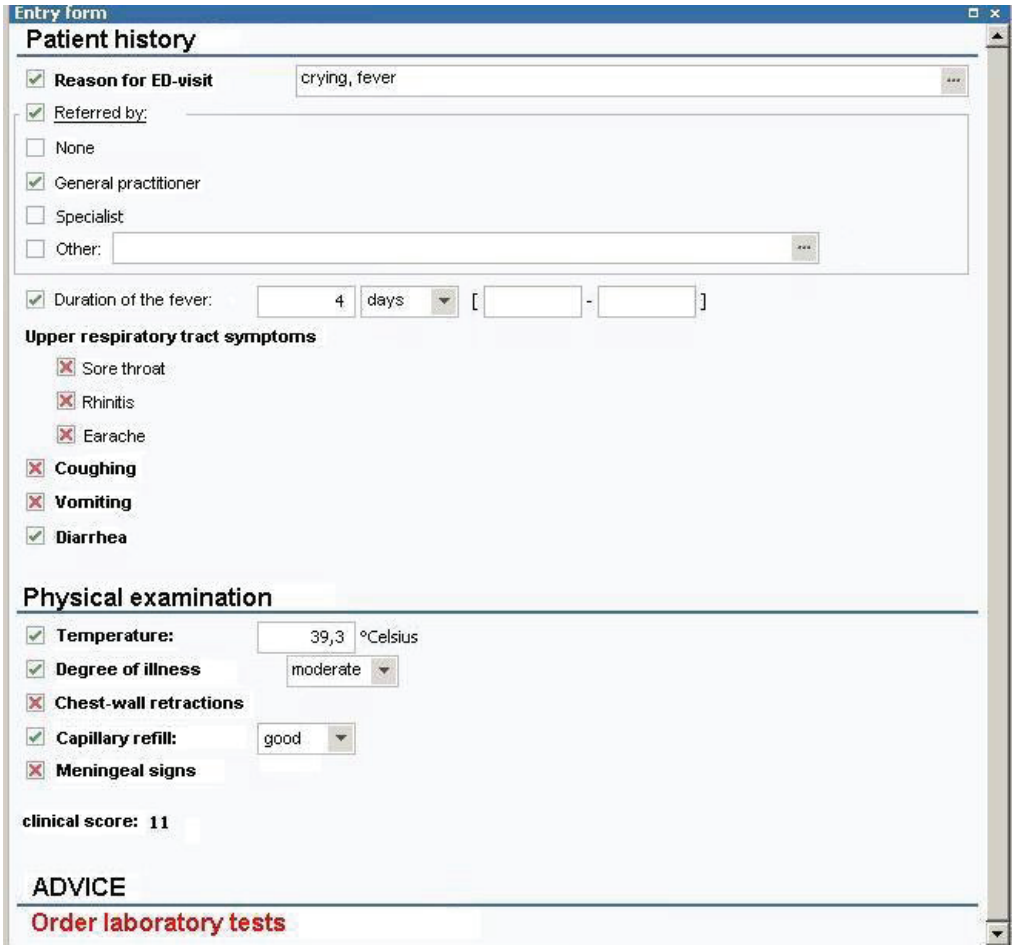
References

1. Baraff LJ. Management of fever without source in infants and children. Annals of Emergency Medicine. 2000 Dec;36(6):602-14.
2. Kucher N, Koo S, Quiroz R, Cooper JM, Paterno MD, Soukonnikov B, et al. Electronic Alerts to Prevent Venous Thromboembolism among Hospitalized Patients. N Engl J Med. 2005 March 10, 2005;352(10):969-77.
3. Armon K, MacFaul R, Hemingway P, Werneke U, Stephenson T. The impact of presenting problem based guidelines for children with medical problems in an accident and emergency department. Arch Dis Child. 2004 February 1, 2004;89(2):159-64.
4. Gorelick MH, Shaw KN. Clinical decision rule to identify febrile young girls at risk for urinary tract infection. Arch Pediatr Adolesc Med. 2000 Apr;154(4):386-90.
5. Haydel MJ, Shembekar AD. Prediction of intracranial injury in children aged five years and older with loss of consciousness after minor head injury due to nontrivial mechanisms. Ann Emerg Med. 2003 Oct;42(4):507-14.
6. Holmes JF, Sokolove PE, Brant WE, Kuppermann N. A clinical decision rule for identifying children with thoracic injuries after blunt torso trauma. Ann Emerg Med. 2002 May;39(5):492-9.
7. Lynch T, Platt R, Gouin S, Larson C, Patenaude Y. Can we predict which children with clinically suspected pneumonia will have the presence of focal infiltrates on chest radiographs? Pediatrics. 2004 Mar;113(3 Pt 1):e186-9.
8. Reilly BM, Evans AT. Translating Clinical Research into Clinical Practice: Impact of Using Prediction Rules To Make Decisions. Ann Intern Med. 2006 February 7, 2006;144(3):201-9.
9. Wyatt J, Spiegelhalter D. Field trials of medical decision-aids: potential problems and solutions. Proc Annu Symp Comput Appl Med Care. 1991:3-7.
10. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. 2005 April 2, 2005;330(7494):765-.
11. Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes: A Systematic Review. JAMA. 2005 March 9, 2005;293(10):1223-38.
12. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of Computer-Based Clinical Decision Support Systems on Physician Performance and Patient Outcomes: A Systematic Review. JAMA. 1998 October 21, 1998;280(15):1339-46.
13. Baker MD, Bell LM, Avner JR. Outpatient Management without Antibiotics of Fever in Selected Infants. N Engl J Med. 1993 November 11, 1993;329(20):1437-41.
14. Lee GM, Harper MB. Risk of bacteremia for febrile young children in the post-Haemophilus influenzae type b era. Archives of Pediatrics & Adolescent Medicine. 1998 Jul;152(7):624-8.
15. Roukema J, Los RK, Bleeker SE, van Ginneken AM, van der Lei J, Moll HA. Paper Versus Computer: Feasibility of an Electronic Medical Record in General Pediatrics. Pediatrics. 2006 January 1, 2006;117(1):15-21.
16. Bleeker SE, Moons KGM, Derksen-Lubsen G, Grobbee DE, Moll HA. Predicting serious bacterial infection in young children with fever without apparent source. Acta Paediatrica. 2001 Nov;90(11):1226-32.
17. Bleeker S, Derksen-Lubsen, G, Grobbee, DE, Donders, ART, Moons KGM, Moll, HA. External validation and update of a prediction rule for serious bacterial infection in young children referred with fever without apparent source. (Abstracts from the European Society for Medical Decision Making Meeting Rotterdam, the Netherlands, June 7-8, 2004). Med Decis Making. 2004;24 (4):423.
18. Bleeker SE. Children with fever without apparent source: diagnosis and dilemmas (Dissertation). Erasmus University Rot-

terdam. 2002.

