Evidence Based Decisions in Nursing 
and their effect on Quality of Care
Evidence Based Decisions in Nursing and their effect on quality of care
Evidence based decisions in nursing and their effect on quality of care.
Dissertation, University of Amsterdam, Amsterdam, The Netherlands

Cover: Jonne Overbeeke-Bakker.
With permission from ‘Goede Doel Bedels’; Almelo (www.goedodoelbedels.nl)
AutismeBedel: Twee werelden samen!
De AutismeBedel toont de wereld, waarin de persoon met autisme een onderdeel is, waarbij de puzzelstukken, die autisme vertegenwoordigen, wijzen op de ingewikkelder en soms verwarrende handicap.

Lay-out: Chris Bor, Medical Photography and Illustration, Academic Medical Center, Amsterdam, The Netherlands
Printed by: Buijten & Schipperheijn, Amsterdam, The Netherlands
Titel: Evidence based decisions in nursing and their effect on quality of care

Copyright ©2012, Marja Storm-Versloot, Hilversum, The Netherlands

All rights reserved. No part of this publication may be reproduced or transmitted in any form by any means, electronic or mechanical, including photocopy, recording or any information storage and retrieval system, without written permission of the author.

Two studies in this thesis were financially supported by a grant from the Netherlands Heart Foundation (grant 2000/216)

The printing of this thesis was financially supported by the Department of Surgery, Academic Medical Center, Amsterdam
Evidence Based Decisions in Nursing and their effect on quality of care

ACADEMISCH PROEFSCHRIFT

ter verkoop van de graad van doctor aan de Universiteit van Amsterdam op gezag van de Rector Magnificus prof. dr. D.C. van den Boom ten overstaan van een door het college voor promoties ingestelde commissie, in het openbaar te verdedigen in de Aula der Universiteit op woensdag 26 september 2012, te 13:00 uur

door

Maria Nellie Versloot

geboren te Velsen Noord
Promotiecommissie

Promotores:  Prof. dr. D.A. Legemate
              Prof. dr. J.C. Goslings

Co-promotores:  Dr. D.T. Ubbink
                Dr. H. Vermeulen

Overige leden:  Prof. dr. P.J.M. Bakker
                Prof. dr. R.J. de Haan
                Prof. dr. R.J.P.M. Scholten
                Prof. dr. A.B. van Vugt
                Prof. dr. T. van Achterberg

Faculteit der Geneeskunde
Aan allen die mij lief zijn
## CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>General introduction and outline of the thesis</td>
<td>9</td>
</tr>
<tr>
<td><strong>PART 1:</strong></td>
<td><strong>EVIDENCE BASED DECISIONS IN NURSING</strong></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Comparison of an informally structured Triage System, the Emergency Severity Index, and the Manchester Triage System to distinguish patient priority in the Emergency Department</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>Observer agreement of the Manchester Triage System and the Emergency Severity Index: a simulation study</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>The effect of a minimal intervention strategy in addition to nicotine replacement therapy to support smoking cessation in cardiovascular outpatients: a randomized clinical trial.</td>
<td>47</td>
</tr>
<tr>
<td>5</td>
<td>Clinical relevance of routinely measured vital signs in hospitalized patients; a systematic review</td>
<td>61</td>
</tr>
<tr>
<td>6</td>
<td>Topical silver for preventing infected wounds: a Cochrane systematic review</td>
<td>85</td>
</tr>
<tr>
<td><strong>PART 2:</strong></td>
<td><strong>EFFECT OF EVIDENCE-BASED DECISIONS ON QUALITY OF CARE</strong></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Influence of the Manchester Triage System on waiting time, treatment time, length of stay and patient satisfaction; a before and after study</td>
<td>133</td>
</tr>
<tr>
<td>8</td>
<td>The number of smokers needed to screen and treat in a smoking cessation programme. Prospective cohort data</td>
<td>147</td>
</tr>
<tr>
<td>9</td>
<td>Long-term adherence to a local guideline on postoperative body temperature measurement: mixed methods analysis</td>
<td>163</td>
</tr>
<tr>
<td>10</td>
<td>Do stakeholders in wound care prefer evidence-based wound care products? A survey in the Netherlands</td>
<td>179</td>
</tr>
<tr>
<td>11</td>
<td>Epilogue</td>
<td>195</td>
</tr>
<tr>
<td>12</td>
<td>Summary in English</td>
<td>207</td>
</tr>
<tr>
<td></td>
<td>Summary in Dutch</td>
<td>215</td>
</tr>
<tr>
<td>Appendices</td>
<td>List of publications</td>
<td>223</td>
</tr>
<tr>
<td></td>
<td>Dankwoord</td>
<td>231</td>
</tr>
<tr>
<td></td>
<td>Curriculum vitae</td>
<td>237</td>
</tr>
</tbody>
</table>
Chapter 1

General introduction and outline of the thesis
INTRODUCTION

In healthcare today, improving the quality of care has gained priority on every agenda. Alarming reports on errors, exploding costs and large variations in practice have inspired this urge for quality improvement in healthcare. \(^1\)\(^-\)\(^3\) From these reports it has become clear that, between the present-day and the desired quality of healthcare, there lies not just a gap, but a chasm. \(^3\)

This quality chasm in healthcare also exists in the Netherlands. Although in this country healthcare has excellent accessibility, it lacks transparency in terms of suitable and accurate information about quality and patient outcomes. \(^4\) Furthermore, the healthcare provided shows a large variation in practice. \(^5\) Improving the quality of care and closing the quality chasm is important not only for policymakers, regulatory authorities, educational institutions and boards of (hospital) directors, but also, and not least, for healthcare professionals and patients themselves. \(^6\)

There is no agreed definition of quality of care, but the most commonly used definition has been formulated by the Institute of Medicine: 'The degree to which health services for individuals and populations increases the likelihood of desired outcomes and are consistent with current professional knowledge'. \(^3\)

In the late 1980s, Evidence-Based Medicine (EBM) was coined as a tool to improve the quality of healthcare among doctors. It is an iterative process to maintain professional knowledge and foster lifelong learning. EBM is defined as the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients, \(^7\) and involves issues such as safety, effectiveness, efficiency, accessibility, and patient centeredness. After its launch, the EBM-paradigm has spread all over the world, and authorities and health services have adopted EBM to improve the quality of care. \(^3\), \(^8\), \(^9\)

Nowadays, improving clinical practices, introducing novel practices, and minimizing practice variation are increasingly being based on best available evidence from the scientific literature. Other healthcare disciplines, like nursing, have also adopted this paradigm in their decision-making regarding patient care. \(^10\) This has led to the more general term Evidence-Based Practice (EBP).

Building a body of state-of-the-art and relevant knowledge is an essential pillar for bridging the quality chasm and making evidence-based decisions possible. \(^11\) Conducting research generates new knowledge, and EBP promotes the integration of this knowledge into practice.

It should be noted that doctors and nurses cannot keep abreast of all newly published research evidence. \(^12\), \(^13\) Therefore, aggregated evidence in the form of evidence-based clinical guidelines can be helpful to allow novel insights to be quickly appreciated and also to improve the process, the structure, and thereby the quality of care. However, evidence-based improvements in the level of patient-related outcomes are rarely investigated and seem to be relatively small. \(^14\) Even when clinical guidelines have been implemented
in practice, 30-50% of patients still do not receive care at the latest standards.\textsuperscript{4,15} This non-adherence to guidelines may lead to unnecessary diagnostics, suboptimal treatment, or even adverse events.\textsuperscript{16,17} Thus, the implementation of guidelines remains a challenge for healthcare professionals responsible for improving the quality of clinical care.\textsuperscript{18}

This thesis is a compilation of efforts to contribute to the body of knowledge in nursing care, to promote evidence-based decision-making and to overcome the challenges in the implementation process. For this purpose, we have collected and investigated evidence for various (novel) clinical practices and routine practices (‘sacred cows’) in nursing and their effect on the quality of care. Although the focus is mainly on nursing care, all the topics investigated have multidisciplinary aspects.

**OUTLINE OF THE THESIS**

This thesis describes in two parts the underpinning of evidence-based decisions in nursing and their effect on the quality of care. **Part 1** describes four different studies on decisions on nursing practices at the emergency department, the out-patient department, and the in-patient hospital care setting. These practices are either novel practices for which evidence is lacking or unclear, or routine clinical practices of which the effectiveness is questioned. **Part 2** then describes the effects of each of these evidence-based decisions on the quality of care. If novel practices are implemented, it is imperative that their effects on the quality of care are investigated in order to weigh their pros and cons. If routine practices have to be changed, it is necessary to understand the difficulties in establishing long-term adherence to the change. Furthermore, understanding the process of disseminating the novel evidence and creating awareness of this evidence among all doctors and nurses can help patients receive the appropriate care.

**PART 1: EVIDENCE-BASED DECISIONS IN NURSING**

**Novel practices**

The first example of a novel practice was the implementation of one of the formally structured triage systems (that is, the Manchester Triage System (MTS) or the Emergency Severity Index (ESI)) at the Emergency Department (ED). To make an evidence-based decision as to which system should be used, the validity and the inter- and intra-observer agreement for both triage systems is investigated. **Chapter 2** deals with a prospective observational comparative study, in which the percentages of undertriage are determined and compared, and the validity of the two systems (MTS and ESI) and the current local informally structured triage system (ISS), and their relation to resource use, hospital admission, and length of stay, are studied. In **Chapter 3**, the inter- and intra-observer agreement of the MTS and the ESI is compared using paper-based clinical scenarios. The
studies were set up to help decide which triage tool to use, and to enlarge the body of knowledge as to validity and agreement.

The second example of a novel practice was the initiation of a nurse-led structured behavioral smoking cessation program, the Minimal Intervention Strategy (MIS), in cardiovascular out-patients. In Chapter 4, the effectiveness of this program combined with nicotine replacement therapy (NRT), as compared with NRT alone, at the cardiovascular out-patient clinics is described. The outcome of this study should help in the process of deciding whether or not to implement the MIS in daily nursing practice in cardiovascular out-patient clinics.

Routine clinical procedures
In medical and surgical hospitalized patients, numerous routine measurements of vital signs, with uncertain effectiveness, are performed. Therefore a systematic review is conducted, as described in Chapter 5, and the clinical relevance of routinely measured vital signs is determined.

Another widely used routine procedure is the use of silver sulfadiazine in burn patients to prevent wound infection and to promote wound healing. However, robust evidence is lacking for many of the outcomes. Hence, in Chapter 6, we describe a Cochrane systematic review on the effects of silver-containing dressings and topical agents for the prevention of wound infection and the promotion of wound healing in uninfected wounds.

The results of both studies are helpful in gaining insight into the value of routine procedures in supporting daily care and clinical decision-making.

PART 2: EFFECT OF EVIDENCE-BASED DECISIONS ON THE QUALITY OF CARE

Effect of decisions on patient care
Based on the results described in Chapters 2 and 3, the MTS was implemented at our ED. Chapter 7 deals with a prospective observational before and after study, in which the effects of the implementation of MTS on waiting times, length of stay and patient satisfaction were determined.

Smoking cessation is an important factor in reducing cardiovascular mortality. The “number needed to treat” (NNT), commonly used to quantify an intervention effect, does not reflect the total effort necessary to identify all patients who could potentially be treated. One of the components in determining the usefulness, relevance and efficiency of screening programs can be expressed by calculating the “number needed to screen” (NNS). Therefore, in Chapter 8 the efficiency of the MIS is studied, and the NNT and NNS are determined, based on the prospectively sampled data from the study described in Chapter 4.
Challenges to change routine care

Two challenges in the implementation process for changes in routine care are described in the last two chapters.

High short-term guideline adherence rates were achieved in decreasing unnecessary routine post-operative body temperature measurements,\textsuperscript{20,21} but regression to old habits is a common human flaw.\textsuperscript{22,23} Hence, the first challenge is to achieve long-term adherence. \textbf{Chapter 9} describes a study in which a mixed-methods analysis was used to determine if adherence was persistent over time and which facilitators and barriers affected long-term adherence.

The second challenge is how to achieve awareness among doctors and nurses of novel research evidence, particularly when this evidence is more and more available in aggregated form in systematic reviews and clinical guidelines. Based on the results of Chapter 6, \textbf{Chapter 10} describes a cross-sectional study investigating the awareness and use of evidence by different stakeholders in wound care in their choice of wound dressings.

Finally, \textbf{Chapter 11} discusses the study results on the body of novel knowledge and its effect on quality of care.
REFERENCES


PART 1

EVIDENCE BASED DECISIONS IN NURSING
Comparison of an informally structured triage system, the Emergency Severity Index, and the Manchester Triage System to distinguish patient priority in the Emergency Department

MN Storm-Versloot
DT Ubbink
J Kappelhof
JSK Luitse

Acad Emerg Med. 2011; 18:822–829
ABSTRACT

Objectives: The objective was to compare the validity of an existing informally structured triage system with the Emergency Severity Index (ESI) and the Manchester Triage System (MTS).

Methods: A total of 900 patients were prospectively triaged by six trained triage nurses using the three systems. Triage ratings of 421 (48%) patients treated only by emergency department (ED) physicians were compared with a reference standard determined by an expert panel. The percentage of undertriage, the sensitivity, and the specificity for each urgency level were calculated. The relationship between urgency level, resource use, hospitalization, and length of stay (LOS) in the 900 triaged patients was determined.

Results: The percentage of undertriage using the ESI (86 of 421; 20%) was significantly higher than in the MTS (48 of 421; 11%). When combining urgency levels 4 and 5, the percentage of undertriage was 8% for the informally structured system (ISS), 14% for the ESI, and 11% for the MTS. In all three systems, sensitivity for all urgency levels was low, but specificity for levels 1 and 2 was high (>92%). Sensitivity and specificity were significantly different between ESI and MTS only in urgency level 4. In all 900 patients triaged, urgency levels across all systems were associated with significantly increased resource use, hospitalization rate, and LOS.

Conclusions: All three triage systems appear to be equally valid. Although the ESI showed the highest percentage of undertriage and the ISS the lowest, it seems preferable to use a verifiable, formally structured triage system.
INTRODUCTION

Patients arriving at the emergency department (ED) are often confronted with long waiting times. These may be caused by arrival volumes, order of arrival, or clinical urgency. Although EDs always use some form of triage, either formal or informal, overcrowding of EDs makes accurate triaging essential to avoid delays in critical patient care, which may result in long waiting times and poor outcomes. There are a number of three-level systems that rank patients as emergent, urgent, or non-urgent. However, as five-level triage systems have been shown to be more reliable and valid than three-level systems, they are likely the systems of choice.

Worldwide, there are four five-level triage systems in use: the Australasian Triage Scale, the Canadian Triage and Acuity Scale, the Manchester Triage System (MTS) and the Emergency Severity Index (ESI). The MTS and the ESI are the most commonly used in the Netherlands.

The MTS was developed in the United Kingdom and is widely used. The MTS contains 52 flow charts, each representing a presenting complaint. The presenting complaint determines which flow chart should be followed. Each flow chart is based on a five-step decision process that uses discriminators at each step to assign patients to one of the five triage categories. A color indicates the level of urgency and its associated maximum waiting time: red = immediate care by a physician; orange = 10 minutes; yellow = 60 minutes; green = 2 hours; and blue = 4 hours. Interobserver and intraobserver agreement on the MTS has been found to be substantial to excellent. Inter- and overtriage as determined by an expert panel ranges from 5 to 25%. Validity has been tested in several studies, however, only in children and in patients with chest pain. Therefore, the overall validity of MTS in daily clinical practice is not supported in the literature.

The ESI system was developed in Boston, Massachusetts. This system uses one algorithm, with ratings ranging from level 1 (the most acutely ill patients) to level 5 (the least resource-intensive patients). The triage nurse estimates the number of resources needed to discharge the patient from the ED in those patients who do not meet ESI level 1 or 2 criteria. The ESI system is valid for both children and adults and has a high interobserver agreement. When triage urgency levels estimated by nurses were compared with the “real” urgency level, version 3 of the ESI had an undertriage rate of 9% and an overtriage rate of 11%.

The two triage systems have been designed for different purposes. The MTS is meant to place patients in order of priority and to assure that patients do not have to wait longer than is safe, given the presenting complaint. The ESI integrates acuity and estimated resource consumption to determine treatment priority. Both ESI and MTS seem to be useful, but to date no studies have compared the validity of these systems within the same patient mix, nor to an informally structured triage practice. We defined validity as the agreement of classifications by the triage systems with a reference standard and
with actual resource utilization. In this prospective observational comparative study, we determined and compared the percentage of undertriage; the validity of both structured triage systems and a local informally structured triage system (ISS); and their relation to resource use, hospital admission, and length of stay (LOS).

METHODS

Study design
This was a prospective, single-center, observational comparative study combined with a retrospective chart review to determine the validity of the ISS, the ESI and the MTS triage systems. The study was explained to patients, and all gave oral informed consent. Our local institutional review board waived the requirement for written informed consent.

Study setting and population
This study was conducted between November and December 2005 at an urban tertiary care academic teaching hospital with a Level 1 trauma center. The ED sees almost 31,000 patients annually, of whom approximately 1,000 (3%) patients are seen in a trauma room. Almost 71% of the patients were self-referrals, while 29% of the patients were referred by a general physician (GP) to a specialist. The overall admission rate was approximately 18%, and 15% of the patients were younger than 15 years.

In 2005, when the study began, no formally structured triage system was in use on our ED, but rather an ISS was in place. This system was based on clinical expertise, but not on explicit criteria and information. When patients were registered, the patients’ appearance and presenting complaints were judged, and the nurse or receptionist implicitly answered the question: “Could this patient wait safely before being seen?” Options were: patient could not wait at all, patient should be seen as soon as possible, or patient could wait.

For the purpose of this study, six ED nurses received a 6-hour combination of didactic and practical training in each triage system (ESI and MTS), in accordance with national standards. At random and on different days of the week between 12 noon and 10 PM, the nurses triaged all patients entering the ED consecutively. Patients already triaged before hospital arrival by ambulance staff, and who met the criteria for treatment at the trauma room according to current guidelines, were not triaged again, but classified as level 1 (ESI) or “red” (MTS) patients. Patients who left the ED without being seen by a physician, or whose records were not available, were excluded from analysis.

Study Protocol
Patients were first registered by the department receptionist and then prospectively classified by one of the trained triage nurses, using all three systems (ISS, ESI and MTS). The treating ED nurse and physician were blinded to the classification codes. If the triage nurse classified the patient as needing to be seen immediately, the patient was turned over
to the treating ED nurse. Otherwise, the patient was sent to the waiting room and followed
the usual procedure.

The information needed to identify the level of acuity is different in each of the triage
systems. For this reason, the triage nurse first classified the patient according to informally
structured practice, which required the least information, then by means of the ESI system,
and finally with the MTS, which required the most information. For triage following the
informally structured practice, based on the patients’ appearance and complaints presented,
the nurse answered the question: “Under difficult circumstances (e.g., an overcrowded
waiting room), what is the maximum possible time that this patient will be able to wait
before being seen?” The answer options were: patient could not wait or was able wait up
to a maximum of 15 minutes, 1 hour, or 4 hours.

Level 1 was defined as the most urgent category, comprising “patient could not wait”
(ISS), level 1 (ESI), and red (MTS). Level 2 comprised the urgency levels “was able to wait
up to a maximum of 15 minutes” (ISS), level 2 (ESI), and orange (MTS); level 3 comprised “1
hour” (ISS), level 3 (ESI), and yellow (MTS); level 4 comprised level 4 (ESI) and green (MTS);
and level 5 comprised “4 hours” (ISS), level 5 (ESI), and blue (MTS).

Reference standard
To determine the effectiveness of the triage systems in those patients treated only by
an ED physician, the classification of these systems was compared with the reference
standard. (Patients referred to meet a specialist were excluded from this analysis, because
they had already undergone some prehospital triage and commitment to specific resource
utilization.) This reference standard was determined by an expert panel consisting of seven
experienced ED physicians. Each physician evaluated all cases individually and was blinded
to the conclusions of the other panel members. The physicians evaluated retrospectively
what the real degree of urgency would have been, based on the ED data, results of
diagnostic tests, and the final diagnosis. Knowing the outcome is prerequisite to determine
which patients were really in danger and were misclassified by the systems. Except for
age and sex, patient data were deidentified. The main question for the panel was: “Under
difficult circumstances what is the maximum possible time that this patient would have
been able to wait before being seen?” The answer options were: patient could not wait or
was able to wait up to a maximum of 15 minutes, 1 hour, 2 hours, or 4 hours. In each case
the decision of the majority was applied, and this decision was defined as the real degree
of urgency. If there were multiple majority decisions, the panel reviewed the case until
consensus was reached.