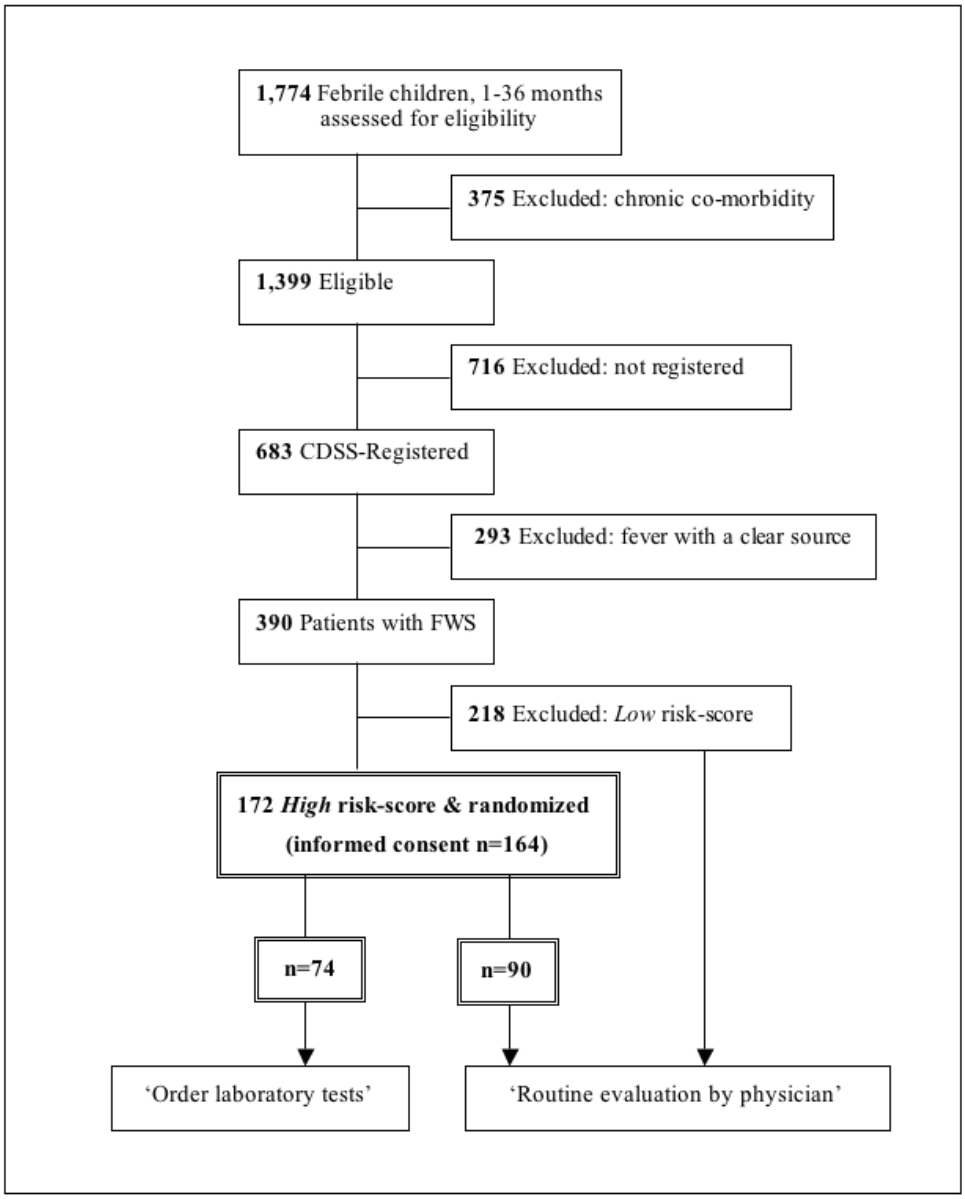
19. Oostenbrink RM, Moons KGMP, Theunissen CCWM, Derksen-Lubsen GMDP, Grobbee DEMDP, Moll HAMDP. Signs of meningeal irritation at the emergency department: How often bacterial meningitis? [Article]. Pediatric Emergency Care June 2001;17(3):161-164. 2001.
20. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al. Clinical Decision Support Systems for the Practice of Evidence-based Medicine. J Am Med Inform Assoc. 2001 November 1, 2001;8(6):527-34.
21. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med. 2004 Aug 30;23(16):2567-86.

Figure 1. CDSS screenshot. The advice is to order laboratory testing for this child.



× = absent
✓ = present

Figure 2. Patient selection and randomization.



FWS = Fever Without apparent Source.
CDSS = Clinical Decision Support System.

Appendix 1. Clinical score chart for referred children

Characteristic	**Points to assign**		**Score**
Duration of fever (days)	*Days*	*Points*	
	½	0	
	1	2	
	1½	4	
	2 - 2½	5	
	3 - 3½	6	
	4 - 4½	7	
	5 - 6	8	
	6½ - 8½	9	
	≥9	10	
History of vomiting	no = 0 / yes = 5 points		
Ill clinical appearance	no = 0 / yes = 4 points		
Chest-wall retractions	no = 0 / yes = 12 points		
Poor peripheral circulation	no = 0 / yes = 7 points		+
			Clinical Score

Low risk-score (≤ 10 points) corresponds with ≤ 12% risk of serious bacterial infection.
High risk-score (> 10 points) corresponds with > 40% risk of serious bacterial infection.

Appendix 2. Clinical score chart for self-referred children

Characteristic	**Points to assign**		**Score**
Duration of fever (days)	For each day of fever 1 point (½ days are rounded up), maximum 10 points		
Age	≤ 1 year = 0		
	> 1 year = -4		
Temperature (°C)	*°C*	*Points*	
	<38*	0	
	> 38	1 point for every 0.5°C	
			Total + 3
			Clinical Score

* If temperature at examination is <38°C, then temperature in history must be ≥38°C.

Low risk-score (< 7 points) corresponds with < 6% risk of serious bacterial infection.
High risk-score (≥ 7 points) corresponds with ≥ 20% risk of serious bacterial infection.

Chapter 9
Summary and future prospects

Summary

The overall aim of the studies described in this thesis was to investigate and optimize the diagnostic process of (febrile) children presenting to the hospital emergency department (ED), and to study aspects of this process as a base for clinical decision support systems. We discussed the use of an electronic medical record with structured data entry, the development of clinical prediction rules for specific diagnostic problems in febrile children attending the hospital ED, the validity of a triage system used for pediatric patients, and the evaluation of a clinical decision support system for diagnostic management of children with fever without apparent source.