Data collection and definitions
The following patient data were collected from case report forms, ED reports, and
electronic hospital information systems: patient demographics, mode of arrival, triage
ratings by the triage nurse, urgency classification by the expert panel of ED physicians, ED
resources used, hospital admission rates (including death), and LOS. LOS was defined as
the time in minutes from registration to discharge or admission. Death in the ED was coded as a death and considered as a hospital admission.

The number of resources was counted in accordance with the ESI (version 3) definitions. Resources used included labs, electrocardiogram, radiology, specialist consultation, intravenous (IV) fluids or hydration, IV or intramuscular medication, simple procedures, and complex procedures. For each patient we documented if one of these resources was used or not.

**Data Analysis**

Descriptive statistics with continuous data are presented either as means with standard deviation (±SD) or as medians, based on the distribution of the data. Categorical data are presented as the percentage frequency occurrence. p values <0.05 were considered to indicate a statistically significant difference. Differences in distribution of urgency levels were tested by means of the Friedman’s test. In all 900 triaged patients, we determined the number of resources used to diagnose the patient, the number of admissions, and LOS in the ED. To evaluate the relationship between triage classification and these aspects, the Spearman’s correlation coefficient was calculated.

Of the subset (patients treated only by emergency physicians), the data of the reference standard were entered into a text file and imported into SPSS, version 16.0 (SPSS Inc., Chicago, IL), for statistical analysis. To determine the validity of all three systems in patients treated by an emergency physician, we compared the ESI and the MTS triage classifications with the reference standard as both five-level and four-level systems. The latter was achieved by combining the urgency levels 4 and 5. The percentage of patients who were under- or overtriaged was calculated. The sensitivity, specificity, predictive values and likelihood ratios and their 95% confidence intervals (CIs) for each of the five urgency levels were calculated. Sensitivity and specificity were defined in terms of correct or overtriage classification, and undertriage was defined as a misclassification.

**RESULTS**

A total of 900 patients were triaged. Of these, 10 patients were lost due to missing ED or triage notes, leaving 890 patients for analysis. Complete triage notes were available for 875 patients (97%) triaged using the ISS, 876 (97%) using the ESI, and 872 (97%) using the MTS.

**Patient characteristics**

The characteristics for all 900 patients are presented in Table 1. Patients referred by the GP to the specialist (mean ± SD age = 48 years ± 27 years) were significantly older than self-referred patients (mean ± SD age = 33 years ± 20 years; p < 0.001). Patients arriving by ambulance (mean ± SD age = 54 years ± 28 years) were significantly older than patients
arriving by private vehicle (mean ± SD age = 34 years ± 21 years; p < 0.001). In self-referred patients, no significant differences in age were found between patients treated by the ED physician only and those referred from ED physician to a specialist.

### Distribution of urgency levels

In all patients available for analysis, the number of patients in each urgency level in each of the triage systems is shown in Figure 1. When the three triage systems were compared as four-level systems, the distribution of urgency levels was shown to be significantly different (Friedman test, p<0.006). The same was true when the ESI and the MTS were compared as five-level systems (Friedman test, p<0.001). In the ISS, more patients were scored “very urgent” than in the ESI and the MTS. Furthermore, according to the MTS, significantly fewer patients belonged to level 5 than according to the other systems.

### Table 1 Patient characteristics

<table>
<thead>
<tr>
<th></th>
<th>All triaged patients (n = 900)</th>
<th>Patients only seen by ED physician (n = 428)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>475</td>
<td>53</td>
</tr>
<tr>
<td>Female</td>
<td>421</td>
<td>47</td>
</tr>
<tr>
<td>Age, yr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (±SD)</td>
<td>37</td>
<td>±23</td>
</tr>
<tr>
<td>Median (range)</td>
<td>36</td>
<td>0-102</td>
</tr>
<tr>
<td>IQR (25-75)</td>
<td>19-53</td>
<td></td>
</tr>
<tr>
<td>Age distribution, yr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 15</td>
<td>163</td>
<td>18</td>
</tr>
<tr>
<td>15 – 30</td>
<td>212</td>
<td>24</td>
</tr>
<tr>
<td>30 – 45</td>
<td>207</td>
<td>23</td>
</tr>
<tr>
<td>45 – 60</td>
<td>170</td>
<td>19</td>
</tr>
<tr>
<td>60 – 75</td>
<td>82</td>
<td>9</td>
</tr>
<tr>
<td>&gt; 75</td>
<td>61</td>
<td>7</td>
</tr>
<tr>
<td>Mode of arrival</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private vehicle or otherwise</td>
<td>784</td>
<td>87</td>
</tr>
<tr>
<td>By ambulance</td>
<td>115</td>
<td>13</td>
</tr>
<tr>
<td>Mode of referral</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-referral</td>
<td>664</td>
<td>74</td>
</tr>
<tr>
<td>Only seen by EP (ED physician)</td>
<td>428</td>
<td></td>
</tr>
<tr>
<td>Referred (by EP) to specialist</td>
<td>236</td>
<td></td>
</tr>
<tr>
<td>Referred by GP or otherwise</td>
<td>231</td>
<td>26</td>
</tr>
<tr>
<td>Number of admissions</td>
<td>162</td>
<td>18</td>
</tr>
<tr>
<td>Number of patients triaged out of hospital during measurement period</td>
<td>27</td>
<td></td>
</tr>
</tbody>
</table>

Data are reported as n (%) unless otherwise specified. GP = General Physician; IQR = interquartile range.
Number of resources

In 890 patients available for analysis, the number of resources was strongly associated with the urgency level in all triage systems. The mean number of resources by urgency level and triage system is presented in Figure 2.
**Number of admissions**

In 890 patients, the number of patient admissions was strongly associated with the urgency level in all triage systems. The distribution of admission by urgency level and triage system is presented in Figure 3.

![Graph showing patient admission rates by urgency level and triage system in 890 patients.](image)

Figure 3. Patient admission rates by urgency (Urg) level and triage system in 890 patients. Spearman’s correlation coefficient: ISS = 0.396; ESI = 0.379; MTS = 0.398; p < 0.001. ESI = Emergency Severity Index; ISS = informally structured system; MTS = Manchester Triage System.

**LOS**

In 890 patients, the length of ED stay was strongly associated with the urgency level in all triage systems. The median LOS per urgency level and triage system is presented in Figure 4. Patients in the highest urgency levels had the longest LOS, except for patients assigned to urgency level 1.

**Reference standard**

Of 890 patients, a total of 428 (48%) were treated by an ED physician only. Seven forms were incomplete and were therefore excluded. Statistically significant differences were found in the percentages of under- and overtriage when all systems were compared as four-level systems. The percentages of undertriage were 8.3% for the ISS, 13.5% for ESI, and 11.2% for the MTS. The highest agreement (64.8%) with the reference standard was found for the ISS, while the highest overtriage (29%) was found for the MTS (Figure 5). When comparing ESI and MTS as five-level systems with the reference standard, agreement decreased and overtriage increased, while significant differences remained.
Figure 4. Median LOS by urgency level and triage system in 890 patients. Pearson’s correlation: ISS = –0.264; ESI = –0.339; MTS = –0.260; p < 0.001. ESI = Emergency Severity Index; ISS = informally structured system; LOS = length of stay; MTS = Manchester Triage System.

Figure 5. Agreement with the reference standard per triage system (n = 421). Reference standard determined retrospectively by an expert panel using all available information including the final diagnosis. ESI 4-level = ESI as a four-level system; levels 4 and 5 combined. MTS 4-level = MTS as a four-level system; levels green and blue combined. Spearman’s correlation coefficient with the reference standard: ISS = 0.466; ESI 4-level = 0.276; and MTS 4-level = 0.272; p<0.001. ESI 5-level and MTS 5-level = ESI and MTS as five-level systems. Spearman’s correlation coefficient with the reference standard: ESI 5-level = 0.172; MTS 5-level = 0.240; p<0.001. ESI = Emergency Severity Index; ISS = informally structured system as a four-level system; MTS = Manchester Triage System.
Details of the sensitivity, specificity, predictive value, and likelihood ratio for each of the five urgency levels are shown in Table 2. Overall, sensitivity and positive predictive values were low for all urgency levels in each system, whereas specificity and negative predictive values were over 95% in urgency levels 1 and 2. The five-level ESI and MTS systems showed significant differences in sensitivity and specificity only in urgency level 4.

<table>
<thead>
<tr>
<th>Cutoff</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR- (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urgency 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISS</td>
<td>40 (5-85)</td>
<td>98 (97-99)</td>
<td>20 (3-56)</td>
<td>99 (98-100)</td>
<td>20.8 (1.7-142)</td>
<td>0.61 (0.15-0.98)</td>
</tr>
<tr>
<td>ESI</td>
<td>NA</td>
<td>100 (99-100)</td>
<td>NA</td>
<td>99 (97-100)</td>
<td>NA</td>
<td>1 (0.5-1)</td>
</tr>
<tr>
<td>MTS</td>
<td>17 (4-64)</td>
<td>100 (99-100)</td>
<td>100</td>
<td>99 (97-100)</td>
<td>NA</td>
<td>0.83 (0.4-0.97)</td>
</tr>
<tr>
<td>Urgency 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISS</td>
<td>47 (29-65)</td>
<td>92 (89-94)</td>
<td>32 (19-47)</td>
<td>95 (93-97)</td>
<td>5.7 (2.6-11.5)</td>
<td>0.58 (0.37-0.8)</td>
</tr>
<tr>
<td>ESI</td>
<td>36 (20-55)</td>
<td>95 (93-97)</td>
<td>40 (23-59)</td>
<td>95 (92-97)</td>
<td>7.8 (2.8-19.6)</td>
<td>0.67 (0.46-0.86)</td>
</tr>
<tr>
<td>MTS</td>
<td>34 (19-53)</td>
<td>95 (93-97)</td>
<td>37 (20-56)</td>
<td>95 (92-97)</td>
<td>7.0 (2.5-17.7)</td>
<td>0.69 (0.48-0.88)</td>
</tr>
<tr>
<td>Urgency 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISS</td>
<td>76 (65-85)</td>
<td>73 (68-78)</td>
<td>41 (33-49)</td>
<td>93 (89-95)</td>
<td>2.8 (2.1-3.8)</td>
<td>0.33 (0.20-0.51)</td>
</tr>
<tr>
<td>ESI</td>
<td>50 (39-61)</td>
<td>74 (69-78)</td>
<td>32 (24-40)</td>
<td>86 (82-90)</td>
<td>1.9 (1.3-2.8)</td>
<td>0.68 (0.50-0.89)</td>
</tr>
<tr>
<td>MTS</td>
<td>60 (49-71)</td>
<td>66 (61-71)</td>
<td>30 (23-38)</td>
<td>87 (83-91)</td>
<td>1.8 (1.3-2.5)</td>
<td>0.60 (0.41-0.83)</td>
</tr>
<tr>
<td>Urgency 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISS</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>ESI</td>
<td>63 (56-71)*</td>
<td>44 (38-50)*</td>
<td>36 (30-42)*</td>
<td>70 (63-77)</td>
<td>1.1 (0.9-1.4)</td>
<td>0.84 (0.58-1.17)</td>
</tr>
<tr>
<td>MTS</td>
<td>98 (94-100)*</td>
<td>2 (1-5)*</td>
<td>34 (29-39)*</td>
<td>67 (30-93)</td>
<td>1.0 (0.9-1.0)</td>
<td>0.97 (0.09-7.50)</td>
</tr>
</tbody>
</table>

ESI = Emergency Severity Index; ISS = Informally structured system; LR = likelihood ratio; MTS = Manchester Triage System; NA = not applicable; NPV = negative predictive value; PPV = positive predictive value.

* p < 0.05

DISCUSSION

In this study we demonstrated that when investigated in an ED setting, the validity of these three triage systems is similar. Undertriage was seen most frequently when using the ESI system. Furthermore, in all triage systems, higher urgency levels were associated with increased resource use, higher rate of hospitalization, and increased LOS. Based on these results, not one of these systems appears superior. However, we do recommend the use of a formally structured triage system in order to obtain verifiable systematic judgments, transparency, and uniformity in triage.

It is difficult to say which level of sensitivity or specificity is acceptable to conclude that a certain triage system is safe. To reach a high sensitivity (i.e., an acceptable degree of
undertriage), the specificity will be so low that the potential for saving resources would be marginal at best.

In our study we found that the ESI had a much lower sensitivity than the 75% found by Travers et al.\textsuperscript{4} They compared ratings by the triage nurse with the triage decisions taken by two reviewers. These decisions were based on the original triage notes. Also, we found the sensitivity of the MTS to be lower than has been reported in previous studies.\textsuperscript{9,15-17} These differences may be explained by the fact that most of these studies were performed in selected patient groups.\textsuperscript{9,15,16} In contrast, we studied the systems’ validity in patients treated by an ED physician only, of whom only 30 (7%) were triaged to urgency levels 1 or 2.

Using both formally-structured systems, the majority of patients deemed by the expert panel to belong in urgency levels 1 or 2 were undertriaged. This difference between the reference standard and the formally structured systems may be due to the fact that the expert panel of ED physicians knew what happened to the patient. Therefore, they may have retrospectively evaluated such patients as being less (or more) urgent than they would otherwise have done using a formally structured system and before knowing the outcome.

Of the total patient group, only a few patients were categorized to level 1 in accordance with the ESI and MTS, because patients triaged in the prehospital setting by ambulance paramedics, and those who met the current criteria for treatment in the trauma room, were treated immediately. For this reason, these patients were not present in our sample.

In addition, it is possible that the ISS and the reference standard contained other priorities deemed to be more important and that caused the users to classify a patient to a higher (or lower) level than was the case with the formally structured triage systems. For example, patients with cerebrovascular accidents should be seen immediately, to start thrombolysis as soon as possible. Therefore, in the ISS these patients are classified to the highest level of urgency. Using the formally structured systems these patients would be classified as level 2 (ESI) or orange (MTS). To state the issue clearly, decisions made following the ISS are not transparent; relevant information remains implicit and cannot be retrieved from the ED form.

In general, predicting hospital admission is difficult.\textsuperscript{17} Nevertheless, the lowest levels of urgency determined by our study were associated with very low admission rates. The predictions of resource use, hospital admission, and LOS per urgency level in each of the three triage systems were consistent with research from the ESI group in Boston, as well as other results from within the Netherlands.\textsuperscript{13,18-20}

**LIMITATIONS**

The limitations of our study design were first, the lack of standardized criteria for determining the reference standard. However, even when working with explicit criteria a high rate of disagreement among experts has been demonstrated.\textsuperscript{21} For the purpose of our study, we tried to compensate for possible disagreement by forming a panel of seven
experienced ED physicians and settling for a majority consensus. As they did not see the patients, the expert panel had to rely on a verbal description. On the other hand, having the diagnosis and diagnostic test results offered the opportunity to identify potentially urgent patients who were not identified by the triage systems.\textsuperscript{22}

Second, for pragmatic reasons we chose to triage at random on different days and between 12 noon and 10 PM. The distribution of urgency levels within this time frame might differ from other times of the day. However, patient age, sex and types of condition were similar to a consecutive series of patients in previous research in our department. Additionally, the distribution of urgency levels in our center (according to the ESI and the MTS classifications) was consistent with previous Dutch reports on both systems.\textsuperscript{18,23}

Third, the fact that three triage systems were applied sequentially in every patient might have biased the results. If the MTS, which was always used last, had received the least attention, it would have shown the lowest validity, but this was not the case. Hence, we do not think this has been an important source of bias.

Fourth, data collected from one center may merely reflect that particular institution’s practice. In a previous study comparing inter- and intraobserver agreement between inexperienced and experienced triage nurses, agreement was found to be the same for the ESI triage nurses, but lower than in experienced MTS triage nurses. However, overall the MTS showed a greater inter- and intraobserver agreement than the ESI.\textsuperscript{8} Still, if the use of nurses extensively experienced in ED practice but inexperienced in triage did result in underestimation of the sensitivity and specificity, this would have affected the results of both formally structured systems equally.

Finally, in our ED we saw a relatively high number of referrals by GPs and percentage of patients sent for specialty services. If this would substantially influence the distribution of the triage levels allocated to these patients, i.e., would lead to generally higher or lower triage levels, this could influence the predictive values we found for the triage systems. In turn, this may be of influence on the generalization of our results. Therefore, the predictive values were determined for the subset of patients treated by the ED physician only. Hence, the reader should check whether this distribution in his or her own ED is similar or dissimilar to ours.

**CONCLUSION**

Informally and formally structured triage systems appear to have equal validity, although the Emergency Severity Index tends to undertriage patients. To ensure transparency and uniformity, a verifiable, formally structured triage system for ED patients is advocated.
REFERENCES


Observer agreement of the Manchester Triage System and the Emergency Severity Index: a simulation study

MN Storm-Versloot
DT Ubbink
V Chin a Choi
JSK Luitse

Emerg Med J 2009;26;556-560
ABSTRACT

Objectives: To compare inter and intra-observer agreement of the Manchester Triage System (MTS) and the Emergency Severity Index (ESI).

Methods: 50 representative emergency department (ED) scenarios derived from actual cases were presented to 18 ED nurses from three different hospitals. Eight of them were familiar with MTS, six with ESI and four were not familiar but trained in both systems. They independently assigned triage scores to each scenario according to the triage system(s) they were familiar with. After 4-6 weeks the same nurses again judged the scenarios in a different order. Unanimity in judgement, and unweighted and quadratic-weighted kappas were calculated.

Results: Unanimity in judgement for MTS was 90% and for ESI 73%. One-level disagreement was found in 8% and 23% of the cases, respectively. Interobserver unweighted kappas were 0.76 (95% CI 0.68 to 0.83) for MTS and 0.46 (95% CI 0.37 to 0.55) for ESI. Quadratic-weighted kappas were 0.82 (95% CI 0.74 to 0.89) and 0.73 (95% CI 0.64 to 0.83), respectively. At 4-6 weeks, one-level intra-observer disagreements were 10% and 22%, and 2-level disagreement 1% and 2%, respectively. Intra-observer unweighted kappas were 0.84 (95% CI 0.73 to 0.94) for MTS and 0.65 (95% CI 0.59 to 0.72) for ESI.

Conclusion: Using paper-based clinical scenarios, MTS was found to have a greater inter and intra-observer agreement than ESI.
INTRODUCTION

Because the demand for emergency services outpaces available resources, emergency department (ED) triage systems face increasing scrutiny. Longer waits for care make the use of reliable, valid triage systems imperative to patient safety. Triage is defined as the initial clinical sorting process in hospital ED. ED generally use some form of triage, either formal or informal, in order to assess the patient’s clinical needs and priority of care. Informal triage systems rely on intuition and clinical experience of the ED nurse and decisions made cannot be tested afterwards. Formal triage systems offers more transparency, but depend on its reliability. This is usually expressed by means of a weighted kappa statistic. However, unanimity in judgement is rarely mentioned, which best illustrates triage uniformity.

Worldwide, four formal, five-level triage systems exist. In The Netherlands, these four systems have been critically appraised by the Dutch National Institute of Quality Control in Healthcare (CBO). Only the Manchester Triage System (MTS) and the Emergency Severity Index (ESI) were found to be applicable in our country. Therefore, both other systems are not in use in The Netherlands.

MTS comprises 52 flowcharts based on patient complaints. The presenting complaint is indicative of the severity and defines which flowchart is to be followed. Each flowchart is based on a five-step decision process that uses discriminators at each step to assign patients to one of the five triage levels. Although existing literature does not allow evaluation of validity, under and overtriage, and interobserver agreement of this system, the CBO has adopted MTS in the current guideline on triage at the ED.

ESI uses one algorithm, with ratings ranging from level 1 (the most acutely ill patients) to level 5 (the least resource-intensive patients). For patients not meeting ESI level 1 or 2 criteria, the triage nurse estimates the number of resources needed to discharge the patient from the ED. The resource usage was shown to correlate well with the different triage levels and showed a high interobserver agreement with quadratic weighted kappas of 0.68–0.89.

Both MTS and ESI seem to be useful, but no studies have compared the diagnostic validity and reliability of these systems. In this study, we focused on the reliability of both systems. We determined and compared the inter and intra-observer agreement by investigating judgement unanimity and the agreement of both triage systems with a reference standard.