In **Chapter 2 and 3**, the use of an electronic medical record (EMR) with structured data entry (SDE), for documentation of patient history and physical examination was described. On the one hand structured data are thought to be necessary to optimally benefit from the EMR, but on the other hand SDE requires an extra effort, while user acceptance is the major barrier to successful EMR implementation. In **Chapter 2**, a comparison was made between documentation of patient information in a paper record and in an EMR. We found that clinicians documented 44% of all available patient information identically in a paper and in an electronic record. 25% of all patient information was documented only in the paper record, and 31% was present only in the electronic record. Interestingly, we found differences in documentation of patient history and physical examination in the electronic record: more information was missing in patient history (38%) compared to physical examination (15%), while physical examination contained more additional information (39%) than patient history (21%). We concluded that physical examination may be better suited for structured data entry than patient history. In a second study (**Chapter 3**), we compared the structured documentation of the same patient information by three different physicians. Of all patient information, only 39% was documented in all three records and 61% was omitted, deduced from or in conflict with the paper record. Hence, the documentation of patient data using structured data entry did not automatically lead to uniformly structured data.

In **Chapter 4**, the development of a clinical prediction rule for radiographic pneumonia was described. Pneumonia is the most frequent serious bacterial infection (SBI) diagnosed in the pediatric ED (i.e. 40-50% of all SBIs). Early identification of children at high or low risk of pneumonia, using readily available characteristics from patient history and physical examination, may support

the diagnostic management. The aim of this study was to develop a clinical prediction rule for pneumonia in children presenting with cough and fever based on available patient characteristics, and to assess the added diagnostic value of white blood cell count (WBC) and C-reactive protein (CRP). As laboratory test results were not available in all patients, multiple imputation was used to estimate the missing values, and to illustrate the effects of this approach on the discriminative ability of the prediction model. All children presenting at the emergency department with fever and cough were included in this study. We found that a combination of information on age, duration of the febrile episode and the presence of tachypnea, could accurately predict the presence of radiological abnormalities, compatible with pneumonia. The added value of WBC and CRP was assessed in two separate models: first, the added value over clinical characteristics was assessed in the patients who had laboratory tests performed. Secondly, in the patients who had no laboratory tests performed, the missing values for WBC and CRP were imputed, based on characteristics that were independently associated with having laboratory tests performed. In both models, WBC and CRP significantly contributed to the diagnosis of pneumonia. The discriminative ability of the model that was based on all patients, with partly imputed laboratory test results, was good with an area under the receiver operating characteristic curve of 0.80 (95% confidence interval [CI] 0.74-0.85). Clinical prediction models, as presented in this study, may be used by emergency department nurses immediately after nurse assessment to initiate the diagnostic process. The prediction model should then be presented as a clinical decision rule, and should also be further validated.

A decreasing incidence of serious bacterial infections (SBI) in children, and the invasiveness of further laboratory testing, justifies a selective diagnostic approach to febrile children. Recognition of children at relatively high risk for SBI early in the diagnostic process may support the decision on further diagnostic or therapeutic management. In **Chapter 5** we assessed the diagnostic value of C-reactive protein (CRP) for SBI in febrile children. Rapid and minimally invasive CRP-tests have become available and may be helpful in the recognition of children at risk for SBI. The added diagnostic value of CRP for SBI over important clinical characteristics (i.e. age, duration of the febrile episode and temperature) and white blood cell count (WBC) was assessed in multivariable logistic regression analyses. We found that CRP is a valuable diagnostic test for SBI in febrile children. The discriminative ability of CRP alone (AUC 0.79, 95%CI 0.75-0.83), was similar to a combination of the clinical characteristics, CRP and WBC. The efficiency of the diagnostic process

for febrile children visiting the ED may be improved by implementation of a rapid CRP-test.

In the development of diagnostic prediction rules, dichotomous logistic regression analyses are generally used, i.e. the probability of a diagnosis of interest being present is estimated versus the probability of its absence. Polytomous logistic regression analyses may however be used to estimate the probabilities of multiple diagnoses simultaneously. In **Chapter 6** we analyzed data of a diagnostic study among 595 children aged 1-36 months, who attended the emergency department with fever without apparent source (FWS), and were thus suspected of having a serious bacterial infection (SBI).

Eleven potential diagnostic predictors were selected based on previous studies and literature (gender, age, duration of fever, changed crying pattern, vomiting, sick impression, pale skin, impaired peripheral circulation, dyspnea, temperature, and weeks of gestation). Three outcome categories to be predicted in our logistic regression analyses were defined: SBI, subdivided in pneumonia and other-SBI, and non-SBI. Four models were developed; model 1 was a polytomous model, estimating probabilities for the three diagnostic outcome categories simultaneously. Model 2 and 3 were sequential dichotomous models, discriminating SBI and non-SBI in step 1, and pneumonia and OSBI in step 2. The two models differed in variable selection method (i.e. model 2 was developed using a fixed set of variables, based on the polytomous model, and in model 3 variables were selected based on contribution to the model). In model 4 each outcome category was opposed to the other two using three separate (non-sequential) dichotomous models. We compared the models with respect to the area under the receiver operating characteristic curve (AUC) and calibration for each of the three outcome categories, and the selected predictors. The AUCs of the three outcome categories were similar for each modeling strategy. The ability to discriminate pneumonia was good, with an AUC of 0.81 (95%CI 0.75-0.87) in each of the derived models, but the ability to discriminate OSBI and non-SBI was moderate with an AUC of approximately 0.69 (95%CI 0.62-0.76) for OSBI, and 0.72 (95%CI 0.66-0.77) for non-SBI. The developed models had comparable calibration and internal validity and only small differences were found in the predictors that were selected in the polytomous and dichotomous models. The definition of homogeneous and clinically relevant outcome categories seems more important than the specific modeling strategy that is chosen.

The evaluations of two clinical decision support systems, used in the ED were presented in **Chapter 7** and **8**. First we evaluated the validity of the Manchester Triage System (MTS) in a pediatric population (**Chapter 7**). The computerized MTS-application supports ED-nurses in objectively assessing the clinical urgency of a patient. The ED is a busy department with many patients attending every day with problems of varying urgency. Following the MTS, every patient can be classified in one of five urgency categories (emergent, very-urgent, urgent, standard, or non-urgent). Clinical urgency can subsequently be used to manage ED patient-flow, with urgent patients having priority over patients with less urgent problems. In this study the MTS urgency-classification was compared with resource-utilization, hospitalization and with a predefined reference urgency classification (based on vital signs, diagnosis, resource-utilization, treatment, and follow-up), as a measure of validity. A representative sample of 1,065 patients, drawn from 18,469 eligible patients who attended the ED of a large inner-city hospital, was studied. The number of patients who used more than 2 resources, increased with a higher level of MTS urgency, and the percentage of hospital admissions increased with higher level of urgency, from 1% in the non-urgent patients to 54% in emergent patients. According to the reference urgency classification, the MTS turned out to be valid regarding the recognition of seriously ill patients, who were all classified as emergent or very urgent. Forty percent of all children was however 'overtriaged', i.e. their urgency was overestimated, which may result in delays in care for patients with truly urgent conditions. Specific pediatric modifications of the MTS should therefore be studied to increase the number of correctly classified pediatric patients.

Finally, in **Chapter 8**, we evaluated the effects of a clinical decision support system (CDSS) for the diagnostic management of young children with fever without apparent source. ED-nurses applied the CDSS, which provided a patient specific diagnostic management advice (i.e. 'order laboratory tests' or 'routine evaluation by physician') in each case. We assessed compliance with the system and evaluated the effects of application of the CDSS on time spent at the ED and amount of performed diagnostic tests in children with FWS. The compliance with registration of febrile children in the CDSS was 50%. Adherence to the management advice was good; in 82% of the CDSS-registered cases the advice to order laboratory tests was followed by the ED-nurse. Against our expectations, it turned out that those children in whom laboratory tests were ordered early in the diagnostic process according to the CDSS-advice, spent 15 minutes more in the ED than the children who were randomized for routine evaluation by the attending ED-physician. The main

explanation for this difference was the fact that significantly more children underwent laboratory evaluation in our intervention group than in our 'routine evaluation' group (82% versus 44%). Our hypothesis was that the limited discriminative ability of the underlying prediction rule for SBI in children with FWS (i.e. the 'knowledge' of the CDSS) may have caused this difference, with an AUC of 0.56 (95%CI 0.48-0.65). This study stressed the importance of including patient outcomes in the evaluation of CDSSs.

Future prospects

The increasing development and implementation of electronic medical records (EMRs) is a major opportunity for the integration of clinical decision rules into practice workflow. To optimally benefit from this opportunity, future research should focus on the applicability and validity of clinical decision rules in EMRs and on the feasibility of structured data entry in an EMR to facilitate automatic generation of management recommendations.

The 'backbone' of a clinical decision support system (CDSS) is the clinical decision rule that is used to generate patient specific recommendations on diagnostic or therapeutic management. Clinical decision rules are a translation of available current evidence from the research literature and practice-based sources and should be developed as computer interpretable algorithms. Typically, after initial derivation and (internal) validation, the validity of decision rules in a new population and, preferably, in other clinical settings is ascertained (external validation). The development of clinical decision rules is guided by a well established methodology. Once developed, however, clinical decision rules should not be regarded as static but rather as dynamic tools that reflect changes in the local clinical setting (e.g. changes in patient population or user profiles). If, for example, a new vaccination strategy is introduced, the validity of the decision rule should be closely monitored and updated if necessary. This subject has received little attention in current literature and there is no established methodology to continuously monitor the validity of decision rules, nor is there a convention when a decision rule needs to be updated, i.e. when is a decision rule not valid anymore? Once we achieve integration of decision rules in daily practice, maintaining the rules up to date and valid will be the major challenge.

Second, patient information should be available in a structured format. The feasibility of routinely collecting patient data in a structured format during patient visits needs further attention, especially in an ED-setting. In our experience, structured data entry (SDE) does not come without a cost: It takes more time for a physician to document all patient information, and differences still exist between documentation of patient information in a paper record and in an EMR. Furthermore, documentation of the same patient information by different physicians using SDE, does not always lead to uniformly structured data. Therefore, both the use of SDE versus free text, and the suitability of structured data for subsequent use in a CDSS, should be further studied.

Regarding the use of SDE in daily practice, one could argue to *optimize* the structured documentation of patient data by, for example, limiting the use of free text or enforcing a certain detail-level. This may at first increase the time and effort to document patient information, but physicians learn to use SDE very quickly, and we expect that the learning effect will continue. The question is, however, whether this scenario is feasible in an emergency department.

Another option is therefore to *limit* the structured documentation of patient data, tailored to specific research questions, to required input for CDSS, or to quality of care benchmarks (e.g. vital signs or final diagnosis). Free text is then used to document the remaining information. If the collection of structured data is driven by the 'information-need' of a CDSS, the effort of capturing structured data may be reduced, while the benefit of clinical decision support still exists. It is unclear which of these two scenarios will fit clinical practice in the ED best.

Concurrent with CDSS-integration in daily practice and maintaining the knowledge base up-to-date, two important aspects of CDSS need further attention: the effects of CDSS utilization and the costs of CDSS implementation and subsequent use. CDSS are shown to have the potential to improve patient care, but in current research the effects of CDSS utilization on clinician performance is often studied instead of effects on patient outcomes [1-3]. The implementation of a CDSS may have unexpected effects on patient outcomes, while at the same time clinician performance seems to be improved [4]. The assessment of CDSS effects on patient outcome should therefore be included in all CDSS-implementation projects. In fact, as with other 'new' interventions in medicine, CDSS should be tested for their clinical effectiveness regarding patient outcomes in randomized trials, prior to wide scale implementation. Furthermore, knowledge of CDSS effects on patient outcomes and clinician performance makes it possible to perform valid cost-effectiveness analyses. Although CDSS are thought to improve practice efficiency and reduce costs, it is unknown whether CDSS are really cost-effective.

Conclusion

Clinical decision support systems (CDSS) may have the potential to improve patient care in the pediatric emergency department. Studies should be initiated to assess integration of CDSS in daily medical practice, CDSS-effects on patient outcomes, and cost-effectiveness of CDSS implementation. The ability to monitor external validity of CDSS recommendations and the possibility

to keep up to date with new evidence and local changes in case mix should be kept in mind during development and implementation of CDSS.

This approach might ultimately contribute to the improvement of the quality of care, by successful development and implementation of clinical decision support systems for the management of febrile children attending the emergency department.

References

1. Hunt, D.L., et al., *Effects of Computer-Based Clinical Decision Support Systems on Physician Performance and Patient Outcomes: A Systematic Review.* JAMA, 1998. **280**(15): p. 1339-1346.
2. Garg, A.X., et al., *Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes: A Systematic Review.* JAMA, 2005. **293**(10): p. 1223-1238.
3. Kawamoto, K., et al., *Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success.* BMJ, 2005. **330**(7494): p. 765-.
4. Han, Y.Y., et al., *Unexpected Increased Mortality After Implementation of a Commercially Sold Computerized Physician Order Entry System.* Pediatrics, 2005. **116**(6): p. 1506-1512.