METHODS

Study design, setting and population

This comparative clinical survey was performed at the Academic Medical Center (AMC) and Onze Lieve Vrouwe Gasthuis (OLVG) in Amsterdam, and the Medical Spectrum Twente (MeSpTw) in Enschede, The Netherlands. The AMC is an urban tertiary care university...
hospital with a level 1 trauma centre. The ED sees almost 31,000 patients annually. The overall admission rate is approximately 18%. Approximately 16% of the patients are younger than 16 years. For triaging patients, ED nurses use an informal system.

The OLVG is an urban teaching hospital with a level 2 trauma centre and approximately 42,000 ED visits a year. The admission rate is approximately 10% and about 16% of the patients are younger than 16 years. Since 2003, ED nurses trained in ESI use this system for triaging their patients.

The MeSpTw is an urban teaching hospital with a level 1 trauma centre in the east of The Netherlands with almost 32,000 ED visits a year at two locations. The admission rate is approximately 20%, but the number of patients younger than 16 years is not known. Since 2003, ED nurses trained in MTS use this system for triaging their patients.

Methods of measurement
Between November and December 2005, a total of 900 ED cases were prospectively collected at the AMC. For the purpose of triaging the patients according to MTS and ESI, six ED nurses received a 6-h combination of didactic and practical training for each system according to national standards. At random days of the week, between 12:00 and 22:00 hours, all consecutive patients entering the ED were triaged. Patients who were triaged by the ambulance staff before presentation at the ED and who met the criteria for treatment in the shock room according to current guidelines were not triaged again, but were classified as “red” (MTS) or “level 1” (ESI) patients. All patients gave oral informed consent for the study as the ethics review board waived the requirement for written informed consent.

Based on the 900 triaged patients, the distribution of the urgency levels was assessed. According to this distribution a representative sample of 50 cases was chosen. These cases were converted into written patient scenarios, using the documented triage notes and ED forms and checked by three nurses from the other contributing centres for comprehensible interpretation, missing data, and feasible judgement. Scenarios included age and gender, chief complaint, patient’s appearance, pain as expressed by the patient and scored by the nurse, history of presenting illness and vital signs such as pulse rate, blood pressure, temperature and oxygen saturation if appropriate. An example is given in box 1.

Box 1: Example of an ED patient scenario.

A 76-year-old male is transported by ambulance to the ED. Yesterday evening he collapsed at home and according to his wife he did not want to get up. The whole night she spent with him on the floor. His consciousness was diminished. He was aphasic and had a paresis of his right arm. His vital signs were heart rate 116 beats/minute, blood pressure 111/73 mm Hg, body temperature 38.4ºC.
For MTS, the scenario writer assigned an urgency level to each scenario. Two independent expert nurses from the MeSpTw did the same. Disagreement existed in three out of 50 scenarios. These were adjudicated by another nurse. The final judgement was regarded as the reference standard. For ESI levels 1 or 2, the reference standard was determined by signs of a critical condition of the patient using the flowchart. For the remaining levels the actual ED resources used (laboratory testing, ECG, radiology, speciality consultations, intravenous fluids or hydration, intravenous or intramuscular medication, simple and complex procedures).

Eight ED nurses of the MeSpTw assigned urgency levels to the 50 patient scenarios using MTS, six ED nurses of the OLVG using ESI and four ED nurses of the AMC using both systems. No discussion was allowed during the assignment. The same judging procedure was repeated with the same scenarios in a different order after an interval of 4-6 weeks. The nurses were kept unaware of their own original assignment and were not allowed to discuss their assignments. This was achieved by supervising the nurses during the assignment sessions.

Data analysis
Inter and intra-observer agreement was calculated using AGREE version 7 (Scienceplus Group, Groningen, The Netherlands), a software program dedicated to calculate kappa values. Kappa values lie between -1 and 1. A kappa value of above 0.8 is called “very good”, between 0.8 and 0.6 “good”, between 0.6 and 0.4 “moderate” and below 0.4 “poor”.\(^\text{19}\) To assess inter and intra-observer agreement among ED nurses and the agreement with the reference standard, pairwise kappa values were calculated, computing an unweighted and quadratic-weighted group kappa for several fixed observers.\(^\text{20,21}\) The group kappa gives a measure of average agreement between all the pairs in excess of chance.\(^\text{22}\) Differences in group kappa values between the hospitals as well as between less or more than 5 years ED experience in the MeSpTw hospital were calculated with Agree.

Because of possible shortcomings of kappa statistics,\(^\text{23,24}\) we also assessed the unanimity of judgements for the first judgement. This was defined as the percentage of scenarios given the same urgency level by all observers and the total number of judgements by all observers given the same urgency level.

Differences in nurse characteristics were analysed using SPSS version 12.

RESULTS
The mean age of all nurses was 39 years (SD 6.7). They had a mean of 9 years (SD 6.1) of ED experience. Nurse characteristics did not differ significantly among the hospitals, except for triage experience: Nurses from the OLVG and MeSpTw had 3 years of triage experience, whereas those of the AMC had none.
MTS scores as given by all observers were unanimous in 23 (46%) of all 50 scenarios; ESI scores in only five (10%) of all scenarios. A one-level urgency disagreement with respect to the triage classification according to MTS occurred in 22 (44%) scenarios and in 34 (68%) scenarios judged by ESI. Two-level disagreement occurred in five (10%) and nine (18%) scenarios, respectively. Three-level disagreement occurred in two (4%) scenarios, but only when using ESI.

In total, 594 (99%) of the 600 (50 scenarios judged by 12 nurses) urgency levels were obtained according to MTS and 498 (99%) of the 500 (50 scenarios judged by 10 nurses) according to ESI. A total of 534 (90%) of MTS-judged urgency levels were unanimous and 363 (73%) of ESI judgements. One-level disagreement occurred in 49 (8%) and 113 (23%) of the judgements, respectively, and two-level disagreements in 10 (2%) and 20 (4%).

Interobserver agreement by using the unweighted kappa was better with MTS and with more triage experience using MTS, but disappeared using the weighted kappa (table 1). For MTS no differences were found in years of ED experience.

Because one of the AMC nurses did not perform a second judgement, intra-observer analysis was calculated for 11 and nine nurses, respectively. Overall agreement between the first and second judgement was 89% for MTS (table 2) and 75% for ESI (table 3). Nearly all disagreements occurred within one level for both systems. Intra-observer agreement followed the same trend as interobserver agreement (table 4).

Compared with the reference standard, a 5% undertriage rate was found using MTS and 13% using ESI. For MTS (dis)agreement between the ED nurses’ judgements and the reference standard followed the same trend as the intra en interobserver agreement (table 5). For ESI, agreement was lower than the intra-observer agreement, but the majority of disagreements was still within one level (table 6).

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Interobserver agreement for MTS and ESI, for each hospital and experience level of ED nurses at the MeSpTw, based on the first judgement.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unweighted kappa (95% CI)</td>
</tr>
<tr>
<td>MTS (n=12)</td>
<td>0.76* (0.68 to 0.83)</td>
</tr>
<tr>
<td>MeSpTw (n=8)</td>
<td>0.85† (0.77 to 0.93)</td>
</tr>
<tr>
<td>&lt; 5 years experience</td>
<td>0.87 (0.78 to 0.96)</td>
</tr>
<tr>
<td>&gt; 5 years experience</td>
<td>0.81 (0.71 to 0.91)</td>
</tr>
<tr>
<td>AMC (n=4)</td>
<td>0.60† (0.48 to 0.73)</td>
</tr>
<tr>
<td>ESI (n=10)</td>
<td>0.46* (0.37 to 0.55)</td>
</tr>
<tr>
<td>OLVG (n=6)</td>
<td>0.48 (0.38 to 0.57)</td>
</tr>
<tr>
<td>AMC (n=4)</td>
<td>0.41 (0.30 to 0.53)</td>
</tr>
</tbody>
</table>

* Significance level p< 0.05, between Manchester Triage System (MTS) and Emergency Severity Index (ESI); † Significance level p< 0.05, between Medical Spectrum Twente, Enschede (MeSpTw) and Academic Medical Centre, Amsterdam (AMC). ED, emergency department. OLVG: Onze Lieve Vrouwe Gasthuis, Amsterdam.
Table 2 MTS: Comparison of ED-nurses’ triage judgements between the first and second judgement

<table>
<thead>
<tr>
<th>ED nurse: second judgement</th>
<th>ED-nurse: first judgement</th>
<th>Red</th>
<th>Orange</th>
<th>Yellow</th>
<th>Green</th>
<th>Blue</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td></td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Orange</td>
<td></td>
<td>0</td>
<td>107</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>113</td>
</tr>
<tr>
<td>Yellow</td>
<td></td>
<td>0</td>
<td>5</td>
<td>173</td>
<td>19</td>
<td>2</td>
<td>199</td>
</tr>
<tr>
<td>Green</td>
<td></td>
<td>0</td>
<td>3</td>
<td>10</td>
<td>195</td>
<td>5</td>
<td>213</td>
</tr>
<tr>
<td>Blue</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1</td>
<td>117</td>
<td>189</td>
<td>221</td>
<td>16</td>
<td>544</td>
</tr>
</tbody>
</table>

Overall agreement Manchester Triage System (MTS): 485/544 (89.2%); one-level disagreement: 52/544 (9.6%); two-level disagreement: 7/544 (1.3%). ED, emergency department.

Table 3 ESI: Comparison of ED nurses’ triage judgements between the first and second judgement

<table>
<thead>
<tr>
<th>ED nurse: second judgement</th>
<th>ED-nurse: first judgement</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Level 5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td>0</td>
<td>90</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>102</td>
</tr>
<tr>
<td>Level 3</td>
<td></td>
<td>0</td>
<td>11</td>
<td>132</td>
<td>10</td>
<td>4</td>
<td>157</td>
</tr>
<tr>
<td>Level 4</td>
<td></td>
<td>0</td>
<td>1</td>
<td>26</td>
<td>31</td>
<td>22</td>
<td>80</td>
</tr>
<tr>
<td>Level 5</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>21</td>
<td>84</td>
<td>108</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>0</td>
<td>103</td>
<td>169</td>
<td>64</td>
<td>111</td>
<td>447</td>
</tr>
</tbody>
</table>

Overall agreement Emergency Severity Index (ESI): 337/447 (75.4%); one-level disagreement: 99/447 (22.1%); two-level disagreement: 9/447 (2.0%); three-level disagreement: 2/447 (0.4%). ED, emergency department.

Table 4 Intra-observer agreement for MTS and ESI, for each hospital and experience level of ED-nurses at the MeSpTw

<table>
<thead>
<tr>
<th></th>
<th>Unweighted kappa (95% CI)</th>
<th>Quadratic-weighted kappa (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTS (n=11)</td>
<td>0.84* (0.73 to 0.94)</td>
<td>0.90 (0.83 to 0.96)</td>
</tr>
<tr>
<td>MeSpTw (n=8)</td>
<td>0.91 (0.83 to 0.98)</td>
<td>0.94 (0.88 to 0.99)</td>
</tr>
<tr>
<td>&lt; 5yrs experience</td>
<td>0.91 (0.74 to 1.00)</td>
<td>0.94 (0.83 to 1.00)</td>
</tr>
<tr>
<td>&gt; 5yrs experience</td>
<td>0.91 (0.78 to 1.00)</td>
<td>0.93 (0.83 to 1.00)</td>
</tr>
<tr>
<td>AMC (n=3)</td>
<td>0.65 (0.30 to 1.00)</td>
<td>0.79 (0.59 to 0.99)</td>
</tr>
<tr>
<td>ESI (n=9)</td>
<td>0.65* (0.59 to 0.72)</td>
<td>0.85 (0.80 to 0.90)</td>
</tr>
<tr>
<td>OLVG (n=6)</td>
<td>0.65 (0.56 to 0.74)</td>
<td>0.84 (0.78 to 0.90)</td>
</tr>
<tr>
<td>AMC (n=3)</td>
<td>0.66 (0.38 to 0.95)</td>
<td>0.88 (0.70 to 1.00)</td>
</tr>
</tbody>
</table>

* Significance level p< 0.05, between Manchester Triage System (MTS) and Emergency Severity Index (ESI). AMC, Academic Medical Centre, Amsterdam; ED, emergency department; MeSpTw, Medical Spectrum Twente, Enschede; OLVG, Onze Lieve Vrouwe Gasthuis, Amsterdam.
LIMITATIONS

The limitations of our study are in the first place the use of standardised and abstracted case scenarios. Case scenarios are artificial, because they do not show the non-verbal clues from the live interview. Therefore, we used prospectively triaged patients and kept subjective information and physical appearance in the scenarios. Intra-observer reliability can only be determined if the first and second judgements are based on identical information, as only scenarios can provide.

Second, the strategy for selecting the reference standard was different for both triage systems. Both systems have a different conceptual foundation. We therefore used the judgement of expert ED nurses as reference standard for MTS in order to determine if the system was used correctly. In contrast, ESI not only scores patient urgency, but also required resources. Therefore, we determined the resources actually used. The number of resources needed depends on local hospital standards. Because of the close collaboration

Table 5 MTS: Comparison between ED-nurses’ triage judgement and the reference standard based on the first judgement

<table>
<thead>
<tr>
<th>ED nurse judgement</th>
<th>Reference standard</th>
<th>Red</th>
<th>Orange</th>
<th>Yellow</th>
<th>Green</th>
<th>Blue</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td></td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Orange</td>
<td></td>
<td>0</td>
<td>111</td>
<td>5</td>
<td>8</td>
<td>0</td>
<td>124</td>
</tr>
<tr>
<td>Yellow</td>
<td></td>
<td>0</td>
<td>5</td>
<td>199</td>
<td>12</td>
<td>2</td>
<td>218</td>
</tr>
<tr>
<td>Green</td>
<td></td>
<td>0</td>
<td>1</td>
<td>22</td>
<td>208</td>
<td>2</td>
<td>233</td>
</tr>
<tr>
<td>Blue</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>0</td>
<td>120</td>
<td>227</td>
<td>235</td>
<td>12</td>
<td>594</td>
</tr>
</tbody>
</table>

Agreement Manchester Triage System (MTS): 526/594 (88.6%); one-level disagreement: 56/594 (9.4%); two-level disagreement: 12/594 (2.0%); undertriage: 32/594 (5.4%); overtriage: 36/594 (6.1%); unweighted kappa: 0.84 (0.80 to 0.87); quadratic weighted kappa: 0.87 (0.70 to 1.00). ED, emergency department.

Table 6 ESI: Comparison between ED-nurses’ triage judgement and the reference standard based on the first judgement

<table>
<thead>
<tr>
<th>ED-nurse judgement</th>
<th>Reference standard</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Level 5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td>0</td>
<td>95</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>111</td>
</tr>
<tr>
<td>Level 3</td>
<td></td>
<td>0</td>
<td>29</td>
<td>60</td>
<td>72</td>
<td>20</td>
<td>181</td>
</tr>
<tr>
<td>Level 4</td>
<td></td>
<td>0</td>
<td>4</td>
<td>16</td>
<td>31</td>
<td>39</td>
<td>90</td>
</tr>
<tr>
<td>Level 5</td>
<td></td>
<td>0</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>98</td>
<td>116</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>0</td>
<td>130</td>
<td>90</td>
<td>118</td>
<td>160</td>
<td>498</td>
</tr>
</tbody>
</table>

Agreement Emergency Severity Index (ESI): 284/498 (57.0%); one-level disagreement: 170/498 (34.1%); two-level disagreement: 39/498 (7.8%); three-level disagreement: 5/498 (1.0%); undertriage: 67/498 (13.5%); overtriage: 147/498 (29.5%); unweighted kappa: 0.43 (0.38 to 0.49); quadratic weighted kappa: 0.71 (0.61 to 0.82). ED, emergency department.
and similarity (patient population, emergency physician training, protocol use and available facilities) between AMC and OLVG hospitals this is not a likely confounder. For the purpose of this study we did not want to compare the diagnostic validity of both systems when the same reference standard should be used.

Third, we did not include level 1 urgency patients. The majority of these patients are already triaged before arrival at the hospital by ambulance staff and are transported directly to the shock room. Moreover, the inclusion of these level 1 cases would have overestimated the kappa values found in each triage group without altering differences between the two groups.

Fourth, the number of observers judging the scenarios in the AMC was rather small. Therefore, the difference in MTS scores found between the hospitals for inter and intra-observer agreement may have occurred due to outliers in the AMC. We did not exclude these outliers, because it reflects actual clinical practice. For ESI, we did not find these effects.

Finally, we used the first version of the MTS and the third version of the ESI, although presently an updated version of both systems exists. At the time of the study, these versions were not available for Dutch hospitals.

DISCUSSION

By means of written case scenarios, we found the MTS to show a high degree of triage unanimity and a good agreement among ED nurses and when compared with a reference standard. For the ESI, the degree of unanimity was lower, but differences were usually not larger than one urgency level.

Our study is the first to compare triage agreement by means of MTS and ESI while distinguishing ED nurses with and without triage experience. Most studies on agreement only report the quadratic-weighted kappa, if at all specified, but rarely exact agreement or unanimity. Unanimity results are more conservative, while weighted kappa values appreciate near disagreement. This seems right for the lower urgency levels, but a one-level difference in the higher urgency levels can delay treatment, which is potentially dangerous to the patient. We did report weighted kappa values to allow comparison between our results and those from other studies.

Few studies report the agreement between triage judgements by ED nurses and a reference standard, mostly based on case scenarios. Our results for ESI are comparable with existing literature, showing good weighted kappa values, ranging from 0.68 to 0.89.11,13-17 Few studies focus on MTS, but an unweighted kappa of 0.60 for the inexperienced hospital in our study matches a value of 0.63 as reported in the guideline on triage.2 The remaining kappa values we found were much better.

We found a significant difference using the unweighted kappa, which disappeared by using the weighted kappa, both for inter and intra-observer agreement and for the
agreement with the reference standard. This is because the weighted kappa accepts one-level disagreement as some form of agreement, which mitigates the differences. A one-level disagreement was also most common in previous studies.\textsuperscript{11,13-16} Unfortunately, we cannot compare the magnitude of disagreement for the inter and intra-observer agreement. We therefore recommend that in future studies both exact agreement and weighted kappa values are presented.

Some studies report undertriage rates varying from 9% to 12%. In these studies the judgements were compared with the “true” urgency ESI triage level, determined by an expert panel.\textsuperscript{14,16} If undertriage occurs, potentially seriously ill patients may be triaged as non-urgent, resulting in an increasing risk of adverse outcomes for these patients. We found an undertriage rate of 13% (67 judgements) for ESI. Of these, only 14 judgements spread over five scenarios showed a two-level disagreement. This may seem serious, but to determine the actual consequences of these undertriage judgements the diagnostic validity of the system also has to be assessed. The relatively low unanimity and high disagreement of ESI might be because the determination of urgency depends on implicit knowledge rather than explicit flowcharts.

For MTS only two judgements in two scenarios showed a two-level disagreement. Cooke and Jinks\textsuperscript{6} reported that almost 20% of incorrect classifications in critically ill patients were due to non-adherence to the MTS guideline; most errors were because of training problems rather than the triage system. We found less non-adherence, but we did not restrict our cases to critically ill patients. All nurses were trained according to standard procedures before using MTS. In a computer-aided environment, adherence is easier to achieve. Unanimity of scoring with MTS can reach 68% without and 96% with computerised decision support.\textsuperscript{2} In our study judgement was performed without computer support, but the flow charts could be consulted. We found a fairly high number of unanimous judgements (90%), although some difference occurred between the AMC (84%) and the MeSpTw (94%). Apparently, nurses need to learn how to use the system correctly. Using a computer aid could help overcome the nurses’ tendency to follow their own line of reasoning in interpreting MTS flowcharts.

CONCLUSIONS

We conclude that MTS has very good agreement and a high unanimous classification rate, whereas ESI has only moderate to good results. For MTS, agreement was not influenced by the ED nurses’ experience, but appeared to be affected by the level of experience with the system. Determination of triage system reliability is a necessary step in establishing its usefulness and is pivotal in any attempt to measure performance in emergency medicine. Beyond triage reliability, as was investigated here, diagnostic validity should be determined of both systems by comparing the triage classifications with an identical reference standard.
REFERENCES


The effect of a Minimal Intervention Strategy in addition to Nicotine Replacement Therapy to support smoking cessation in cardiovascular outpatients: a randomized clinical trial
ABSTRACT

Background: Smoking is an important risk factor for recurrent events in cardiovascular patients. Evidence exists that nicotine replacement therapy (NRT) approximately doubles smoking cessation rates. The minimal intervention strategy (MIS) has been used successfully to assist patients to quit smoking in general practice, and was recently adapted for cardiology inpatients (C-MIS). It is hypothesized that in cardiovascular outpatients the combination of C-MIS and NRT significantly increases the number of quitters compared to NRT alone.