Chapter 10
Samenvatting

Samenvatting

Het doel van de studies die beschreven worden in dit proefschrift was het onderzoeken en optimaliseren van het diagnostische proces bij kinderen, die zich (met koorts) presenteren op de spoedeisende hulp afdeling (SEH) van het ziekenhuis en het beschrijven van aspecten van dit proces als basis voor beslissingsondersteunende systemen. We bespreken het gebruik van een elektronisch medisch dossier met gestructureerde gegevensinvoer, evenals de ontwikkeling van klinische predictieregels voor specifieke diagnostische uitkomsten bij kinderen die zich met koorts op de SEH presenteren. Daarnaast werden de validiteit van een triage systeem in een pediatrische populatie en een evaluatiestudie naar een klinisch beslissingsondersteunend systeem voor diagnostiek bij kinderen met koorts zonder focus beschreven.

In **Hoofdstuk 2 en 3** werd het gebruik van een elektronisch medisch dossier (EMD) met gestructureerde gegevensinvoer (structured data entry, SDE) voor de documentatie van anamnese en lichamelijk onderzoek beschreven. Er wordt over het algemeen aangenomen dat gestructureerde gegevens noodzakelijk zijn om optimaal te kunnen profiteren van een EMD. Echter, gestructureerde gegevensinvoer kost extra tijd en vereist daarom een extra inspanning van de gebruiker. Acceptatie door de gebruiker is juist de belangrijkste factor voor succesvolle implementatie van een EMD. In **Hoofdstuk 2** werd een vergelijking gemaakt tussen documentatie van patiëntgegevens in een papieren dossier en in een EMD met gestructureerde gegevensinvoer. Het bleek dat artsen 44% van de beschikbare patiëntgegevens identiek documenteerden in het papieren dossier en in het EMD. 25% van de patiëntgegevens werd alleen in het papieren dossier gedocumenteerd en 31% alleen in het EMD. Er werden verschillen gevonden in de documentatie van anamnese en lichamelijk onderzoek in het EMD: meer patiëntgegevens uit de anamnese bleken niet gedocumenteerd te zijn in het EMD (38%) dan uit het lichamelijk onderzoek (15%), terwijl de documentatie van lichamelijk onderzoek in het EMD juist méér extra gegevens bleek te bevatten (39%) dan de anamnese (21%). Wij concludeerden dat lichamelijk onderzoek beter geschikt is voor gestructureerde documentatie in het EMD dan de anamnese. In een tweede studie (**Hoofdstuk 3**), vergeleken we de gestructureerde documentatie van dezelfde patiëntgegevens door drie verschillende artsen. Van alle beschikbare patiëntgegevens werd slechts 39% in alle drie EMD's gedocumenteerd. 61% van de patiëntgegevens werd door minstens één van de drie artsen vergeten, onjuist geïnterpreteerd, of in conflict met de oorspronkelijke informatie gedocumenteerd. Het gebruik van een EMD met gestructureerde gege-

vensinvoer leidt dus niet automatisch tot uniform gestructureerde patiëntgegevens.

In **Hoofdstuk 4** beschreven we de ontwikkeling van een klinische predictieregel voor een pneumonie. Pneumonie is de meest voorkomende ernstige bacteriële infectie (EBI) onder kinderen die zich met koorts op de SEH presenteren (40-50% van alle EBI's). Een vroege herkenning van kinderen met een verhoogd of juist een verlaagd risico op pneumonie, op basis van beschikbare patiëntkarakteristieken, zou het diagnostische proces kunnen ondersteunen. Het doel van deze studie was het ontwikkelen van een predictieregel voor een pneumonie voor kinderen die zich met koorts en hoesten op de SEH presenteren en om de toegevoegde waarde van het leukocytengetal in het bloed (WBC) en C-reactief proteïne (CRP) voor de diagnose pneumonie te evalueren. Omdat laboratoriumuitslagen niet voor alle patiënten beschikbaar waren (er werd niet bij alle patiënten laboratoriumonderzoek verricht), gebruikten we 'multipele-imputatie' methoden om de missende gegevens aan te vullen. Hierdoor kon geïllustreerd worden wat de effecten van deze methode zijn op het discriminerende vermogen van het predictiemodel. In deze studie werden alle kinderen die zich met koorts en hoesten op de SEH presenteerden geïncludeerd.

Een combinatie van informatie over de leeftijd van het kind, de duur van de koortsepisode en de aanwezigheid van tachypnoe, kon de aanwezigheid van een pneumonie, gedefinieerd als de radiologische afwijkingen op een thoraxfoto, betrouwbaar voorspellen. De toegevoegde waarde van WBC en CRP werd geëvalueerd in twee verschillende modellen: eerst werd de toegevoegde waarde onderzocht in de dataset van kinderen bij wie laboratoriumonderzoek werd verricht. Vervolgens werd hetzelfde gedaan in de totale dataset, waarin missende laboratoriumuitslagen werden geïmputeerd. In beide modellen hadden WBC en CRP duidelijk toegevoegde waarde voor de diagnose pneumonie. Met name het model waarin de dataset met deels geïmputeerde gegevens werd gebruikt had een goed discriminerend vermogen (ROC-oppervlakte 0.80, 95% betrouwbaarheidsinterval [BI] 0.74-0.85). Klinische predictieregels, zoals gepresenteerd in deze studie, zouden door SEH-verpleegkundigen gebruikt kunnen worden om het diagnostische proces te initiëren. Het predictiemodel moet in die situatie gepresenteerd worden als een klinische beslisregel en dienst eerst gevalideerd te worden.

De afnemende incidentie van EBI's onder kinderen en het invasieve karakter van aanvullend onderzoek rechtvaardigt een selectief diagnostisch beleid bij kinderen met koorts. De herkenning van kinderen met een relatief groot risico op een EBI, vroeg in het diagnostische proces, zou de

beslissingen met betrekking tot het te volgen diagnostische en/of therapeutische beleid kunnen ondersteunen. In **Hoofdstuk 5** bestudeerden wij de diagnostische waarde van C-reactief proteïne (CRP) voor EBI's bij kinderen met koorts. Snelle en minimaal invasieve CRP-tests zijn tegenwoordig beschikbaar en zouden bij kunnen dragen aan de vroege herkenning van kinderen met een verhoogd risico op een EBI. De toegevoegde waarde van CRP naast belangrijke klinische patiëntkarakteristieken (zoals leeftijd, hoogte van de koorts, duur van de koorts) en het leukocytengetal in het bloed (WBC) werd onderzocht met behulp van multivariabele logistische regressie analyses. Het bleek dat CRP een waardevolle diagnostische test is voor het inschatten van EBI bij kinderen met koorts. Het onderscheidende vermogen (wel/geen EBI) van CRP alleen (ROC-oppervlakte 0.79, 95%BI 0.75-0.83) was vergelijkbaar met het onderscheidende vermogen van patiëntkarakteristieken, CRP en WBC gecombineerd. De efficiency van het diagnostische proces voor kinderen met koorts die de SEH bezoeken zou wellicht verbeterd kunnen worden door de implementatie van een snelle CRP-test.

Meestal worden voor de ontwikkeling van klinische predictieregels dichotome logistische regressie analyses gebruikt; de kans dat een bepaalde uitkomst/diagnose aanwezig is wordt geschat versus de kans dat deze afwezig is. Met behulp van polytome logistische regressie analyses kunnen de kansen op meerdere verschillende uitkomsten in één keer geschat worden. In **Hoofdstuk 6** hebben we de gegevens geanalyseerd van een diagnostische studie onder 595 kinderen van 1 tot 36 maanden oud met koorts zonder focus, die de SEH bezochten en werden beoordeeld voor de aanwezigheid van een EBI. Elf potentiële predictor variabelen werden geselecteerd op basis van eerder onderzoek en literatuur (leeftijd, geslacht, duur van de koorts, veranderd huilpatroon, overgeven, zieke indruk, bleke huid, verminderde perifere circulatie, dyspnoe, temperatuur en zwangerschapsduur). In deze studie definieerden wij drie verschillende diagnostische uitkomsten: EBI, opgesplitst in pneumonie en overige-EBI, en non-EBI. Vier predictiemodellen werden vervolgens gedefinieerd; model 1 was een polytoom model, waarmee de kans op elk van de drie uitkomsten in één keer geschat werd. Model 2 en 3 waren sequentiële dichotome modellen waarmee in stap 1 de kans op EBI en op non-EBI geschat werd en in stap 2 de kans op pneumonie en overige-EBI binnen de EBI-groep. Model 2 en 3 verschilden in de selectie van predictor variabelen (model 2 werd ontwikkeld met een vaste set variabelen, gebaseerd op het polytome model en in model 3 werden predictor variabelen geselecteerd op basis van hun bijdrage aan het model). In model 4 werden drie (niet-sequentiële) dichotome

modellen ontwikkeld, waarin telkens één van de uitkomsten werd afgezet tegen de andere twee. We vergeleken de vier modellen met betrekking tot ROC-oppervlakte voor elk van de drie uitkomsten en met betrekking tot de geselecteerde predictor variabelen. De ROC-oppervlakten voor de verschillende uitkomsten bleken vergelijkbaar voor elk van de gekozen modelleringstrategieën. Het onderscheidende vermogen voor pneumonie was in elk van de ontwikkelde modellen goed, met een ROC-oppervlakte van 0.81 (95%BI 0.75-0.87). Het onderscheidende vermogen voor overige-EBI en non-EBI was echter matig met een ROC-oppervlakte van ongeveer 0.69 (95%BI 0.62-0.76) voor overige-EBI en 0.72 (95%BI 0.66-0.77) voor non-EBI. De ontwikkelde modellen hadden een vergelijkbare calibratie en interne validiteit. Er werden slechts kleine verschillen gevonden in de predictor variabelen die geselecteerd werden in de polytome en dichotome modellen. De conclusie was dat het definiëren van homogene en klinisch relevante diagnostische uitkomstcategorieën belangrijker lijkt te zijn dan de keuze voor een specifieke modelleringstrategie.

In **Hoofdstuk 7** en 8 werden twee studies gepresenteerd waarin klinische beslissingsondersteunende systemen werden geëvalueerd op de SEH. In **Hoofdstuk 7** hebben we de validiteit van het Manchester Triage System (Manchester Triage Systeem, MTS) in een populatie kinderen, die de SEH bezochten, onderzocht. Dit geautomatiseerde triage systeem ondersteunt SEH-verpleegkundigen om op een objectieve wijze de urgentie van behandeling van een patiënt te bepalen. De SEH afdeling is over het algemeen een drukke afdeling waar zich dagelijks vele patiënten presenteren met klachten die sterk variëren in urgentie. Met behulp van het MTS worden alle patiënten geclassificeerd in één van vijf urgentiecategorieën (acuut levensbedreigend, zeer urgent, urgent, standaard, niet-urgent). De vastgestelde urgentie wordt vervolgens gebruikt om de volgorde van behandeling van patiënten op de SEH te bepalen, waarbij patiënten met een hoge urgentie eerder worden geholpen dan patiënten met een lage urgentie. In deze studie werd de validiteit van het MTS in een kindergeneeskundige populatie onderzocht door de urgentie volgens het MTS te vergelijken met de actuele zorgbehoefte van de patiënt (diagnostiek, therapie), de noodzakelijkheid van opname in het ziekenhuis en met een vooraf gedefinieerde urgentieclassificatie, welke als referentiestandaard diende (gebaseerd op vitale kenmerken, diagnose, diagnostiek, therapie en follow-up). Er werd een representatieve steekproef van 1,065 patiënten genomen uit een totale groep van 18,469 patiënten van 0 tot 16 jaar die de SEH van een groot algemeen opleidingsziekenhuis hadden bezocht. Het aantal patiënten dat twee of meer diagnostische tests of therapeutische interventies onderging was

gerelateerd aan een hogere MTS urgentieclassificatie. Het percentage patiënten dat opgenomen moest worden in het ziekenhuis nam toe naarmate de urgentie hoger was, van 1% van de niet-urgente patiënten tot 54% van de acuut levensbedreigende patiënten. Het MTS bleek valide met betrekking tot de herkenning van acuut zieke patiënten; alle patiënten die volgens de referentiestandaard acuut levensbedreigend waren, werden in het MTS als acuut bedreigd of zeer urgent geclassificeerd. Veertig procent van alle kinderen echter, werd 'overgetrieerd', ofwel, de urgentie volgens het MTS was hoger dan volgens de referentiestandaard. Dit zou kunnen resulteren in vertraging van de zorg voor patiënten met een 'echt' urgent probleem. Specifieke kindergeneeskundige aanpassingen van het MTS zouden daarom overwogen moeten worden om het aantal correct geclassificeerde patiënten te verhogen.