Methods: A randomized clinical trial in 385 smoking patients who attended the cardiovascular outpatient departments in the Academic Medical Centre, Amsterdam for the treatment of atherosclerotic disease. Patients were allocated to either NRT + C-MIS or NRT alone. Self-reported and biochemically validated abstinence rates were measured at 12 months’ follow-up.

Results: Including patients with incomplete follow-up as smokers, abstinence was reported by 19% of the NRT + C-MIS group and 14% of the NRT group [absolute risk reduction (ARR) =0.05; 95% confidence interval (CI) -0.02 to 0.12]. According to biochemical markers, abstinence rates were 28 and 24%, respectively (ARR= 0.04, 95% CI -0.06 to 0.14). Hence, no significant differences between groups were found. The number of cigarettes smoked a day decreased significantly at 12 months: from 21 to 15 a day in the experimental group, and from 21 to 14 in the control group (P< 0.001), but did not differ between groups (P = 0.32).

Conclusions: The effectiveness of a minimal contact intervention was investigated in order to reach as many cardiovascular patients as possible in the setting of outpatient departments. This intervention was not found to be effective.
INTRODUCTION

Although smoking is one of the most important risk factors for recurrent events in cardiovascular patients, smoking behaviour receives only limited attention from treating physicians.\(^1\)\(^-\)\(^3\) This is mainly due to lack of time and inadequate smoking cessation programmes. In 1998 the Health Education Authority in the UK developed evidence-based recommendations\(^4\)\(^,\)\(^5\) regarding smoking cessation, based on evidence from the US Agency for Health Care Policy and Research (AHCPR) guideline\(^6\) and Cochrane Reviews. According to these guidelines, key components of successful cessation programmes are the combination of nicotine replacement therapy (NRT) and supportive care. Prescription of NRT to smokers who attempt to quit should be standard practice, since it increases the odds of quitting approximately 1.5 to 2-fold, regardless of setting.\(^7\)

Supportive care is an additional tool for influencing patients’ smoking behaviour. In the absence of an effective and feasible behavioural smoking cessation programme as a primary prevention tool, the minimal intervention strategy (MIS) was developed. The MIS is an individualized short intervention, and has proven its efficacy and feasibility in general practice in a large, randomized, controlled trial. It resulted in a point prevalence abstinence rate of 11.9%, compared to 3.8% in the control group at 6 months, and 13.4 versus 7.3% at 12 months.\(^8\) The theoretical assumptions of the attitude–social influence–efficacy model\(^9\)\(^,\)\(^10\) and the transtheoretical model\(^11\)\(^,\)\(^12\) (suggesting that patients go through motivational stages before they change their health behaviour) form the basis of the MIS. The intervention is in line with the smoking cessation guidelines of the UK.\(^4\)\(^,\)\(^5\)

These days, the MIS is propagated by several health institutions [e.g. the Dutch Expertise Centre on Tobacco Control (STIVORO), and the Dutch College of General Practitioners (NHG)] to support smoking cessation in specific patient populations. For the purpose of secondary prevention, the MIS programme has been adapted for cardiology inpatients (C-MIS)\(^13\) with nurses largely being responsible for the intervention.\(^13\)\(^,\)\(^14\) Currently, 45% of the 121 Dutch cardiology inpatient wards are using the C-MIS.\(^15\) However, its incremental effect in addition to NRT in cardiovascular outpatients is unknown.

A Cochrane Review\(^16\) also showed that insufficient data are available on the combination of nurse-led interventions and NRT as compared to NRT alone.

The aim of the current study was to contribute to the secondary prevention of cardiovascular disease by attempting to improve smoking cessation rates. We thereby carried out a randomized clinical trial and tested the hypothesis that the combination of C-MIS and NRT, as performed by a nurse in the outpatient clinic, significantly increases the number of quitters as compared to NRT alone.
METHODS

Participants
Consecutive patients were recruited at the outpatient departments of vascular surgery, cardiology and vascular medicine of the Academic Medical Centre, Amsterdam, The Netherlands. Patients were eligible if they were ≥ 18 years old, smoked ≥ 5 cigarettes a day, and suffered from documented peripheral arterial disease (PAD) or coronary artery disease (CAD). Exclusion criteria were having an acute myocardial infarction in the previous month, unstable angina, serious arrhythmia, recent stroke, skin allergy, pregnancy and insufficient comprehension of the Dutch language.

Randomization
Patients just diagnosed with a smoking-related illness may be more receptive to participate in a smoking cessation programme, compared to patients who attend the outpatient clinic for a routine follow-up visit. Therefore, a distinction was made between types of clinic attendance. Patients diagnosed with PAD/CAD within the first three visits were considered ‘newly diagnosed/first visit’. All other potential cardiovascular patients were identified on the basis of the hospital information system, and the presence of PAD/CAD was verified by physicians.

During a regular consultation, eligible patients received advice to quit smoking from their treating physician and were invited to participate in the study. Consenting patients were referred to a nurse practitioner, and received information about the study procedure. Patients were not informed about the behavioural intervention, in order to avoid a Hawthorne effect. After follow-up, patients received a letter containing this withheld information. While patients completed their baseline questionnaire (and signed a written informed consent), nurses randomly assigned patients to either the control or the experimental group. A computerized balanced randomization programme was designed, taking prognostic factors (e.g. clinic attendance, age and gender) into account. Randomization was stratified by clinic attendance and outpatient department (vascular medicine, vascular surgery or cardiology).

Interventions
Following randomization, 8 weeks of free NRT [transdermal nicotine patches: < 20 cigarettes/day (20 cm²=14mg/24 h) or ≥ 20 cigarettes/day (30 cm²=21mg/24 h)] was offered to all patients, accompanied with intensive application instructions from the nurse practitioner. NRT was only presented when patients were planning to quit, because smoking and using NRT simultaneously may cause serious health problems in patients with cardiovascular disease.

Patients in the control group received usual care only, i.e. no additional motivational counselling or self-help materials. In the experimental group, patients were offered the behavioural intervention (C-MIS). In the C-MIS, during a 15–30-min counselling session,
six steps were performed by the nurse practitioner. First, the nurse assessed the patients’ smoking profile, including nicotine dependency and motivation to quit. Based on patients’ readiness to change, the nurse would: increase motivation by stressing the adverse effects of smoking and the benefits of quitting; discuss perceived barriers of quitting; set a date to quit smoking; offer NRT and self-help materials such as brochures and information on support groups; and plan at least one follow-up contact by telephone 2 weeks following the quit date. A second behavioural counselling session was provided on patients’ request.

Outcomes
The primary endpoint was point prevalence abstinence at 12 months’ follow-up as indicated by patients’ self-report and biochemically validated measures (nicotine, cotinine and thiocyanate levels from urine and saliva samples). Secondary outcomes included possible interaction effects of treatment condition and baseline characteristics (education, type of disease, nicotine dependence, clinic attendance) on abstinence, and change in the number of cigarettes/day.

Measures
Measurements of patients’ baseline characteristics included age, gender, level of education, marital status, type of disease (PAD, CAD), number of cigarettes smoked a day, severity of nicotine dependence, and outpatient clinic attendance (first or routine follow-up visit). We used the Fagerström Test for Nicotine Dependence (FTND), which measures smoking habits with six questions (Cronbach’s $\alpha = 0.62$). Sum scores (range 0–10) are used to categorize patients into low ($\leq 5$) and high ($\geq 6$) nicotine dependency.$^{19,20}$

Smoking status was assessed at baseline and 12 months follow-up using a 7-day point prevalence abstinence measure: ‘Have you been smoking during the past 7 days?’ (yes; yes, but only one puff; no). Patients were considered to be abstinent when they had not been smoking, not even one puff. When patients did not respond to the 12 months’ follow-up questionnaire or the reminder (both by postal mail), smoking status was obtained by telephone interview. Follow-up was blind to allocation.

With the 12-month follow-up questionnaire, patients were asked to bring along a urine sample, and/or to provide saliva when they arrived at the outpatient clinic for a routine follow-up visit. The samples were frozen at $-20^\circ C$ until analysis. Nicotine, cotinine and thiocyanate are widely used biomarkers, but misclassification (such as classifying passive smokers as daily smokers) occurs relatively frequent when applying a single biomarker for the analysis of smoking status.$^{21}$ Therefore, Sastre Torano and van Kan$^{22}$ developed an analytical method for the simultaneous determination of three biomarkers in urine and two nicotine-related biomarkers (nicotine and cotinine) in saliva (93.2% correct classification). We used their discriminant function$^{22}$ to distinguish objectively between smokers and non-smokers. Due to financial restrictions, saliva samples were only used for classification when no urine sample was available.
Sample size
The power calculation was based on data in Thorax,\textsuperscript{4} and the British Medical Journal,\textsuperscript{5} as well as data from the Cochrane Reviews\textsuperscript{7,16} and MIS studies.\textsuperscript{12–14} We hypothesized an 11\% cessation rate for NRT alone, and 21\% for the combination of NRT and C-MIS. The hypothesis was tested one-sided, i.e. we did not expect the C-MIS to increase the number of smoking patients. If $\alpha = 0.05$, 188 patients per treatment group (total of 376) are needed to obtain a statistical power of 80\%. Analyses were based on the intention-to-treat principle. Patients with incomplete follow-up were considered to be persisting smokers.

Statistics
Point prevalence of abstinence at 12 months was calculated by means of a logistic regression. Accordingly, we calculated absolute risk reductions (ARR), number needed to treat (NNT), odds ratios (OR) and associated confidence intervals (CI). Possible interaction effects of treatment condition and baseline characteristics (education, type of disease, nicotine dependence, clinic attendance and number of cigarettes/day) on abstinence were tested with logistic regression analyses. To investigate whether the C-MIS affected the number of cigarettes/day, a repeated-measures analysis of variance was used.

RESULTS

Response and sample
Of the approximately 1375 smoking atherosclerotic patients who were screened for eligibility from September 2001 until May 2004, 385 (28\%) met the inclusion criteria and agreed to participate in this study (Fig 1).

After randomization, one patient was withdrawn from the study because of cognitive problems, and eight patients died during follow-up, leaving 376 patients for analyses.

The trial ended in May 2005 when we received the 12-months’ follow-up questionnaire of the last included patient. Questionnaire response rates were 99\% (n=372/376) at baseline, and 70\% (n=263/376) at 12 months. Response rates in both treatment groups were comparable. Those who did not respond to the 12-month questionnaire were reached through telephone interview (18\%; n=68/376), or were lost to follow-up (12\%; n=45/376). No differences in baseline characteristics were found between those who did (n=331/376) and those who did not respond at 12 months (lost to follow-up, n=45/376), except for a small effect of marital status: more unmarried patients were lost to follow-up (chi-squared (1 df)=2.84, $P=0.06$). We collected 265 urine and 237 saliva samples (71 and 63\%, respectively). A total of 269 samples were used for analysis (265 urine and four saliva). Table 1 presents patients’ baseline characteristics.

Data on the number of patients who had used NRT and whether they used the patches as prescribed, have been reported elsewhere.\textsuperscript{23}
At follow-up, 21% (n=35/168) of the experimental group and 17% (n=27/163) of the control group reported abstinence. Including patients lost to follow-up as smokers, these rates are 19% (n=35/188) and 14% (n=27/188), respectively: 81% smokers in the experimental and 86% in the control group yields an ARR of 0.05 (95% CI= -0.02 to 0.12). Point prevalence abstinence rates were not significantly different between treatment groups (Table 2).

Experimental patients were more, although not significantly, likely to quit smoking than control patients (OR=1.44, 95% CI=0.83 to 2.50; one-sided Fisher’s exact test, P=0.17).

No interaction effects of treatment and patients’ characteristics on abstinence were found. The number of cigarettes/day decreased significantly at 12 months: from 21 to 15 cigarettes in the experimental group, and from 21 to 14 in the control group [F(1,194)=90.2, P<0.001]. The effect did not differ between the two treatment groups [F(1,194)=1, P=0.32].

Abstinence rates according to self-report measures

At follow-up, 21% (n=35/168) of the experimental group and 17% (n=27/163) of the control group reported abstinence. Including patients lost to follow-up as smokers, these rates are 19% (n=35/188) and 14% (n=27/188), respectively: 81% smokers in the experimental and 86% in the control group yields an ARR of 0.05 (95% CI= -0.02 to 0.12). Point prevalence abstinence rates were not significantly different between treatment groups (Table 2). Experimental patients were more, although not significantly, likely to quit smoking than control patients (OR=1.44, 95% CI=0.83 to 2.50; one-sided Fisher’s exact test, P=0.17).

No interaction effects of treatment and patients’ characteristics on abstinence were found. The number of cigarettes/day decreased significantly at 12 months: from 21 to 15 cigarettes in the experimental group, and from 21 to 14 in the control group [F(1,194)=90.2, P<0.001]. The effect did not differ between the two treatment groups [F(1,194)=1, P=0.32].

* The number of screened eligible patients (N=1375) is an extrapolation from percentages for the total cohort of patients (N=2725). Of this total cohort, 28% (119/421) of the smoking atherosclerotic patients agreed to participate. In this trial, we included 385 smoking atherosclerotic patients. Therefore, approximately 1375 patients (385/28 x 100=1375) were screened for their eligibility. CRF, case record form; NRT, nicotine replacement therapy; C-MIS, minimal intervention strategy adapted for cardiology inpatients.

Figure 1: Flow chart of participants through the trial.
Abstinence rates according to urine or saliva samples

According to biochemical markers, abstinence rates were 28% (n=38/137) in the experimental and 24% (n=32/132) in the control group (ARR=0.04, 95% CI= -0.06 to 0.14) (Table 2). Table 3 presents a comparison of self-reported and biochemical validated abstinence. The classification model for the biochemical markers matched self-reported abstinence.

### Table 1: Baseline characteristics of patients assigned to C-MIS+NRT or NRT.

<table>
<thead>
<tr>
<th></th>
<th>Experimental (n=186) (%)</th>
<th>Control (n=186) (%)</th>
<th>p*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean, sd)</td>
<td>59 (12)</td>
<td>58 (12)</td>
<td>.76#</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td>.42</td>
</tr>
<tr>
<td>Male</td>
<td>118 (63)</td>
<td>115 (62)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>68 (37)</td>
<td>71 (38)</td>
<td></td>
</tr>
<tr>
<td>Education†</td>
<td></td>
<td></td>
<td>.14</td>
</tr>
<tr>
<td>Low</td>
<td>107 (57)</td>
<td>125 (67)</td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>57 (31)</td>
<td>42 (23)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>22 (12)</td>
<td>19 (10)</td>
<td></td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td>.42</td>
</tr>
<tr>
<td>Married</td>
<td>105 (57)</td>
<td>108 (58)</td>
<td></td>
</tr>
<tr>
<td>Unmarried</td>
<td>81 (43)</td>
<td>78 (42)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td>.50</td>
</tr>
<tr>
<td>Dutch</td>
<td>167 (90)</td>
<td>166 (89)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>19 (10)</td>
<td>20 (11)</td>
<td></td>
</tr>
<tr>
<td>Type of disease</td>
<td></td>
<td></td>
<td>.38</td>
</tr>
<tr>
<td>PAD</td>
<td>115 (62)</td>
<td>111 (60)</td>
<td></td>
</tr>
<tr>
<td>CAD</td>
<td>71 (38)</td>
<td>75 (40)</td>
<td></td>
</tr>
<tr>
<td>Outpatient attendance</td>
<td></td>
<td></td>
<td>.46</td>
</tr>
<tr>
<td>Routine follow-up visit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First visit</td>
<td>131 (70)</td>
<td>128 (69)</td>
<td></td>
</tr>
<tr>
<td>Nicotine dependency</td>
<td></td>
<td></td>
<td>.30</td>
</tr>
<tr>
<td>Yes</td>
<td>79 (43)</td>
<td>73 (39)</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>107 (57)</td>
<td>113 (61)</td>
<td></td>
</tr>
<tr>
<td>No. cigarettes/day (mean, sd)</td>
<td></td>
<td></td>
<td>.53#</td>
</tr>
<tr>
<td></td>
<td>21 (10)</td>
<td>21 (10)</td>
<td></td>
</tr>
</tbody>
</table>

NRT, nicotine replacement therapy; C-MIS, minimal intervention strategy adapted for cardiology inpatients; PAD, peripheral arterial disease; CAD, coronary artery disease. * P-value was obtained from chi-squared test unless indicated otherwise; # P-value was obtained from t-test. † Low, vocational training; middle, advanced vocational training; high, high vocational or university training.

Note: n=372 instead of 376 because four patients did not respond to the baseline questionnaire.

### Abstinence rates according to urine or saliva samples

According to biochemical markers, abstinence rates were 28% (n=38/137) in the experimental and 24% (n=32/132) in the control group (ARR=0.04, 95% CI= -0.06 to 0.14) (Table 2). Table 3 presents a comparison of self-reported and biochemical validated abstinence. The classification model for the biochemical markers matched self-reported abstinence.

### Table 2: Abstinence rates at 12 months.

<table>
<thead>
<tr>
<th>Abstinence</th>
<th>Quitters (n; %)</th>
<th>ARR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experimental</td>
<td>Control</td>
</tr>
<tr>
<td>Self-report</td>
<td>35/168 (21)</td>
<td>27/163 (17)</td>
</tr>
<tr>
<td>Self-report + lost to follow-up</td>
<td>35/188 (19)</td>
<td>27/188 (14)</td>
</tr>
<tr>
<td>Urine†</td>
<td>38/137 (28)</td>
<td>32/132 (24)</td>
</tr>
</tbody>
</table>

ARR, absolute risk reduction; NNT, number needed to treat; OR, odds ratio; CI, confidence interval. * Probability associated with a one-sided Fisher’s Exact Test; † including four saliva samples because of the absence of urine.
abstinence and biochemical validated abstinence in 87% (71% + 16%) of the 269 cases. When taking self-reported smoking status as a reference, 3% (9/269) were false positive, whereas 10% (26/269) were false negative. One of these misclassified (false negative) patients reported using NRT at 12 months’ follow-up.

To gain insight into the unexpectedly high percentage of misclassified patients, we decided to analyse saliva (when available) from discrepant cases. Of the 35 misclassified patients, 25 saliva samples were available. Abstinence obtained from the saliva test corresponded in 12% of the cases (3/25) with the urine test, and with self-reported abstinence in 88% of the cases (19/25). Only three false positive cases remained. In case of reclassification of the 25 misclassified patients on the basis of their saliva, 7% (20/269) misclassifications remained (Table 3).

**DISCUSSION**

We could not prove that the nurse-led C-MIS offered to cardiovascular outpatients was effective. Based on the literature, we assumed an 11% cessation rate for NRT alone and 21% for the combination of C-MIS and NRT, but found 14 and 19% – a non-significant difference. However, according to the wide confidence intervals around the point estimates, we cannot definitely exclude that the C-MIS might have an effect.

In a previous C-MIS study, a significant difference between the C-MIS and usual care with respect to point prevalence abstinence was only found when patients lost to follow-up were excluded from the analyses (OR=1.63; 95% CI=1.13 to 2.34). A second C-MIS study

<table>
<thead>
<tr>
<th>Table 3: A comparison of self-report and biochemical validated abstinence (n=269)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Urine†</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Urine† + 25 saliva samples</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

† including 4 saliva samples.

<table>
<thead>
<tr>
<th>NNT (95% CI)</th>
<th>OR (95% CI)</th>
<th>p*</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 (8 to infinity)</td>
<td>1.30 (0.75 to 2.25)</td>
<td>0.20</td>
</tr>
<tr>
<td>20 (9 to infinity)</td>
<td>1.44 (0.83 to 2.50)</td>
<td>0.17</td>
</tr>
<tr>
<td>25 (7 to infinity)</td>
<td>1.23 (0.71 to 2.13)</td>
<td>0.30</td>
</tr>
</tbody>
</table>
reported no difference in quitting rates after 12 months among cardiac outpatients (22% in experimental versus 20% in the control group, \(P=0.95\)). Comparison with these studies is hindered by differences in applied methodologies. Both studies compared the C-MIS with a usual care group that did not include pharmaceutical support such as NRT, and no biochemical validation was used. Also, one study\(^{13}\) involved cardiac inpatients instead of outpatients, and did not randomize the complete sample.