In **Hoofdstuk 8** hebben we een klinisch beslissingsondersteunend systeem (Clinical Decision Support System, CDSS) voor de diagnostiek bij jonge kinderen met koorts zonder focus geëvalueerd. SEH-verpleegkundigen gebruikten het CDSS bij alle kinderen die zich met koorts presenteerden. Het CDSS leverde op basis van de ingevoerde gegevens een patiëntspecifiek advies met betrekking tot het te volgen diagnostisch beleid ('laboratoriumonderzoek inzetten' of 'standaard beoordeling door arts'). We onderzochten de compliance van SEH-verpleegkundigen met het systeem en evalueerden het effect van het CDSS-gebruik op de tijd die patiënten doorbrachten op de SEH en op de hoeveelheid laboratoriumdiagnostiek die werd verricht. De compliance met registratie van kinderen met koorts was 50%. Het CDSS-advies laboratoriumdiagnostiek te verrichten werd in 82% van de gevallen opgevolgd door de SEH-verpleegkundige. Tegen de verwachting in, bleek dat de kinderen bij wie door de SEH-verpleegkundige laboratoriumonderzoek was aangevraagd volgens het CDSS-advies, gemiddeld 15 minuten langer op de SEH verbleven dan de kinderen die gerandomiseerd waren voor 'standaard beoordeling door arts'. De belangrijkste verklaring voor dit verschil was het feit dat er in de interventiegroep (laboratoriumonderzoek aangevraagd volgens CDSS-advies) bij significant meer kinderen laboratoriumdiagnostiek werd gedaan dan in de groep kinderen met 'standaard beoordeling door arts' (82% versus 44%). Dit verschil zou ontstaan kunnen zijn door een te laag discriminerend vermogen van het onderliggende predictiemodel voor EBI in kinderen met KZF (de 'kennis' van het CDSS). De oppervlakte onder de 'receiver-operating characteristic curve' (ROC-curve) was 0.56 (95%BI 0.48-0.65). Deze studie onderstreept het belang van het gebruik van uitkomstmaten op patiëntniveau in de evaluatie van klinische beslissingsondersteunende systemen.

Dankwoord

Eindelijk is het dan zover. Als alles ongeveer 'af' is mag je aan het dankwoord beginnen. Ik verkeer in de gelukkige positie dat ik dat in de zon middenin een rustig wijnveld in de Languedoc mag doen. Een mooi moment om terug te kijken op vier jaar werken aan verschillende onderzoeksprojecten, waarvan de verslaglegging hier voor u ligt.

Terugkijkend realiseer ik me eens te meer dat er ontzettend veel mensen bij dit project betrokken zijn geweest zonder wie het niet tot een dergelijk mooi einde had kunnen komen. Een aantal van hen wil ik daarom hier noemen.

Dr. Henriette Moll

Beste Henriette, om met de deur in huis te vallen, ik had me geen betere co-promotor kunnen wensen. Toen ik vier jaar geleden kennis kwam maken was ik al onder de indruk van je inspiratie, motivatie en wetenschappelijke vooruitstrevendheid. De afgelopen jaren zijn voor mij erg leuke leerzame jaren geweest. Ik ben blij dat je me hebt weten te overtuigen van het belang van een gedegen epidemiologische opleiding en dankbaar dat je me daartoe vervolgens ook de mogelijkheden hebt geboden.

Prof. Dr. J. van der Lei

Beste Johan, omdat de afdeling Medische Informatica zo'n belangrijke rol heeft gespeeld bij dit onderzoeksproject, ben ik erg blij dat je mijn promotor hebt willen zijn. Het snelle en heldere overzicht dat je telkens hebt en je scherpe en eerlijke commentaar hebben mij enorm geholpen. Ik hoop dat we zo nu en dan nog eens een goed glas (franse?) wijn drinken!

Prof. Dr. H.A. Büller

Zeer geachte professor Büller, ondanks uw volle agenda heeft u tijd willen vrijmaken om mijn promotor te zijn, waarvoor mijn grote dank. U zult begrijpen dat het mij verheugd dat het ErasmusMC de ontwikkeling en implementatie van het elektronisch medisch dossier zo voortvarend aanpakt, en het belang ervan onderkent.

Dr. E.W. Steyerberg

Beste Ewout, ten eerste wil ik je bedanken voor het plaatsnemen in de kleine promotiecommissie en daarin ook secretaris hebt willen zijn. De afgelopen jaren hebben we veel samengewerkt, niet alleen aan de lopende onderzoeksprojecten maar ook met het (succesvol) schrijven van een tweetal nieuwe projectvoorstellen. Ik ben blij dat ik zoveel van je heb kunnen leren, al raakte ik de draad nog wel eens kwijt. Ik heb er grote bewondering voor hoe jij de kliniek en het onderzoek zo goed bij elkaar weet te krijgen.

Prof. Dr. H.M. Oudesluys-Murphy

Zeer geachte professor Oudesluys, dank voor het plaatsnemen in de kleine promotiecommissie. Als nieuwe 'baas' van mij in het LUMC, hoop ik nog veel van u te kunnen leren in de komende jaren.

Prof. Dr. P. Patka

Zeer geachte professor Patka, het is een eer dat u, als hoofd van een grote SEH-afdeling, plaats heeft willen nemen in de kleine promotiecommissie en het manuscript heeft willen beoordelen.

Prof. Dr. M.G.M. Hunink en Prof. Dr. J.C. de Jongste wil ik bedanken voor het plaatsnemen in de grote promotiecommissie.

Naast de reeds genoemde mensen zijn er nog een aantal zonder wie dit proefschrift er niet was geweest, zoals Marcel de Wilde. Beste Marcel, vanaf m'n eerste dag in het Sophia heb ik het genoegen gehad met je mogen samenwerken. Van Structured Data Entry tot implementatie van beslisregels en tot het ontwikkelen van een MTS-applicatie. Stuk voor stuk zijn het cruciale elementen geweest voor dit onderzoeksproject. Ik hoop dat je in Rotterdam gelegenheid zult hebben om nog heel veel van je ideeën uit te werken, want ik ben ervan overtuigd dat we er nog heel erg veel aan kunnen hebben.

Dan Dr. Renske Los, data-extractie uit SDE, ik vond het wat ambitieus (verzamel eerst maar

eens de data dacht ik), maar uiteindelijk lagen onze onderzoeken toch wel heel erg dicht bij elkaar. Uiteindelijk ' delen' we zelfs een hoofdstuk in onze proefschriften. Wellicht zie ik je nog een keer als EMD-consultant?