Biochemically validated abstinence was higher than self-reported, but still did not differ between intervention groups. The discriminant function we used after urine analysis was previously shown to have high discriminating power (93.2% correct classification). However, the number of discrepancies in our study was substantial (13%). This resulted in higher estimates of overall abstinence rates. Different explanations for these observations can be given. First, although the method of determining three biomarkers in urine, or two in saliva simultaneously seems to be better than applying only one single biomarker, the model still may not be sensitive enough. Second, patients in the false-negative group (10%) (reported to be smokers, but classified as quitters) did not collect morning urine, but afternoon or evening urine instead. In this case, biomarker concentrations can be considerably lower, potentially classifying smokers as quitters. Third, patients in the false-positive group (3%) (reported to be quitters, but classified as smokers) did not report their true smoking status. It is not likely that patients who quit smoking report that they are still smoking (false-negative group). This idea is supported by a study of Caraballo et al.\(^{25}\) They found that variation in smoking patterns, including the extent of nicotine dosing, may explain most of the false negatives, whereas deception regarding smoking status may explain most of the false positives.

A few possible explanations for the lack of effectiveness of the C-MIS can be given. First, patients may not have received the treatment they were allocated to, or nurse specialists did not deliver the C-MIS correctly. However, evaluation of the C-MIS\(^{26}\) showed that patients in the experimental group were offered all required components of the C-MIS, whereas patients in the control group were not \([F(1;58)=392.27, P<0.001]\). Also, no differences in counselling quality between nurse specialists \([F(2;57)=0.07, P=0.93]\) and in C-MIS delivery over time were found \([F(4;55)=0.31, P=0.87]\), which demonstrates that the first as well as the last treated patients were exposed to the same intervention.\(^{26}\)

Second, since the C-MIS attempts to change patients’ cognition in order to achieve smoking cessation, the extent to which cognitions are indeed affected is of interest and was subject of our previous study.\(^{27}\) In that study\(^{27}\) we found no main effects of the C-MIS on the development of cognitions, but we did find that higher-educated patients, those who received the C-MIS, had higher intentions to quit smoking and higher self-efficacy levels than patients who did not receive the C-MIS. We therefore assume that the low or average level of education of our population (88%) might contribute to the ineffectiveness of C-MIS. Our assumption is supported by a study of Wray et al.\(^{28}\) who examined the extent to which education influences the decision to quit in middle-aged adults following
a myocardial infarction. They found that each additional year of educational attainment beyond high school raised the probability of quitting. Escobedo et al. postulated that higher-educated persons are more aware of the detrimental effects of smoking on health and that smoking has become less socially accepted. They also argued that in order to better reach individuals of lower socio-economic status with the quit-smoking message, it is necessary to understand why these individuals continue to smoke. Finally, our population consists of relatively old patients (mean age=58 (SD=12)). Even if research has shown that older smokers are interested in cessation, the majority of them hold the belief that quitting smoking will provide few additional health benefits.

Some limitations of our investigation need to be addressed. First, the absence of an additional control group in which only usual care was offered. Consequently we were unable to compare abstinence rates of patients who only received usual care to rates of those who additionally received NRT. Second, our biochemical validation of self-reported abstinence was limited because we did not collect urine and/or saliva from the entire study population. Unfortunately, only 72% of the patients responded to the request to provide urine and saliva at 12 months’ follow-up.

In conclusion, we set out to investigate the effectiveness of a short, relatively easy to implement intervention. With such an intervention we reach as many patients as possible in the setting of a cardiovascular outpatient department. We found this intervention to be feasible. However, it did not lead to a significant increase in cessation rates. Taking these disappointing results and other C-MIS studies into account, we conclude that C-MIS is not effective in supporting cardiovascular patients in their efforts to quit smoking.
REFERENCES


Clinical relevance of routinely measured vital signs in hospitalized patients: a systematic review

MN Storm-Versloot
L Verweij
C Lucas
J Ludikhuize
JC Goslings
DA Legemate
H Vermeulen

SUBMITTED
ABSTRACT

**Context:** Conflicting evidence exists on the effectiveness of routinely measured vital signs on the early detection of increased probability of adverse events.

**Objective:** To assess the clinical relevance of routinely measured vital signs in medically and surgically hospitalized patients through a systematic review.

**Data sources:** MEDLINE, EMBASE, CENTRAL, CINAHL, and MEDION were searched to October 2011.

**Study selection:** Prospective studies evaluating the clinical relevance of routine vital sign measurements (temperature, heart rate, blood pressure, oxygen saturation and respiratory rate) of hospitalized patients, in relation to mortality, septic or circulatory shock, ICU admission, bleeding, reoperation or infection.

**Data extraction:** Using structured forms, two reviewers independently extracted data and assessed the potential for bias. From each study, 2x2 tables were constructed to calculate Likelihood Ratios and predictive values, including 95% confidence intervals. If reported, area under the curve (AUC) data was also extracted.

**Data synthesis:** Of 13,806 citations, 14 studies were eligible, totaling 4972 participants. Studies were performed in medical (6), surgical (4), or the combined patient population (4). Only three studies were relatively free from potential bias. For temperature, the positive LR (LR+) ranged from 0 to 9.88 (median 1.78; 9 studies); heart rate 0.82 to 6.79 (median 1.51; 5 studies); blood pressure 0.72 to 4.7 (median 2.97; 4 studies); oxygen saturation 1.96 (1 study); and respiratory rate 1.27 to 1.89 (3 studies). Overall, three studies reported AUC data, ranging from 59 to 76. Two studies reported on combined vital signs, in which one study found a LR+ of 47.0, but in the other the AUC was not influenced.

**Conclusion:** Some discriminative LR+ were found, suggesting the clinical relevance of routine vital sign measurements. However, the subject is poorly studied, and many studies have methodological flaws. Further rigorous research is needed on the value of routinely measuring vital signs in order to detect adverse events early.
CONTEXT

Doctors and nurses have traditionally been taught that routine monitoring of vital signs is an important way of measuring physiological functioning and determining the probability of clinical deterioration and adverse events.\(^1\)\(^2\) Instructions to monitor vital signs are widely found in textbooks, in clinical teaching and on ward rounds during which patients’ vital signs charts on are studied and discussed. Although routine monitoring is daily practice in hospitals, its diagnostic effectiveness has been a point of debate for many years.

Many older studies conclude that measuring vital signs is useful. This is in direct contrast with more recent studies which question the relevance of routine measurements. Older studies suggest that changes in vital signs occur hours prior to adverse events and clinical deterioration.\(^3\)-\(^5\) On the other hand some newer studies have come to the conclusion that changes in vital signs either do not occur or do not occur early enough to determine the probability of adverse events in general hospital patients.\(^6\)-\(^8\)

The prevention and early detection of clinical deterioration and adverse events is currently a major topic in quality assurance programs. Worldwide several governmental institutes have developed guidelines on the identification of acutely ill medical patients which recommend the use of early warning scores or related systems, in which vital sign measurements are combined in an overall score.\(^9\),\(^10\) Implementation of these guidelines has led to a substantial increase in the measuring of vital signs. Recent literature has mainly focused on the accuracy of these early warning models which provide clinicians a tool for severity assessment.\(^11\),\(^12\) Knowledge of the positive likelihood ratio (LR+) for the different thresholds of each vital sign within these models is important in order to interpret them.

OBJECTIVE

This systematic review was initiated to identify and summarize those studies that have examined the clinical relevance of each routinely measured vital sign in detecting adverse events (mortality, septic shock, circulatory shock, admission to ICU, bleeding, reoperation and infection) in medically and surgically hospitalized patients.

METHODS

Data sources

Search strategy

Systematic and comprehensive searches were developed with a clinical librarian. The electronic databases MEDLINE, EMBASE, CENTRAL, CINAHL, and MEDION were searched up to October 2011. Search terms and strategies for each database are described in Appendix A.
In addition to these searches one author (JL) contacted experts in the field to identify relevant ongoing publications. However, we did not search abstract books of conference proceedings.

**Study selection**

We performed a three-phase selection process. The search identified 13,806 citations of potential relevance. In the first phase two reviewers (LV, MNSV) independently checked a 25% random sample of the abstracts of these references with broad inclusion criteria (Appendix B). If there was disagreement the abstracts were discussed. As agreement was more than 95%, one reviewer (MNSV) continued with the remaining abstracts.

Two reviewers (JL, HV) independently performed a second selection of the remaining 442 abstracts with narrower inclusion criteria for mortality, septic shock, circulatory shock, admission to ICU, bleeding, reoperation and infection outcomes. It also had to be possible to create 2x2 tables from data reported in the abstract. Any disagreement was resolved through discussion.

The overall selection revealed 52 potentially relevant studies for which the full text was independently read in the third phase by LV and MNSV, who created 2x2 tables of extracted data as a final inclusion criterion. In cases of disagreement, the final decision regarding inclusion was made by a third reviewer (HV).

**Criteria for inclusion**

*Types of studies:* Any prospective study which evaluated routine measurements of vital signs in an original hospitalized patient series, had been published as full text and written in English, French, German, Dutch or Spanish. If a number of studies had been published on the same series, the most complete study on outcomes was used.

*Types of participants:* Surgical and medical patients aged over 18 and admitted to general hospital wards. Patients admitted to specialized wards, such as intensive care unit, cardiac care unit, neurology and cardiology were excluded, as were patients only measuring vital signs within the first 24 hours after admission for medical treatment or surgical intervention.

*Types of interventions:* Studies measuring vital signs (temperature, heart rate, blood pressure, oxygen saturation and respiratory rate) on a routine basis. The vital signs had to be compared to a reference standard, which could be a single test or a combination of different tests or could be defined as the presence or absence of an adverse event.\(^{13}\)

*Types of outcome measures:* Adverse events of interest were mortality, septic or circulatory shock, admission to ICU, bleeding, reoperation or infection. Information on vital signs in relation to adverse events had to be reported as sensitivity (Sens), specificity (Spec), predictive values, likelihood-ratios (LR), area under the curve (AUC); alternatively the 2x2 table could be calculated from information given in the report. These clinical relevance parameters reflect the increased probability of having outcomes of interest.
Data extraction
Two reviewers (LV, MNSV) independently extracted data using a data extraction sheet. To describe the included studies in sufficient detail and to identify clinical heterogeneity, we extracted the following data: study characteristics (study design, year of publication, specialism and country of origin); patient demographics (age, gender, type of disease, reason for admission and co-morbidity); vital sign measurements (frequencies, threshold used, moment of measurement); adverse events, with their definition and prevalence (mortality, septic or circulatory shock, admission to ICU, bleeding, reoperation and infection); reference tests; and the above-mentioned types of outcome measures or raw study data to calculate clinical relevance parameters.

Quality assessment
A valid checklist for prognostic studies has been lacking until now, however six domains with potentially useful criteria are described by Hayden et al. and Minne et al.14,15 Two of these domains are relevant to our study and are congruent with four relevant quality items of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool (Appendix C).16 These four criteria comprise a defined representative patient spectrum, a defined outcome such as presence or absence of an adverse event, independent measurements, and blind outcome assessment. The methodological quality was assessed independently by two reviewers (LV, MNSV) and is summarized in Table 1.

Analysis
From each study 2x2 tables were extracted for each routinely measured vital sign and for each threshold, in order to calculate clinical relevance parameters: sensitivity (Sens), specificity (Spec), positive predictive value (PPV), negative predictive value (NPV), positive likelihood-ratio (LR+) and negative likelihood-ratio (LR-). Additionally, a 95% confidence interval (95% CI) was calculated for each. The AUC data was extracted from studies when it was reported.

RESULTS

Characteristics of included studies
Fourteen studies4,6,7,17-27 met the inclusion criteria (Figure 1). All were published between 1986 and 2010, with a total of 4972 participants. Details are summarized in Table 2.

Data synthesis
Studies were too heterogeneous to pool data. Therefore the results are presented as clinical relevance parameters according to each vital sign in relation to the outcomes of interest.
Table 1: Quality assessment

<table>
<thead>
<tr>
<th>Author</th>
<th>Spectrum</th>
<th>Outcome</th>
<th>Independent</th>
<th>Blinded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chalmers</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Hoogewerf</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Kline</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Conen</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Lighthall</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Arnell</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Gomez</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>Gonen</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Mato</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mato</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Barbier</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Payman</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Vermeulen</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Madan</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

| a | Is there a defined representative patient spectrum |
| b | Is there a defined outcome such as presence or absence of adverse event |
| c | Are reported measurements independently measured from reference test (outcome) |
| d | Are reported measurements influenced by knowledge of the outcome |

Figure 1: Flow chart

10869 excluded:
- No routine measurements
- No vital signs measured
- Missing outcomes

390 excluded:
- Missing clinical relevance values

38 excluded:
- No possibility of data-extraction N=21
- No routine measurements N=11
- Retrospective / review N=5
- Not inhospital N=1

13806 citations
2495 duplicates

11311 titles and abstracts screened for relevance

442 abstracts screened for potential to create clinical relevance values

52 full text

14 included for analysis
Clinical relevance parameters of vital signs

Temperature (Table 3)
Nine studies\textsuperscript{4,7,18,20-22,25-27} reported on temperature, for which the most common outcomes were infection or septic shock. LR+ ranged from 0 to 9.88 (median 1.78). The highest LR+ was found for reoperation outcomes, with a threshold of $\geq38.6^\circ$C. The accompanying PPV increased from 3.3% to 25%.\textsuperscript{25} The lowest LR+ was found in the primary diagnostic design study\textsuperscript{7} for infection outcomes with a threshold $\geq39^\circ$C. The accompanying PPV decreased from 6.7% to 0%. One study\textsuperscript{26} reported AUC data, which ranged from 59 to 61.

There were no distinctive differences between medical and surgical study populations.

Heart rate (Table 4)
Five studies\textsuperscript{17,20,22,25,26} reported on heart rate, for which the most common outcome was ICU admission. LR+ ranged from 0.79 to 6.79 (median 1.51). The highest LR+ was found for ICU-admission outcomes, with a threshold $\geq100$ bpm\textsuperscript{17}. The accompanying PPV increased from 6.9% to 33%. It was notable that using the same threshold, a LR+ of 0.82 was found for reoperation outcomes.\textsuperscript{25} One study\textsuperscript{26} reported AUC data, which ranged from 64 to 71, dependent on which threshold was used.

One study\textsuperscript{25} reported on a surgical population. There were no distinctive differences between medical, surgical and unknown study populations.

Blood pressure (Table 5)
Four studies\textsuperscript{6,19,20,22} reported on blood pressure, in which outcomes were mortality and ICU admission. LR+ ranged from 0.72 to 4.70 (median 2.97). The highest LR+ was found for ICU admissions, with a threshold for systolic blood pressure of $<90$ mmHg.\textsuperscript{19} The accompanying PPV increased from 10.2% to 35%. The same study\textsuperscript{19} reported AUC data, which ranged from 59 to 70. It was notable that a difference of 20 mmHg from baseline systolic blood pressure showed a LR+ of 0.72.\textsuperscript{6}

No study specifically reported on a surgical population. There were no distinctive differences between medical and combined study populations.

Peripheral oxygen saturation
One study\textsuperscript{23} reported on oxygen saturation, for which a combined outcome was measured. LR+ was 1.96 (95%CI: 1.42 to 2.69) for the threshold of $<95%$ SpO2. The accompanying PPV was 43%, but pre-test probability of outcome was unknown.

Respiratory rate (Table 6)
Three studies\textsuperscript{20,22,26} reported on respiratory rate, with different outcomes. LR+ ranged from 1.27 to 1.89. The highest LR+ was found for ICU admission, with a threshold of $>24$/min.\textsuperscript{20} The accompanying PPV increased from 40.1% to 43%. One study\textsuperscript{26} reported an AUC of 59.

No study specifically reported on a surgical population.
### Table 2: Characteristics of included studies

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Country</th>
<th>Outcome</th>
<th>Vital signs</th>
<th>Threshold vital signs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chalmers</td>
<td>2008</td>
<td>UK</td>
<td>mortality, ICU admission</td>
<td>BP</td>
<td>yes</td>
</tr>
<tr>
<td>Hoogewerf</td>
<td>2006</td>
<td>Netherlands</td>
<td>mortality, ICU admission⁵</td>
<td>T, HR, BP, RR</td>
<td>yes</td>
</tr>
<tr>
<td>Kline⁶</td>
<td>2006</td>
<td>USA</td>
<td>mortality, ICU admission circ shock⁶</td>
<td>SpO₂</td>
<td>yes</td>
</tr>
<tr>
<td>Conen⁶</td>
<td>2006</td>
<td>Switzerland</td>
<td>ICU admission, bleeding, infection⁶</td>
<td>BP</td>
<td>yes</td>
</tr>
<tr>
<td>Lighthall⁷</td>
<td>2009</td>
<td>USA</td>
<td>mortality, ICU admission</td>
<td>HR, BP, RR, SpO₂b</td>
<td>yes</td>
</tr>
<tr>
<td>Arnell⁷</td>
<td>1996</td>
<td>USA</td>
<td>ICU admission</td>
<td>HR</td>
<td>yes</td>
</tr>
<tr>
<td>Gomez⁰</td>
<td>2006</td>
<td>Argentina</td>
<td>ICU admission</td>
<td>T, HR, BP, RR</td>
<td>yes</td>
</tr>
<tr>
<td>Gonen²¹</td>
<td>2008</td>
<td>Turkey</td>
<td>septic shock</td>
<td>T</td>
<td>yes</td>
</tr>
<tr>
<td>Mato²⁶</td>
<td>2009</td>
<td>USA</td>
<td>septic shock</td>
<td>T, HR, RR²⁶</td>
<td>yes</td>
</tr>
<tr>
<td>Mato²⁷</td>
<td>2010</td>
<td>USA</td>
<td>septic shock</td>
<td>T, HR, RR²⁷</td>
<td>yes</td>
</tr>
<tr>
<td>Barbier¹⁸</td>
<td>1986</td>
<td>France</td>
<td>infection</td>
<td>T</td>
<td>yes</td>
</tr>
<tr>
<td>Payman⁴</td>
<td>1989</td>
<td>UK</td>
<td>infection</td>
<td>T</td>
<td>yes</td>
</tr>
<tr>
<td>Vermeulen⁷</td>
<td>2005</td>
<td>Netherlands</td>
<td>infection</td>
<td>T</td>
<td>yes</td>
</tr>
<tr>
<td>Madan²⁵</td>
<td>2007</td>
<td>USA</td>
<td>reoperation</td>
<td>T, HR</td>
<td>yes</td>
</tr>
</tbody>
</table>

T = temperature, HR = heart rate, BP = blood pressure, SpO₂ = peripheral oxygen saturation, RR = respiratory rate; I = intervention group, C = comparison group, SD = standard deviation; NA = not applicable; ⁵ = combined outcomes, ⁶ = combined vital signs, ⁷ = single and combined vital signs

### Table 3: Clinical relevance parameters of temperature in relation to defined outcomes

<table>
<thead>
<tr>
<th>Author</th>
<th>Specialism</th>
<th>Threshold °C</th>
<th>N</th>
<th>Outcome</th>
<th>Prev %</th>
<th>Sens % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoogewerf²²</td>
<td>Medical</td>
<td>&lt;35 or &gt;40</td>
<td>260</td>
<td>mortality, ICU⁶</td>
<td>30.8</td>
<td>14 (6 - 21)</td>
</tr>
<tr>
<td>Gomez²⁰</td>
<td>Medical</td>
<td>≥39</td>
<td>167</td>
<td>ICU</td>
<td>40.1</td>
<td>28 (18 - 39)</td>
</tr>
<tr>
<td>Gonen²¹</td>
<td>Medical</td>
<td>≥38</td>
<td>61</td>
<td>septic shock</td>
<td>1.6</td>
<td>100</td>
</tr>
<tr>
<td>Mato²⁶</td>
<td>Medical</td>
<td>&lt;36 or &gt;38</td>
<td>547</td>
<td>septic shock</td>
<td>8.4</td>
<td>NA</td>
</tr>
<tr>
<td>Mato²⁷</td>
<td>Medical</td>
<td>≥37.9</td>
<td>547</td>
<td>septic shock</td>
<td>8.4</td>
<td>NA</td>
</tr>
<tr>
<td>Barbier¹⁸</td>
<td>Surgical</td>
<td>≥38</td>
<td>100</td>
<td>infection</td>
<td>3.0</td>
<td>100</td>
</tr>
<tr>
<td>Payman⁴</td>
<td>Surgical</td>
<td>temp oral ≥37.4</td>
<td>41</td>
<td>infection</td>
<td>29.3</td>
<td>58 (30 - 86)</td>
</tr>
<tr>
<td>Payman⁴</td>
<td>Surgical</td>
<td>temp aural ≥37.8</td>
<td>41</td>
<td>infection</td>
<td>29.3</td>
<td>58 (30 - 86)</td>
</tr>
<tr>
<td>Vermeulen⁷</td>
<td>Surgical</td>
<td>≥38.5</td>
<td>284</td>
<td>infection</td>
<td>6.7</td>
<td>37 (15 - 59)</td>
</tr>
<tr>
<td>Vermeulen⁷</td>
<td>Surgical</td>
<td>≥39</td>
<td>284</td>
<td>infection</td>
<td>6.7</td>
<td>5 (0 - 15)</td>
</tr>
<tr>
<td>Vermeulen⁷</td>
<td>Surgical</td>
<td>≥38.0</td>
<td>284</td>
<td>infection</td>
<td>6.7</td>
<td>11 (0 - 24)</td>
</tr>
<tr>
<td>Vermeulen⁷</td>
<td>Surgical</td>
<td>≥38.5</td>
<td>228²⁸</td>
<td>infection</td>
<td>6.7</td>
<td>6 (3 - 9)</td>
</tr>
<tr>
<td>Vermeulen⁷</td>
<td>Surgical</td>
<td>≥39.0</td>
<td>228²⁸</td>
<td>infection</td>
<td>6.7</td>
<td>0</td>
</tr>
<tr>
<td>Madan²⁵</td>
<td>Surgical</td>
<td>≥38.6</td>
<td>245</td>
<td>reoperation</td>
<td>3.3</td>
<td>25 (0 - 55)</td>
</tr>
</tbody>
</table>

Prev = prevalence, Sens = sensitivity, Spec = specificity, PPV = positive predictive value, NPV = negative predictive value, LR⁺ = positive likelihood ratio, LR⁻ = negative likelihood ratio, AUC = area under the curve. CI = confidence interval, NA = not applicable; ²⁸ = body temperature in degrees Celsius; ²⁸ = Measurements, ²⁸ = combined outcome
Table 2: Relevance of vital signs

<table>
<thead>
<tr>
<th>Mean age</th>
<th>N</th>
<th>Male</th>
<th>Specialism</th>
<th>Reason for admission</th>
</tr>
</thead>
<tbody>
<tr>
<td>66</td>
<td>1007</td>
<td>500</td>
<td>Medical</td>
<td>community acquired pneumonia</td>
</tr>
<tr>
<td>I 67.9, C 69.2</td>
<td>260</td>
<td>180</td>
<td>Medical</td>
<td>community acquired pneumonia</td>
</tr>
<tr>
<td>53</td>
<td>200</td>
<td>86</td>
<td>Combined</td>
<td>pulmonary embolism</td>
</tr>
<tr>
<td>64</td>
<td>639</td>
<td>322</td>
<td>Combined</td>
<td>all cause</td>
</tr>
<tr>
<td>66</td>
<td>1089</td>
<td>&gt;96%</td>
<td>Combined</td>
<td>all cause</td>
</tr>
<tr>
<td>37</td>
<td>102</td>
<td>12</td>
<td>Combined</td>
<td>gallstone pancreatitis</td>
</tr>
<tr>
<td>48</td>
<td>167</td>
<td>NA</td>
<td>Medical</td>
<td>neutropenia with fever</td>
</tr>
<tr>
<td>I 47, C 46.5</td>
<td>61</td>
<td>NA</td>
<td>Medical</td>
<td>kidney stones</td>
</tr>
<tr>
<td>I 54, C 51</td>
<td>547</td>
<td>NA</td>
<td>Medical</td>
<td>hematologic malignancies</td>
</tr>
<tr>
<td>I 54, C 51</td>
<td>230</td>
<td>126</td>
<td>Medical</td>
<td>hematologic malignancies</td>
</tr>
<tr>
<td>63.5</td>
<td>100</td>
<td>35</td>
<td>Surgical</td>
<td>hip arthroplasty</td>
</tr>
<tr>
<td>56.9</td>
<td>41</td>
<td>32</td>
<td>Surgical</td>
<td>elective surgery</td>
</tr>
<tr>
<td>55.3</td>
<td>284</td>
<td>134</td>
<td>Surgical</td>
<td>all elective surgery patients</td>
</tr>
<tr>
<td>39</td>
<td>245</td>
<td>NA</td>
<td>Surgical</td>
<td>laparoscopic gastric bypass</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Spec</th>
<th>PPV</th>
<th>NPV</th>
<th>LR+</th>
<th>LR-</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>% (95% CI)</td>
<td>% (95% CI)</td>
<td>% (95% CI)</td>
<td>(95% CI)</td>
<td>(95% CI)</td>
<td>% (95% CI)</td>
</tr>
<tr>
<td>87 (82 - 92)</td>
<td>31 (16 - 47)</td>
<td>69 (63 - 75)</td>
<td>1.03 (0.53 - 2.00)</td>
<td>1.0 (0.90 - 1.11)</td>
<td>NA</td>
</tr>
<tr>
<td>77 (71 - 84)</td>
<td>33 (21 - 45)</td>
<td>73 (67 - 80)</td>
<td>1.24 (0.78 - 1.99)</td>
<td>0.93 (0.78 - 1.10)</td>
<td>NA</td>
</tr>
<tr>
<td>85 (76 - 94)</td>
<td>10 (0 - 29)</td>
<td>100</td>
<td>6.67 (3.65 - 12.18)</td>
<td>0.000</td>
<td>NA</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>61 (54 - 68)</td>
</tr>
<tr>
<td>44 (39 - 48)</td>
<td>14 (10 - 18)</td>
<td>100</td>
<td>1.78 (1.65 - 1.92)</td>
<td>0.00</td>
<td>NA</td>
</tr>
<tr>
<td>8 (3 - 14)</td>
<td>3 (0 - 7)</td>
<td>100</td>
<td>1.09 (1.03 - 1.16)</td>
<td>0.00</td>
<td>NA</td>
</tr>
<tr>
<td>72 (56 - 89)</td>
<td>47 (21 - 72)</td>
<td>81 (66 - 96)</td>
<td>2.12 (0.99 - 4.52)</td>
<td>0.58 (0.28 - 1.17)</td>
<td>NA</td>
</tr>
<tr>
<td>69 (52 - 86)</td>
<td>44 (19 - 68)</td>
<td>80 (64 - 96)</td>
<td>1.88 (0.91 - 3.87)</td>
<td>0.60 (0.30 - 1.23)</td>
<td>NA</td>
</tr>
<tr>
<td>80 (75 - 85)</td>
<td>12 (4 - 20)</td>
<td>95 (92 - 98)</td>
<td>1.81 (0.96 - 3.41)</td>
<td>0.79 (0.56 - 1.12)</td>
<td>NA</td>
</tr>
<tr>
<td>93 (90 - 96)</td>
<td>5 (0 - 15)</td>
<td>93 (90 - 96)</td>
<td>0.73 (0.10 - 5.19)</td>
<td>1.02 (0.91 - 1.14)</td>
<td>NA</td>
</tr>
<tr>
<td>98 (97 - 100)</td>
<td>0</td>
<td>93 (90 - 96)</td>
<td>0.00</td>
<td>1.02 (1.00 - 1.04)</td>
<td>NA</td>
</tr>
<tr>
<td>92 (88 - 95)</td>
<td>8 (0 - 19)</td>
<td>94 (91 - 97)</td>
<td>1.27 (0.32 - 4.99)</td>
<td>0.98 (0.83 - 1.14)</td>
<td>NA</td>
</tr>
<tr>
<td>92 (91 - 94)</td>
<td>8 (4 - 12)</td>
<td>90 (89 - 91)</td>
<td>0.81 (0.48 - 1.38)</td>
<td>1.02 (0.98 - 1.05)</td>
<td>NA</td>
</tr>
<tr>
<td>98 (98 - 99)</td>
<td>3 (0 - 7)</td>
<td>90 (89 - 91)</td>
<td>0.23 (0.03 - 1.69)</td>
<td>1.02 (1.00 - 1.03)</td>
<td>NA</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>90 (89 - 91)</td>
<td>0.00</td>
<td>1.00</td>
<td>NA</td>
</tr>
<tr>
<td>98 (96 - 100)</td>
<td>25 (0 - 55)</td>
<td>98 (96 - 100)</td>
<td>9.88 (2.05 - 41.6)</td>
<td>0.77 (0.52 - 1.15)</td>
<td>NA</td>
</tr>
</tbody>
</table>
Table 4: Clinical relevance parameters of heart rate in relation to defined outcomes

<table>
<thead>
<tr>
<th>Author</th>
<th>Specialism</th>
<th>Threshold bpm</th>
<th>N</th>
<th>Outcome</th>
<th>Prev %</th>
<th>Sens % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoogewerf22</td>
<td>Medical</td>
<td>&gt;125</td>
<td>260</td>
<td>mortality, ICU\textsuperscript{a}</td>
<td>30.8</td>
<td>26 (17 - 36)</td>
</tr>
<tr>
<td>Arnell\textsuperscript{17}</td>
<td>Combined</td>
<td>≥100</td>
<td>102</td>
<td>ICU</td>
<td>6.9</td>
<td>86 (60 - 100)</td>
</tr>
<tr>
<td>Gomez\textsuperscript{20}</td>
<td>Medical</td>
<td>&gt;120</td>
<td>167</td>
<td>ICU</td>
<td>40.1</td>
<td>27 (16 - 38)</td>
</tr>
<tr>
<td>Mato\textsuperscript{26}</td>
<td>Medical</td>
<td>&gt;90</td>
<td>547</td>
<td>septic shock</td>
<td>8.4</td>
<td>NA</td>
</tr>
<tr>
<td>Mato\textsuperscript{26}</td>
<td>Medical</td>
<td>≥99</td>
<td>547</td>
<td>septic shock</td>
<td>8.4</td>
<td>NA</td>
</tr>
<tr>
<td>Madan\textsuperscript{25}</td>
<td>Surgical</td>
<td>&gt;100</td>
<td>245</td>
<td>reoperation</td>
<td>3.3</td>
<td>13 (0 - 35)</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Combined outcomes

Prevalence, Sens = sensitivity, Spec = specificity, PPV = positive predictive value, NPV = negative predictive value, LR+ = positive likelihood ratio, LR- = negative likelihood ratio, AUC = area under the curve, CI = confidence interval, NA = not applicable; bpm = beats per minute; \textsuperscript{a} combined outcomes.

Table 5: Clinical relevance parameters of blood pressure in relation to defined outcomes

<table>
<thead>
<tr>
<th>Author</th>
<th>Specialism</th>
<th>Threshold mmHg</th>
<th>N</th>
<th>Outcome</th>
<th>Prev %</th>
<th>Sens % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chalmers\textsuperscript{19}</td>
<td>Medical</td>
<td>syst &lt;90</td>
<td>1007</td>
<td>mortality</td>
<td>9.6</td>
<td>41 (31 - 51)</td>
</tr>
<tr>
<td>Chalmers\textsuperscript{19}</td>
<td>Medical</td>
<td>diast ≤60</td>
<td>1007</td>
<td>mortality</td>
<td>9.6</td>
<td>54 (44 - 64)</td>
</tr>
<tr>
<td>Conen\textsuperscript{6}</td>
<td>Combined</td>
<td>syst &gt;10\textsuperscript{a}</td>
<td>639</td>
<td>mortality, ICU, bleeding, infection\textsuperscript{b}</td>
<td>19.1</td>
<td>41 (32 - 50)</td>
</tr>
<tr>
<td>Conen\textsuperscript{6}</td>
<td>Combined</td>
<td>syst &gt;20\textsuperscript{a}</td>
<td>639</td>
<td>mortality, ICU, bleeding, infection\textsuperscript{b}</td>
<td>19.1</td>
<td>10 (5 - 16)</td>
</tr>
<tr>
<td>Conen\textsuperscript{6}</td>
<td>Combined</td>
<td>diast &gt;10\textsuperscript{a}</td>
<td>639</td>
<td>mortality, ICU, bleeding, infection\textsuperscript{b}</td>
<td>19.1</td>
<td>24 (16 - 31)</td>
</tr>
<tr>
<td>Conen\textsuperscript{6}</td>
<td>Combined</td>
<td>diast &gt;20\textsuperscript{a}</td>
<td>639</td>
<td>mortality, ICU, bleeding, infection\textsuperscript{b}</td>
<td>19.1</td>
<td>5 (1 - 9)</td>
</tr>
<tr>
<td>Hoogewerf\textsuperscript{22}</td>
<td>Medical</td>
<td>syst &lt;90</td>
<td>260</td>
<td>mortality, ICU\textsuperscript{b}</td>
<td>30.8</td>
<td>4 (0 - 8)</td>
</tr>
<tr>
<td>Chalmers\textsuperscript{19}</td>
<td>Medical</td>
<td>syst &lt;90</td>
<td>1007</td>
<td>ICU</td>
<td>10.2</td>
<td>45 (35 - 54)</td>
</tr>
<tr>
<td>Chalmers\textsuperscript{19}</td>
<td>Medical</td>
<td>diast ≤60</td>
<td>1007</td>
<td>ICU</td>
<td>10.2</td>
<td>65 (56 - 74)</td>
</tr>
<tr>
<td>Gomez\textsuperscript{20}</td>
<td>Medical</td>
<td>syst &lt;90</td>
<td>167</td>
<td>ICU</td>
<td>40.1</td>
<td>22 (12 - 32)</td>
</tr>
</tbody>
</table>

Prevalence, Sens = sensitivity, Spec = specificity, PPV = positive predictive value, NPV = negative predictive value, LR+ = positive likelihood ratio, LR- = negative likelihood ratio, AUC = area under the curve, CI = confidence interval, NA = not applicable; mmHg = blood pressure in millimeters mercury; \textsuperscript{a} difference in mmHg from baseline blood pressure; \textsuperscript{b}combined outcomes.

Table 6: Clinical relevance parameters of respiratory rate in relation to defined outcomes

<table>
<thead>
<tr>
<th>Author</th>
<th>Specialism</th>
<th>Threshold /min</th>
<th>N</th>
<th>Outcome</th>
<th>Prev %</th>
<th>Sens % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoogewerf\textsuperscript{22}</td>
<td>Medical</td>
<td>&gt;30</td>
<td>260</td>
<td>mortality, ICU\textsuperscript{a}</td>
<td>30.8</td>
<td>39 (28 - 49)</td>
</tr>
<tr>
<td>Gomez\textsuperscript{20}</td>
<td>Medical</td>
<td>&gt;24</td>
<td>167</td>
<td>ICU</td>
<td>40.1</td>
<td>25 (15 - 36)</td>
</tr>
<tr>
<td>Mato\textsuperscript{26}</td>
<td>Medical</td>
<td>&gt;20</td>
<td>547</td>
<td>septic shock</td>
<td>8.4</td>
<td>NA</td>
</tr>
</tbody>
</table>

Prevalence, Sens = sensitivity, Spec = specificity, PPV = positive predictive value, NPV = negative predictive value, LR+ = positive likelihood ratio, LR- = negative likelihood ratio, AUC = area under the curve, CI = confidence interval, NA = not applicable; / min = respiratory rate per minute; \textsuperscript{a} combined outcomes.
### Table 4: Relevance of vital signs

<table>
<thead>
<tr>
<th>Spec % (95% CI)</th>
<th>PPV % (95% CI)</th>
<th>NPV % (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR- (95% CI)</th>
<th>AUC % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>78 (72 - 84)</td>
<td>35 (23 - 47)</td>
<td>71 (64 - 77)</td>
<td>1.21 (0.76 - 1.92)</td>
<td>0.94 (0.81 - 1.10)</td>
<td>NA</td>
</tr>
<tr>
<td>87 (81 - 94)</td>
<td>33 (12 - 55)</td>
<td>99 (97 - 100)</td>
<td>6.79 (3.69 - 12.48)</td>
<td>0.16 (0.03 - 1.01)</td>
<td>NA</td>
</tr>
<tr>
<td>81 (75 - 87)</td>
<td>35 (22 - 48)</td>
<td>74 (68 - 80)</td>
<td>1.81 (0.84 - 3.60)</td>
<td>0.79 (0.77 - 1.07)</td>
<td>NA</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>64 (57 - 71)</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>71 (63 - 78)</td>
</tr>
<tr>
<td>85 (80 - 89)</td>
<td>3 (0 - 8)</td>
<td>97 (94 - 99)</td>
<td>0.82 (0.13 - 5.28)</td>
<td>1.03 (0.79 - 1.35)</td>
<td>NA</td>
</tr>
</tbody>
</table>

### Table 5: Relevance of heart rate in relation to defined outcomes

<table>
<thead>
<tr>
<th>Spec % (95% CI)</th>
<th>PPV % (95% CI)</th>
<th>NPV % (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR- (95% CI)</th>
<th>AUC % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>90 (88 - 92)</td>
<td>31 (23 - 38)</td>
<td>93 (92 - 95)</td>
<td>4.07 (3.00 - 5.54)</td>
<td>0.66 (0.55 - 0.77)</td>
<td>70 (67 - 74)</td>
</tr>
<tr>
<td>65 (62 - 68)</td>
<td>14 (10 - 18)</td>
<td>93 (91 - 95)</td>
<td>1.52 (1.24 - 1.87)</td>
<td>0.72 (0.58 - 0.89)</td>
<td>59 (56 - 62)</td>
</tr>
<tr>
<td>59 (55 - 63)</td>
<td>18 (14 - 23)</td>
<td>81 (77 - 85)</td>
<td>0.99 (0.78 - 1.26)</td>
<td>1.01 (0.86 - 1.19)</td>
<td>NA</td>
</tr>
<tr>
<td>86 (83 - 89)</td>
<td>14 (7 - 22)</td>
<td>81 (77 - 84)</td>
<td>0.72 (0.41 - 1.28)</td>
<td>1.05 (0.98 - 1.12)</td>
<td>NA</td>
</tr>
<tr>
<td>76 (72 - 79)</td>
<td>18 (12 - 24)</td>
<td>81 (78 - 85)</td>
<td>0.97 (0.68 - 1.39)</td>
<td>1.01 (0.90 - 1.13)</td>
<td>NA</td>
</tr>
<tr>
<td>95 (93 - 97)</td>
<td>19 (5 - 33)</td>
<td>82 (78 - 97)</td>
<td>1.05 (0.44 - 2.51)</td>
<td>1.0 (0.95 - 1.04)</td>
<td>NA</td>
</tr>
<tr>
<td>98 (96 - 100)</td>
<td>43 (6 - 80)</td>
<td>70 (63 - 75)</td>
<td>1.69 (0.39 - 7.37)</td>
<td>0.98 (0.94 - 1.03)</td>
<td>NA</td>
</tr>
<tr>
<td>91 (89 - 92)</td>
<td>35 (27 - 43)</td>
<td>93 (92 - 95)</td>
<td>4.70 (3.50 - 6.30)</td>
<td>0.61 (0.51 - 0.73)</td>
<td>70 (67 - 73)</td>
</tr>
<tr>
<td>66 (63 - 69)</td>
<td>18 (14 - 22)</td>
<td>94 (93 - 96)</td>
<td>1.93 (1.62 - 2.27)</td>
<td>0.52 (0.40 - 0.69)</td>
<td>68 (65 - 72)</td>
</tr>
<tr>
<td>95 (91 - 98)</td>
<td>63 (43 - 82)</td>
<td>76 (70 - 81)</td>
<td>4.25 (1.96 - 9.25)</td>
<td>0.82 (0.72 - 0.94)</td>
<td>NA</td>
</tr>
</tbody>
</table>

### Table 6: Relevance of blood pressure in relation to defined outcomes

<table>
<thead>
<tr>
<th>Spec % (95% CI)</th>
<th>PPV % (95% CI)</th>
<th>NPV % (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR- (95% CI)</th>
<th>AUC % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>69 (63 - 76)</td>
<td>36 (26 - 46)</td>
<td>72 (65 - 79)</td>
<td>1.27 (0.89 - 1.81)</td>
<td>0.88 (0.72 - 1.08)</td>
<td>NA</td>
</tr>
<tr>
<td>87 (84 - 92)</td>
<td>43 (27 - 58)</td>
<td>75 (69 - 81)</td>
<td>1.89 (1.08 - 3.30)</td>
<td>0.86 (0.74 - 1.00)</td>
<td>NA</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>59 (53 - 65)</td>
</tr>
</tbody>
</table>
Single and combinations of deteriorated vital signs (Table 7)

One study reported on abnormal single or combinations of vital signs, with outcomes of a combination of mortality and admission to the ICU. For one abnormal sign, LR+ ranged from 4.28 to 6.91. For two abnormal signs, LR+ ranged from 18.10 to 47.00. It was notable that although the accompanying PPV was increasing, the Sens was decreasing. In other words, of those patients who had an adverse event, 28% had two abnormal signs.