De afdeling SEH van het Sophia Kinderziekenhuis wil ik bedanken voor alle inzet met betrekking tot het koortsonderzoek en het gebruik van de beslisregels. Bovendien mag ik me gelukkig prijzen dat ik als arts-assistent met jullie heb mogen samenwerken. Met name wil ik Jan van der Linden en Sonaida Pieternella bedanken voor de tijd en energie die jullie in het koortsproject en in het MTS hebben willen steken.

De afgelopen jaren hebben een aantal medisch studenten, elk met een eigen onderzoeksvraag, in grote mate bijgedragen aan de vorderingen van dit onderzoeksproject, en voor een leuke samenwerking gezorgd. Fleur Velders, Sofie Benneheij, Yvonne Maat, Marianne van der Windt, Marielle Vollebregt, Eefje Dons, Alet Eising, Myrthe Koster, Mienke de Weijs, Rhiannon van Loenhout en Laura Melenhorst wil daarvoor hartelijk bedanken. Ik hoop jullie enthousiast zijn geworden voor het doen van onderzoek en wellicht zie ik jullie nog eens terug in de kindergeneeskunde?

Vier jaar geleden begon ik 'alleen' op Sp1545, met een stapel verhuisdozen van m'n voorgangers. Ik ben ontzettend blij dat ik al snel roomies kreeg.

Als eerste Marinka, jaren zaten we rug-aan-rug in Rotterdam en nu staan we zij-aan-zij in Leiden. Ik ben blij dat we collega's blijven! Idse, man van het scherpe commentaar, Barry White liefhebber, SPSS-vraagbaak en uitstekende kok, ik zal je missen (oef, dramatisch). Ik hoop dat jullie het naar je zin gaan hebben in Zwolle! (ik kom nogal eens langs Zwolle op weg naar het noorden des lands, bezwaar als ik zo nu en dan eens dag kom zeggen?) Mandy, alweer geruime tijd geleden gepromoveerd en aan de opleiding begonnen. Gelukkig kwam je af en toe nog even om de hoek kijken. Ook wij blijven eigenlijk gewoon collega's op afstand!

Sacha Bleeker, m'n voorgangster/voorbeeld (echt waar), die weer terugkwam naar Rotterdam. Ik ben blij dat we nog weer je oude kamer hebben kunnen delen en zal je humor en opgewektheid missen. Mirjam, mijn 'MTS-opvolgster', ik wens je alle succes de komende jaren en ben ervan overtuigd dat het MTS met jouw aanpassingen over niet al te lange tijd de Nederlandse standaard voor triage bij kinderen zal worden!

Een aantal arts-onderzoekers wil ik nog in het bijzonder bedanken voor een ontzettend leuke

tijd in het Sophia: Jochem, Gwenda, José, Matthijs, Robert, Leon, Pieter, Edwin, Clementien en Carmelo en last but not least, Karen Corver, de eerste jaren in Rotterdam was jij m'n maatje met noordelijke nuchterheid. Je maakte me wegwijs in het Sophia en we reisden samen vaak heen-en-weer naar Den Haag. Ik vond het erg jammer dat je naar Engeland verhuisde, maar ben blij dat we elkaar nog geregeld zien!

Dan mijn paranimfen Sander IJtsma en Reinoud Roukema. Sander, vanaf het moment dat we in 1996 aan onze studie geneeskunde begonnen in Groningen, zijn we goede vrienden geworden en gebleven. Allebei aan het promoveren (wie had dat toen gedacht) en aan de opleiding tot specialist begonnen. Ik hoop dat we de tijd kunnen blijven vinden om elkaar nog vaak te zien. Ik vind het supermooi dat je, 10 jaar nadat we elkaar hebben leren kennen, naast me staat als paranimf. Beste Reinoud, broertje van om de hoek en inmiddels vader van m'n prachtige neefje Jens, ongeveer een maand geleden stond ik aan jouw zijde als getuige op je huwelijk en nu sta je naast mij; toch een prettig idee om je grote broer bij je te hebben.

Natuurlijk wil ik hier ook mijn ouders en zusjes Janneke en Annemarie bedanken voor alle interesse, steun en aandacht de afgelopen tijd. Pap, mam, wat mogen jullie er ontzettend trots op zijn dat we zo'n gezellig en gelukkig gezin zijn met z'n allen (Bart en Natas incluis!). Dit proefschrift geeft hopelijk eindelijk een goed beeld van waar ik de afgelopen jaren nu eigenlijk zo druk mee ben geweest.

Lieve Mariek, je moest eens weten hoe belangrijk je voor me bent geweest en nog steeds bent. De afgelopen maanden waren erg zwaar, met het afronden van dit proefschrift, de start van m'n opleiding in Leiden en al het geregel rondom ons nieuwe huis. We gaan een geweldige toekomst tegemoet!

Curriculum Vitae

Jolt Roukema werd geboren (23 oktober 1977) in Emmen en groeide op in Assen. In 1996 behaalde hij zijn VWO diploma aan de Christelijke Scholengemeenschap 'Vincent van Gogh' te Assen. Aansluitend studeerde hij Geneeskunde aan de Rijksuniversiteit Groningen. In 1997 behaalde hij zijn propedeutisch examen. De co-schappen werden doorlopen in het Martini Ziekenhuis te Groningen, waarna hij zijn wetenschappelijke stage en keuzecoschap deed in het Juliana Kinderziekenhuis te Den Haag (Dr. G. Derksen-Lubsen). Hij behaalde zijn doctoraal- en artsexamen in 2002. Na het behalen van het artsexamen werkte hij als arts-onderzoeker in het Erasmus Medisch Centrum, Sophia Kinderziekenhuis te Rotterdam. Het onderzoeksproject waaraan hij werkte groeide uit tot een promotieonderzoek waarvan het eindresultaat hier voor u ligt. Tijdens zijn promotieonderzoek behaalde hij de graad Master of Science in de Klinische Epidemiologie (Netherlands Institute of Health Sciences (Nihes), Erasmus Universiteit Rotterdam). Per 1 juli 2006 is hij begonnen aan de opleiding Kindergeneeskunde in het Leids Universitair Medisch Centrum (opleider Dr. R.N. Sukhai).

List of abbreviations

AUC = Area under the (receiver operating characteristic) curve

CDR = Clinical decision rule

CDSS = Clinical decision support system

CI = Confidence interval

CRP = C-reactive protein

EBM = Evidence based medicine

ED = Emergency department

EMR = Electronic medical record

FWS = Fever without apparent source

MTS = Manchester Triage System

NPV = Negative predictive value

OR = Odds ratio

OSBI = Other serious bacterial infections

PPV = Positive predictive value

ROC-area = Area under the receiver operating characteristic curve

SBI = Serious bacterial infection

SDE = Structured data entry

WBC = White blood cell count