Two studies reported on the relationship between multiple combinations of deteriorated temperature, heart rate, respiratory rate and septic shock. The AUC ranged from 66 to 76. There was no difference in discriminating power between two and three deteriorated vital signs.

No study specifically reported on a surgical population.

Table 7: Clinical relevance parameters of combined vital signs in relation to defined outcomes

<table>
<thead>
<tr>
<th>Author</th>
<th>Specialism</th>
<th>Threshold</th>
<th>N</th>
<th>Outcome</th>
<th>Prev %</th>
<th>Sens % (95% CI)</th>
<th>Spec % (95% CI)</th>
<th>PPV % (95% CI)</th>
<th>NPV % (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR- (95% CI)</th>
<th>AUC% (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lighthall</td>
<td>Combined</td>
<td>one abnormal sign†</td>
<td>1089</td>
<td>mortality, ICU cardiac arrest</td>
<td>7.6</td>
<td>72 (61 - 81)</td>
<td>89 (87 - 91)</td>
<td>35 (28 - 42)</td>
<td>97 (96 - 98)</td>
<td>6.61 (5.30 - 8.24)</td>
<td>0.31 (0.22 - 0.44)</td>
<td>NA</td>
</tr>
<tr>
<td>Lighthall</td>
<td>Combined</td>
<td>one abnormal sign†</td>
<td>1089</td>
<td>mortality 30 days</td>
<td>4.2</td>
<td>59 (43 - 73)</td>
<td>86 (84 - 88)</td>
<td>16 (11 - 22)</td>
<td>98 (97 - 99)</td>
<td>4.28 (3.22 - 5.70)</td>
<td>0.48 (0.34 - 0.68)</td>
<td>NA</td>
</tr>
<tr>
<td>Lighthall</td>
<td>Combined</td>
<td>one abnormal sign†</td>
<td>1089</td>
<td>ICU</td>
<td>3.5</td>
<td>89 (75 - 97)</td>
<td>87 (85 - 89)</td>
<td>20 (14 - 27)</td>
<td>99 (98 -100)</td>
<td>6.91 (5.71 - 8.37)</td>
<td>0.12 (0.05 - 0.31)</td>
<td>NA</td>
</tr>
<tr>
<td>Lighthall</td>
<td>Combined</td>
<td>≥ two abnormal signs†</td>
<td>1089</td>
<td>mortality, ICU, cardiac arrest</td>
<td>7.6</td>
<td>28 (18 - 39)</td>
<td>99 (99 - 99)</td>
<td>78 (57 - 91)</td>
<td>NA</td>
<td>47.00</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Lighthall</td>
<td>Combined</td>
<td>≥ two abnormal signs†</td>
<td>1089</td>
<td>mortality 30 days</td>
<td>4.2</td>
<td>26 (14 - 41)</td>
<td>99 (97 - 99)</td>
<td>44 (25 - 65)</td>
<td>NA</td>
<td>18.10</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Lighthall</td>
<td>Combined</td>
<td>≥ two abnormal signs†</td>
<td>1089</td>
<td>ICU</td>
<td>3.5</td>
<td>37 (22 - 54)</td>
<td>99 (98 - 99)</td>
<td>52 (32 - 71)</td>
<td>NA</td>
<td>29.80</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Prev = prevalence, Sens = sensitivity, Spec = specificity, PPV = positive predictive value, NPV = negative predictive value, LR+ = positive likelihood ratio, LR- = negative likelihood ratio, AUC = area under the curve, CI = confidence interval, NA = not applicable. temp °C = body temperature in degrees Celsius, HR bpm = heart rate in beats per minute, BP mmHg = blood pressure in millimeters mercury, SpO2 % = peripheral oxygen saturation in percentage, RR / min = respiratory rate per minute. † HR bpm <40 or >110, systolic BP mmHg <90, SpO2 % <90, RR / min <8 or >26. ° combined outcomes
DISCUSSION

Measurement of vital signs is a widely used and accepted routine in hospitalized patients. However, our review reveals that this common practice is poorly studied. Only 13 observational and one diagnostic study were identified, from which only mediocre evidence could be drawn on clinical relevance underpinning this daily practice, especially as most studies were designed for purposes other than our primary study objective and were thus not free of potential bias. In general all LR+ were low, but some interesting discriminative LR+ for single or combined vital signs were found.

This is illustrated by the study of Chalmers et al. They reported a promising discriminative LR+ of 4.07 for the single vital sign systolic blood pressure <90 mmHg. The pre-test probability of 10% rose to a post-test probability of 31%, a change of 21%. However, even when vital signs deviate from the normal value and a relative discriminative LR+ is found, post-test probability remains rather low and the additional value to the clinician is doubtful.

Mato et al. reported on combined vital signs (temperature, heart rate and respiratory rate). The AUC ranged from 66 to 76, which was slightly higher than their reports of single

<table>
<thead>
<tr>
<th>Spec % (95% CI)</th>
<th>PPV % (95% CI)</th>
<th>NPV % (95% CI)</th>
<th>LR+ (95% CI)</th>
<th>LR- (95% CI)</th>
<th>AUC% (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>89 (87 - 91)</td>
<td>35 (28 - 42)</td>
<td>97 (96 - 98)</td>
<td>6.61 (5.30 - 8.24)</td>
<td>0.31 (0.22 - 0.44)</td>
<td>NA</td>
</tr>
<tr>
<td>86 (84 - 88)</td>
<td>16 (11 - 22)</td>
<td>98 (97 - 99)</td>
<td>4.28 (3.22 - 5.70)</td>
<td>0.48 (0.34 - 0.68)</td>
<td>NA</td>
</tr>
<tr>
<td>87 (85 - 89)</td>
<td>20 (14 - 27)</td>
<td>99 (98 - 100)</td>
<td>6.91 (5.71 - 8.37)</td>
<td>0.12 (0.05 - 0.31)</td>
<td>NA</td>
</tr>
<tr>
<td>99 (99 - 99)</td>
<td>78 (57 - 91)</td>
<td>NA</td>
<td>47.00</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>99 (97 - 99)</td>
<td>44 (25 - 65)</td>
<td>NA</td>
<td>18.10</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>99 (98 - 99)</td>
<td>52 (32 - 71)</td>
<td>NA</td>
<td>29.80</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>73 (64 - 80)</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>75 (67 - 82)</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>68 (60 - 75)</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>76 (68 - 84)</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>74 (66 - 81)</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>66 (59 - 73)</td>
</tr>
<tr>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>68 (61 - 77)</td>
</tr>
</tbody>
</table>

Relevance of vital signs

15
vital signs (AUC range 59 to 71). Lighthall et al.\textsuperscript{24} showed that when two or more abnormal vital signs were present simultaneously, a remarkably high LR+ of 47 was accompanied by a PPV of 78%. However, there is a high proportion of false negative rates, e.g. of patients having an adverse event, only 28% had two or more abnormal vital signs.

As well as the poor post-test probabilities and false negative rates, it is worth discussing the thresholds used and generalizability. The clinical relevance of discriminative LRs found in some of the studies included are questionable, as thresholds of vital signs used are extreme; systolic blood pressure <90 mmHg,\textsuperscript{19,20,22,24} oxygen saturation <90%,\textsuperscript{24} and respiratory rate <8 or ≥26/min.\textsuperscript{24} Patients with these extremes generally have easily identifiable clinical signs of deterioration for doctors and nurses with trained assessment skills or clinical judgment. Furthermore, in the study of Goldhill et al.,\textsuperscript{28} which was conducted in patients seen by an intensive care outreach service, similar discriminative LR+ were found when using the same extreme thresholds of vital signs. They also reported on less extreme thresholds, showing that differences between pre-test and post-test probability vanished.

For daily practice it is important to differentiate between thresholds. In our review, only two studies provided results for different thresholds.\textsuperscript{7,26} Mato et al.\textsuperscript{26} used thresholds for heart rate >90/min or ≥99/min, with no significant differences in the AUC. Vermeulen et al.\textsuperscript{7} conducted a diagnostic study and reported on different thresholds of body temperature measurement (BTM) in relation to infection. Results show that BTM is of limited value in the early detection or exclusion of an infection, and the false negative rate was rather high.

The clinical relevance or generalizability of some studies can be questioned, since specific groups of patients (e.g. community acquired pneumonia,\textsuperscript{19,22} and neutropenia with fever\textsuperscript{20}) with high pre-test probability of mortality and ICU admission were studied. Although Chalmers et al.\textsuperscript{19} found some moderate differences from pre-test to post-test, two other studies found none.\textsuperscript{20,22} The same contradiction can be seen in excluded studies for this review: Goldhill et al.\textsuperscript{28} show that in patients seen by an intensive care outreach service, an increasing number of deviating vital signs was associated with higher hospital mortality. In contrast, Pedersen et al.\textsuperscript{29} found that early detection of hypoxemia in perioperative patients did not reduce either transfer to ICU or mortality. Thus, results are contradicting and can be due to differences in pre-test probability.

Our review demonstrates that there is still a lack of well-designed diagnostic and large observational studies specifically intended to investigate the clinical relevance of routine measurements for patients admitted to general hospital wards. Also the definitions of ‘routine measurements’ or ‘non-routine measurements’ are open for debate. This is in line with findings from other literature reviews of vital sign measurements, which conclude that there is a lack of explicit knowledge based on quantitative research.\textsuperscript{1,29,30} This suggests that much of the current practice of routinely measuring vital signs in general hospitalized patients (as well as the accuracy, frequency, and usefulness for detecting clinically relevant outcomes) is based on tradition and not yet on evidence from research.
Despite the lack of evidence, the monitoring of vital signs, or models mainly based on vital signs, currently receives a great deal of attention as part of quality and safety programs such as ‘The Survival Sepsis Campaign’ and campaigns to detect critically ill patients.\textsuperscript{31,32} Although observational studies show a relationship between outcomes and the number of patients with deviated vital signs, diagnostic studies can reveal the predictive value of vital signs.

CONCLUSION

In this review some discriminative positive likelihood ratios were found, suggesting the clinical relevance of routine vital sign measurements. However, these results must be interpreted with caution as the number of studies is limited, patient groups studied varied largely as to pre-test probability for adverse events, and almost all studies have methodological flaws. The daily routine of measuring vital signs in hospitalized patients can therefore still be questioned. We challenge researchers in the field of vital signs and early warning scores to consider designing and performing diagnostic studies. With the present-day popularity of models predicting patients’ adverse events and critical illness, it is important to develop further evidence of the contribution of each routinely measured vital sign.
REFERENCES


APPENDIX A

Search strategies

PUBMED (Medline)

((((((blood pressure[mh] OR blood pressure determination[mh] OR blood pressure[tiab]
OR pulse pressure[tiab] OR systolic pressure[tiab] OR diastolic pressure[tiab] OR
body temperature[mh] OR body temperature*[tiab] OR organ temperature*[tiab] OR
temperature[tiab] OR respiration[mh] OR respiratory rate*[tiab] OR respiratory[tiab]
OR respiratory rate[mh] OR heart rate[mh] OR heart rate*[tiab] OR pulse rate*[tiab] OR
pulse[tiab] OR cardiac rate*[tiab] OR heart frequency[tiab] OR cardiac frequency[tiab]
OR vital signs[majr] OR vital sign*[tiab] OR Oximetry[mh] OR oxygen saturation[tiab]
OR oximetr*[tiab]) AND (Hospitalization[mh] OR Hospital[mh] OR hospitalized[tiab]
OR hospitalised patient*[tiab] OR hospitalization[tiab] OR hospitalisation[tiab] OR
perioperative[tiab] OR “Perioperative Care”[Mesh]) AND (routine*[tw] OR ritual*[tw]
OR diagnostic tests, routine[mh] OR “Monitoring, Physiologic”[Mesh] OR physiologic
monitoring[tiab] OR continuous monitoring[tiab]))) NOT (Intraoperative care[mh] NOT
Postoperative care[mh])))) NOT ((adolescent[mh] OR child[mh] OR infant[mh]) NOT
(adult[mh])))

EMBASE (Ovid)

1. blood pressure/
2. exp blood pressure measurement/
3. (blood pressure or pulse pressure or systolic pressure or diastolic pressure).ti,ab.
4. exp body temperature/
5. (body temperature* or organ temperature* or temperature or respiratory rate* or
respiratory).ti,ab.
6. exp breathing/
7. breathing rate/
8. heart rate/
9. (heart rate* or pulse rate* or pulse or cardiac rate* or heart frequency or cardiac
frequency).ti,ab.
10. *vital sign/
11. vital sign*.ti,ab.
12. exp oximetry/
13. (oxygen saturation or oximetr*).ti,ab.
14. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13
15. hospitalization/
16. exp hospital/
17. (hospital* adj patient*).tw.
18. (hospitalized patient* or hospitalised patient* or hospitalized or hospitalization or hospitalisation).ti,ab.
19. hospital patient/ or aged hospital patient/
20. inpatient$.ti,ab.
21. (postoperative or perioperative).ti,ab.
22. perioperative period/
23. 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22
24. routine*.tw.
25. ritual*.tw.
26. diagnostic test/
27. monitoring/ or exp patient monitoring/
28. (physiologic* monitoring or continuous monitoring).ti,ab.
29. 24 or 25 or 26 or 27 or 28
30. peroperative care/ not postoperative care/
31. (adolescent/ or child/ or infant/) not adult/
32. 14 and 23 and 29
33. 32 not 30 not 31

CENTRAL

**ID Search**

#1 MeSH descriptor **Blood Pressure** explode all trees
#2 MeSH descriptor **Blood Pressure Determination** explode all trees
#3 (blood pressure or pulse pressure or systolic pressure or diastolic pressure): ti,ab,kw
#4 MeSH descriptor **Body Temperature** explode all trees
#5 (body temperature* or organ temperature* or temperature or respiratory rate* or respiratory): ti,ab,kw
#6 MeSH descriptor **Respiratory Rate** explode all trees
#7 MeSH descriptor **Respiration** explode all trees
#8 MeSH descriptor **Heart Rate** explode all trees
#9 (heart rate* or pulse rate* or pulse or cardiac rate* or heart frequency or cardiac frequency): ti,ab,kw
#10 MeSH descriptor **Vital Signs** explode all trees
#11 (vital sign*): ti,ab,kw
#12 MeSH descriptor **Oximetry** explode all trees
#13 (oxygen saturation or oximetr*): ti,ab,kw
#14 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13)

#15 MeSH descriptor Hospitalization explode all trees

#16 MeSH descriptor Hospitals explode all trees

#17 (hospitalized patient* or hospitalised patient* or hospitalized or hospitalization or hospitalisation): ti,ab,kw

#18 MeSH descriptor Inpatients explode all trees

#19 (inpatient or inpatients): ti,ab,kw

#20 (postoperative or perioperative): ti,ab,kw

#21 MeSH descriptor Perioperative Care explode all trees

#22 (#15 OR #16 OR #17 OR #18 OR #20 OR #21)

#23 (routine*): ti,ab,kw

#24 (ritual*): ti,ab,kw

#25 MeSH descriptor Diagnostic Tests, Routine explode all trees

#26 MeSH descriptor Monitoring, Physiologic explode all trees

#27 (physiologic* monitoring or continuous monitoring): ti,ab,kw

#28 (#23 OR #24 OR #25 OR #26 OR #27)

#29 MeSH descriptor Intraoperative Care explode all trees

#30 MeSH descriptor Postoperative Care explode all trees

#31 (#29 AND NOT #30)

#32 MeSH descriptor Adolescent explode all trees

#33 MeSH descriptor Child explode all trees

#34 MeSH descriptor Infant explode all trees

#35 MeSH descriptor Adult explode all trees

#36 (#32 OR #33 OR #34) AND NOT #35

#37 (#14 AND #22 AND #28)

#38 (#37 AND NOT #31 AND NOT #36)

CINAHL (Ebsco)

Query

S31 S30 not S28 not S29

S30 S14 and S22 and S27

S29 (MH “Adolescence+” OR MH “Child+” OR MH “Infant+”) not MH “Adult+

S28 MH “Intraoperative Care+” not MH “Postoperative Care+

S27 S23 or S24 or S25 or S26

S26 (TI physiologic* monitoring or AB physiologic* monitoring) or (TI continuous monitoring or AB continuous monitoring)

S25 MH “Monitoring, Physiologic+

S24 MH “Diagnostic Tests, Routine”
Relevance of vital signs

S23 TX routine* or TX ritual*
S22 S15 or S16 or S17 or S18 or S19 or S20 or S21
S21 TI (perioperative or postoperative) or AB (perioperative or postoperative)
S20 MH “Perioperative Care+”
S19 TI inpatient* or AB inpatient*
S18 MH “Inpatients”
S17 (TI hospitali?ation or AB hospitali?ation) or (TI hospitali?ed W1 patient* or AB hospitali?ed W1 patient*) or (TI hospitalized or AB hospitalized)
S16 (MH “Hospitals+”)
S15 (MH “Hospitalization+”)
S14 S1 or S2 or S3 or S4 or S5 or S6 or S7 or S8 or S9 or S10 or S11 or S12 or S13
S13 TI (oxygen saturation or oximetr*) or AB (oxygen saturation or oximetr*)
S12 MH “Oximetry+”
S11 TI vital sign* or AB vital sign*
S10 MM “Vital Signs+”
S9 TI (heart rate* or pulse rate* or cardiac rate* or heart frequency or cardiac frequency) or AB (heart rate* or pulse rate* or cardiac rate* or heart frequency or cardiac frequency)
S8 MH “Heart Rate”
S7 (MH “Respiration+”)
S6 MH “Respiratory Rate”
S5 TI (body temperature* or organ temperature* or temperature or respiratory rate* or respiratory) or AB (body temperature* or organ temperature* or temperature or respiratory rate* or respiratory)
S4 MH “Body Temperature+”
S3 TI (blood pressure or pulse pressure or systolic pressure or diastolic pressure) or AB (blood pressure or pulse pressure or systolic pressure or diastolic pressure)
S2 MH “Blood Pressure Determination+”
S1 MH “Blood Pressure+”
APPENDIX B

First selection

<table>
<thead>
<tr>
<th>Inclusion</th>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prospective studies</td>
<td>Admitted to specialized wards, such as ICU, Cardiac Care Unit, Emergency Department</td>
</tr>
<tr>
<td>Routine vital sign measurement: body temperature, heart rate, blood pressure, oxygen saturation, or respiratory rate</td>
<td>Primary treatment for illnesses requiring vital sign measurements</td>
</tr>
<tr>
<td>Medical and surgical inhospital patients</td>
<td>Primary Cardiology, Obstetrics, Neurology</td>
</tr>
<tr>
<td>Adults</td>
<td>Absence of an abstract in a Western European language</td>
</tr>
<tr>
<td>Complication, adverse event, unpredicted medical event as outcome</td>
<td></td>
</tr>
</tbody>
</table>

Second selection

Criteria of first selection and adverse events specified: mortality, septic shock, ICU admittance, bleeding, reoperation and infection.
Clinical relevance parameters (sensitivity, specificity, predictive values, likelihood-ratio’s, or area under the curve) in abstract or data available to create 2x2 tables

Third selection

Criteria of second selection screened in full text. Additionally studies were excluded which only described patients during the first 24 hours after a medical treatment or surgery. Articles were also excluded if no complete published reports were available or no clinical relevance parameters could be extracted.

APPENDIX C

A: The QUADAS tool

Item

1. Was the spectrum of patients representative of the patients who will receive the test in practice?
2. Were selection criteria clearly described?
3. Is the reference standard likely to correctly classify the target condition?
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?
6. Did patients receive the same reference standard regardless of the index test result?
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?
8. Was the execution of the index test described in sufficient detail to permit replication of the test?
9. Was the execution of the reference standard described in sufficient detail to permit its replication?
10. Were the index test results interpreted without knowledge of the results of the reference standard?
11. Were the reference standard results interpreted without knowledge of the results of the index test?
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?
13. Were uninterpretable/intermediate test results reported?
14. Were withdrawals from the study explained?

B: Quality assessment framework for systematic reviews of prognostic studies

1. Study participation
   Description of setting and study period
   Description of inclusion and exclusion criteria
   Description of patient mix
   Number of patients reported
   Number of patients >100
   Mortality rate reported
   Description of patient characteristics
   Study population represents source population

2. Prognostic factor measurement
   Definition of all prognostic factor(s) evaluated
   Description of type of model(s)
   Description of % of participants with complete data and handling of missing values

3. Outcome measurement
   Definition of outcome of interest

4. Analysis
   Description of all evaluation measures
   Description of model building strategy
   Description of test method
   Both aspects of discrimination and calibration evaluated
   Separate test set used for testing
   Sufficient presentation of data to assess adequacy of the analysis
   No selective reporting of results
   Comparison to reference model
Topical silver for preventing wound infection: a systematic review

MN Storm-Versloot
CG Vos
DT Ubbink
H Vermeulen

Adapted from: Cochrane Database of Systematic Reviews 2010, Issue 3
ABSTRACT

**Background:** Silver-containing treatments are popular and used in wound treatments to combat a broad spectrum of pathogens, but evidence of their effectiveness in preventing wound infection or promoting healing is lacking.

**Objectives:** To establish the effects of silver-containing wound dressings and topical agents in preventing wound infection and healing of wounds.

**Search strategy:** We searched the Cochrane Wounds Group Specialised Register (6 May 2009); The Cochrane Central Register of Controlled Trials (CENTRAL) (2009 Issue 2); Ovid MEDLINE (1950 to April Week 4 2009); Ovid EMBASE (1980 to 2009 Week 18); EBSCO CINAHL (1982 to April Week 4 2009) and Digital Dissertations (to May 2009) for relevant trials. We contacted manufacturers and distributors.

**Selection criteria:** Randomised controlled trials (RCTs) comparing silver-containing wound dressings and topical agents with silver-containing and non-silver-containing comparators on uninfected wounds.

**Data collection and analysis:** Two authors independently selected trials, assessed risk of bias, and extracted data.

**Main results:** We identified 26 RCTs (2066 patients). Heterogeneity of treatments and outcomes precluded meta-analysis. We grouped results according to wound type, and silver preparation.

**Burns:** Thirteen trials compared topical silver (in a variety of formulations - including silver sulfadiazine (SSD) cream) with non-silver dressings. One trial showed fewer infections with silver nitrate when compared with a non-silver dressing, but three trials showed significantly more infection with SSD than with the non-silver dressing.

Six trials compared SSD cream with silver-containing dressings. One showed significantly fewer infections with the silver-containing dressing (Hydron AgSD) compared with SSD, the remaining five found no evidence of a difference.

One trial compared two silver-containing dressings, and showed a significantly lower infection rate with silver-coated gauze (Acticoat®) than with silver nitrate gauze.

**Other wounds:** Six trials compared SSD/silver-containing dressings with non-silver dressings (nine dressings in total). Most comparisons (seven) found no significant differences in infection rates; one trial in a variety of wounds exhibited significantly fewer infections with SSD/hydrocolloid, but another, in acute wounds, found significantly more infections with SSD. Only one comparison showed a significant reduction in healing time associated with a silver-containing hydrofibre dressing in diabetic foot ulcers.

**Authors’ conclusions:** There is insufficient evidence to establish whether silver-containing dressings or topical agents promote wound healing or prevent wound infection; some poor quality evidence suggests SSD is harmful rather than beneficial.
BACKGROUND

Description of the condition
Wounds are a prevalent clinical problem and a burden to many patients, resulting in pain, discomfort, longer hospital stay, and considerable economic costs for the healthcare system. Wounds are either acute or chronic, and can result from venous or arterial insufficiency, diabetes, burns, trauma, chronic pressure or surgery. If wounds become contaminated with bacteria or clinically infected, wound healing is likely to be impaired. This holds true for both acute and chronic wounds. In addition, wound infection is one of the most common surgical complications, and leads to significant mortality and morbidity. The focus in wound care, therefore, is to prevent wound infection and to promote wound healing.

Prevention of wound infection has always been a challenge. It was not until the late eighteenth century that micro-organisms were recognised as the cause of infectious diseases, and the principles of asepsis and hygiene began to be more fully understood (germ theory, as developed by Pasteur during the period 1860 to 1863, and Lister’s development of antiseptic surgery). Good hygiene and use of antiseptics were initially considered effective strategies for the prevention of infection, including wound infection. Nurses developed stringent hygiene rules for dressing changes, and physicians experimented with various antiseptics. Some of these preventative actions have been investigated for their effectiveness in various types of wounds, including aseptic dressing techniques, hand-rubbing, sterile gloving, shaving, and skin disinfection.

Description of the intervention
Several antiseptic dressings or agents are available, each claiming advantages regarding wound healing or prevention of wound infection. The effectiveness of antiseptics such as povidone iodine, chlorhexidine, alcohol, and silver-based compounds against microorganisms has been studied in vitro as well as in vivo. In particular, silver-based compounds (e.g. silver sulfadiazine cream (SSD)) have been widely used on burns since the 1960s in an attempt to overcome the problem of wound infection, and increasingly, silver-containing dressings and topical applications are being used to prevent infection in non-burn wounds such as leg ulcers, diabetic foot ulcers, fingertips, and pressure ulcers. There is a growing number of silver-containing dressings and topical agents available for the treatment of skin wounds, including creams such as SSD, silver salts such as silver nitrate, alginate (e.g. Silvacerel®), foams (e.g. Avance, Contreet Ag), hydrofibres (e.g. Aquacel® Ag), hydrocolloids (e.g. SSD/hydrocolloid, Contreet Ag) and polymeric films and meshes (e.g. Argaees), including metallic, nanocrystalline (e.g. Acticoat®) or ionic silver (Aquacel® Ag).

How the intervention might work
Silver ions bind to the DNA of bacteria and bacterial spores, thus reducing their ability to replicate. Furthermore, silver is reported to be effective against all known bacteria,
fungi and some viruses. Few bacteria have been shown to develop resistance to silver (resistance is a major problem associated with use of antibiotics). Silver has also been described as effective against malodour. The various silver-containing dressings differ in the way the Ag+ ions are released. Mostly, Ag+ ions are released from the dressing through oxidation when the silver atoms come into contact with fluid. The silver can be incorporated as complex silver molecules in creams, ointments, hydrocolloids, hydrogels or foam dressings, which regulate the speed of delivery. Recent products have been produced in an attempt to ensure a more controlled and prolonged release of small (nanocrystalline) silver particles into the wound area. This nanocrystalline form releases silver ions faster than the normal silver materials, and, therefore, is claimed to have increased antimicrobial activity.

Why it is important to do this review
Silver-containing dressings have become popular despite the absence of a robust summary of the evidence for their role in preventing wound infection, and encouraging wound healing. The effect of silver-containing wound dressings and topical applications as treatments for infected wounds is the subject of a related review, which identified little evidence of effectiveness. It is timely, therefore, to conduct a systematic review of the effects of silver-containing dressings and topical agents for the prevention of wound infection and the promotion of wound healing in uninfected wounds.

OBJECTIVES
To summarise the evidence for the effects of silver-containing dressings and topical agents compared with non-silver dressings and topical agents in terms of preventing of wound infections and/ or promoting wound healing.

METHODS
Criteria for considering studies for this review
Types of studies
We considered all randomised controlled trials (RCTs), both published and unpublished, that evaluated the effects of silver-containing dressings and topical agents (used alone or in combination with other dressings/agents), in preventing infection or promoting the healing, or both, of uninfected wounds of any aetiology (cause) and in any care setting.

Types of participants
Men and women aged 18 years and over with any type of wound (not diagnosed as infected at baseline) in any care setting.
**Types of interventions**
Wound dressings and topical applications containing silver.
Eligible comparisons were:
1. Topical silver-containing agents compared with topical agents without silver;
2. Dressings containing silver compared with any dressings without silver (including dressings containing other antiseptics);
3. Comparisons between alternative topical preparations of silver (e.g. SSD cream);
4. Comparisons between alternative silver-containing dressings, including dose comparisons.

**Types of outcome measures**

*Primary outcomes*
1. Wound infection rate:\(^{41-43}\) Infection was defined as localised pain and swelling, spreading erythema (redness), appearance of a purulent exudate, odour, and the presence of a positive bacterial culture with more than \(10^5\) colony-forming units per mm\(^3\) tissue.\(^{42}\) Trial authors’ definitions of infection (e.g. critical colonisation) were also accepted.
2. Wound healing: this was measured as time to complete healing, rate of change in wound area or volume, or both, or time to skin grafting.

*Secondary outcomes*
Adverse events; rate of use of systemic antibiotics; pain; patient satisfaction; health related quality of life (HRQoL); length of hospital stay (LOS); costs.

**Search methods for identification of studies**

*Electronic searches*
The following electronic databases were searched: Cochrane Wounds Group Specialised Register (Searched 6 May 2009); The Cochrane Central Register of Controlled Trials (CENTRAL) (*The Cochrane Library* Issue 2, 2009); Ovid MEDLINE (1950 to April Week 4 2009); Ovid EMBASE (1980 to 2009 Week 18); EBSCO CINAHL (1982 to April Week 4 2009); Digital dissertations at http://www.umi.com (to October 2008).

The MEDLINE search was combined with the Cochrane Highly Sensitive Search Strategy for identifying reports of randomised controlled trials in MEDLINE (the sensitivity- and precision-maximising version (2008 revision)) Ovid format.\(^{44}\) The EMBASE and CINAHL searches were combined with the trial filters developed by the Scottish Intercollegiate Guidelines Network.\(^{45}\) No date or language restrictions were applied.

*Searching other resources*
We also contacted companies, manufacturers and distributors of silver dressings for details of unpublished and ongoing trials and scrutinised citations within all obtained trials and major review articles to identify any additional trials.
Data collection and analysis

Selection of studies
Two review authors (HV and DU) independently assessed the titles and abstracts of studies identified from the search in terms of their relevance and design. Full text versions of articles were obtained if, from the initial assessment, it was suggested they might meet the inclusion criteria. Another review author (either CV or MS) assessed those studies where there was disagreement.

Data extraction and management
Details of selected trials were extracted and summarised using a data extraction sheet. Data from trials published in duplicate were included only once. Data extraction was undertaken by one review author (CV), and checked for accuracy by a second (MS). Any discrepancy was resolved by discussion.

We extracted the following data.

- Characteristics of the trial (method of randomisation, setting, location of care, country, source of funding).
- Participants (number, type of wound(s), definition used to determine infection, wound size, duration of wound, length of follow-up, co-morbidities).
- Intervention (type of silver dressing or topical silver, dose of silver, frequency of dressing changes, co-interventions).
- Comparative intervention (type of dressing or topical application, dose of silver (where applicable), number of dressing changes, co-interventions).
- Primary outcomes: rate of wound infection; wound healing.
- Secondary outcomes: number and proportion of adverse events; rate of use of systemic antibiotics; pain; patient satisfaction; quality of life (QoL); length of hospital stay (LOS), and cost of treatment.

Assessment of risk of bias in included studies
Two review authors (CV and MS) independently assessed the risk of bias of each trial using the criteria outlined in the Cochrane Handbook for Systematic Reviews of Interventions. Any disagreement was referred to a third review author (DU) for adjudication.

The following criteria were applied: Sequence generation; Allocation concealment; Blinding of participant, care provider and outcome assessor; Incomplete outcome data; Drop-out rate (i.e. < 20%); Intention-to-treat analysis; Groups similar at baseline for the most important prognostic indicators; Sponsoring by a manufacturer who had a potential interest in the results; Co-interventions avoided or given to all groups?

We completed the risk of bias table for each eligible study and present an assessment of risk of bias using a 'risk of bias summary figure'. This display of internal validity indicates the weight readers may give the results of each study.
Data synthesis
Quantitative data were entered into RevMan 5 by one review author (CV) and were checked by a second (MS). Summary estimates of treatment effect (with 95% confidence intervals (CI)) were calculated for each outcome and every comparison. For continuous outcomes, the mean difference (MD) is presented. For dichotomous outcomes, the risk difference (RD) is presented; this is an absolute effect measure that expresses the difference between the experimental and control event rates, and allows calculation of the number needed to treat (NNT). We refrained from a sensitivity analysis because of the lack of replication of comparisons.

Subgroup analysis and investigation of heterogeneity
We conducted predefined subgroup analyses for different wound types: burns, acute (e.g. surgical), chronic (e.g. ulcers) and mixed wound types.

Where studies evaluated similar interventions in a similar population we assessed statistical heterogeneity using the Chi² test and estimated the amount of heterogeneity using I². Where pooling seemed appropriate in view of clinical and methodological similarities between studies, we planned to use a fixed-effect model where I² was below 25%. We did not intend to pool studies where inter-study heterogeneity was high (I² greater than 75%), and we intended to use a random-effects model when I² was between 25% and 75%. We constructed a funnel plot to test for publication bias.

RESULTS

Description of studies
See Characteristics of included studies (Table 1).

Results of the search
The search identified 313 titles of potential relevance. Discrepancy in judgement regarding suitability occurred in approximately 10% of all abstracts, but was resolved after adjudication by a third review author. After the first screening, 59 citations were considered potentially relevant. Full text articles were obtained and screened by two review authors independently against the inclusion criteria (Figure 1). One ongoing trial was identified, and four trials are awaiting assessment.

Trials were excluded if no infection or healing parameters were reported; or if silver-containing agents were not used in one of the treatment arms; if the trials were not RCTs; or if trials were published in abstract form only and no additional information could be retrieved from the trial authors to allow a decision regarding eligibility for inclusion to be made.
Included studies
Twenty-six trials\textsuperscript{49-74} met the inclusion criteria (Table 1). All 26 were published between 1980 and 2008. Study sizes ranged from 14 to 465 participants, and a total of 2066 participants were enrolled. The majority of trials (i.e. 21 of the 26 (81\%)) included fewer than 80 participants. Burns were the most frequently studied wound type (20 out of 26 (77\%)), and there was substantial variation between trials in the percentage of total body surface area (TBSA) and depth of burn studied (14 trials studied partial-thickness or superficial burns, six studied full-thickness burns). One trial included a range of types of wound (i.e. venous leg ulcers, partial-thickness burns and donor sites).\textsuperscript{59} The remaining trials included minor soft tissue injuries,\textsuperscript{53} open surgical or traumatic wounds,\textsuperscript{64} venous leg ulcers,\textsuperscript{62,63} and diabetic foot ulcers.\textsuperscript{62,63} Around half of the trials (14 out of 26 (54\%)) compared 1\% SSD cream with another topical agent or dressing without silver.\textsuperscript{49,50,64,55-59,62,66,69-74} Six trials (23\%) compared 1\% SSD with other silver-containing topical agents or dressings such as Acticoat\textsuperscript{®}, Aquacel\textsuperscript{®} Ag, Hydron\textsuperscript{®} AgSD, Sildimac\textsuperscript{®}, SSD-cerium nitrate, and SSD with chlorhexidine digluconate cream.\textsuperscript{51,52,54,60,67,68} One trial compared a silver-coated gauze dressing (Acticoat\textsuperscript{®}) with another topical agent or dressing without silver,\textsuperscript{61} and one trial compared a silver-coated gauze dressing (Acticoat\textsuperscript{®}) with 0.5\% silver nitrate solution.\textsuperscript{72} One trial compared an activated charcoal dressing containing silver (Actisorb Plus\textsuperscript{®}) with other topical agents.\textsuperscript{73} Two trials compared a hydrofibre dressing, containing ionic silver (Aquacel\textsuperscript{®}), with other topical agents.\textsuperscript{63,64} One trial compared a 0.5\% silver nitrate solution with two other agents.\textsuperscript{65}
While most of the trials had two treatment arms, two trials had three treatment arms, and one trial had four treatment arms. All but two trials reported infection rates, but the definitions of infection varied. Four trials (15%) defined infection as the presence of more than $10^5$ organisms per gram of tissue; 15 trials (58%) accepted positive wound swabs or clinical signs of infection as evidence of infection. Seven trials (27%) provided no definition of infection. Twenty-one trials (81%) reported healing rates predominantly in terms of days to complete healing, or time to complete re-epithelialisation. Pain was the secondary outcome measure most frequently reported. Three trials reported a sample size calculation. It was not clear whether informed consent was obtained in 11 trials, and in 13 trials the ethics review board approval was not reported.

Excluded studies
Eighteen trials did not meet the inclusion criteria. Six trials were not RCTs, five trials were only published in abstract form with no further information forthcoming from the study authors, and in four trials wounds were already infected. The three remaining trials were excluded because they did not compare dressings; no data was reported on the effect of silver; and the silver compound was not the comparator under investigation rather it was the type of bag covering the hand.

Risk of bias in included studies
A summary of the assessment of risk of bias based on the criteria outlined in Higgins is given in Figure 2 and Figure 3. In general, the overall methodological quality of the included trials was relatively poor, although a few trials were at low risk of bias.

Analysis of time to healing as a continuous variable
It is not appropriate to analyse time-to-event data - such as time to healing – using methods for continuous outcomes (e.g. using mean times-to-event) as the relevant times are only known for the subset of participants who have experienced the event (e.g. healing). The most appropriate way of summarising time-to-event data is to use methods of survival analysis and express the intervention effect as a hazard ratio. A hazard ratio is interpreted in a similar way to a risk ratio, as it describes how many times more (or less) likely a participant is to experience the event at a particular point in time if they receive the experimental rather than the control intervention. Inappropriate analysis of outcome data can introduce bias in the interpretation of the results.

Effects of interventions
Diverse interventions were evaluated in the 26 included trials, and, as a result, pooling was possible for only two trials. We have presented the results according to wound type, i.e. acute wounds (first burns and then other wounds), chronic wounds, and mixed wounds. Within each wound type we investigated the following comparisons:
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Setting</th>
<th>N (n)</th>
<th>Definition infection</th>
<th>Unit of allocation</th>
<th>wound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afilalo</td>
<td>Canada</td>
<td>ED</td>
<td>48</td>
<td>nr</td>
<td>patient</td>
<td>burns &lt; 15% TBSA</td>
</tr>
<tr>
<td>Carneiro</td>
<td>Tanzania</td>
<td>Surgical ward</td>
<td>64</td>
<td>cultures</td>
<td>patient</td>
<td>2nd degree burns &lt; 30% TBSA</td>
</tr>
<tr>
<td>Caruso</td>
<td>USA</td>
<td>Burn centre</td>
<td>84</td>
<td>nr</td>
<td>patient</td>
<td>superficial, mid-dermal, mixed partial-thickness burns 5-40% TBSA</td>
</tr>
<tr>
<td>De Gracia</td>
<td>Philippines</td>
<td>Burn service</td>
<td>60</td>
<td>cultures clinical criteria</td>
<td>patient</td>
<td>moderate and severe burns &gt; 15% TBSA</td>
</tr>
<tr>
<td>Dire</td>
<td>Texas</td>
<td>ED</td>
<td>465</td>
<td>clinical criteria</td>
<td>patient</td>
<td>minor, uncomplicated soft-tissue necessitating suturing</td>
</tr>
<tr>
<td>Fang</td>
<td>China</td>
<td>Nr</td>
<td>27 (54)</td>
<td>cultures</td>
<td>wounds</td>
<td>2nd degree burns</td>
</tr>
<tr>
<td>Gerding</td>
<td>USA</td>
<td>Burn centre</td>
<td>47 (50)</td>
<td>cultures clinical criteria</td>
<td>wounds</td>
<td>partial-thickness burns</td>
</tr>
<tr>
<td>Gerding</td>
<td>USA</td>
<td>ED: Burn centre</td>
<td>64 (56)</td>
<td>clinical criteria</td>
<td>wounds</td>
<td>partial-thickness thermal burns &lt; 10% TBSA</td>
</tr>
<tr>
<td>Hansbrough</td>
<td>USA</td>
<td>nr</td>
<td>79 (158)</td>
<td>nr</td>
<td>wounds</td>
<td>partial-thickness burns 1-25% TBSA</td>
</tr>
<tr>
<td>Homann</td>
<td>Germany</td>
<td>Burn centre</td>
<td>47</td>
<td>clinical criteria</td>
<td>wounds</td>
<td>partial-thickness burns &lt; 50% TBSA</td>
</tr>
<tr>
<td>Hutchinson</td>
<td>USA, UK, Netherlands</td>
<td>nr</td>
<td>292</td>
<td>clinical criteria</td>
<td>patient</td>
<td>partial-thickness burns partial-thickness burns donor sites venous leg ulcer</td>
</tr>
<tr>
<td>Inman</td>
<td>Canada</td>
<td>Burn unit</td>
<td>121</td>
<td>clinical criteria accompanied with cultures &gt;10^5 organisms gram/tissue</td>
<td>patient</td>
<td>full-thickness burns</td>
</tr>
<tr>
<td>Innes</td>
<td>Canada</td>
<td>admitted</td>
<td>17 (32)</td>
<td>cultures clinical criteria</td>
<td>wounds</td>
<td>Burns requiring split skin graft</td>
</tr>
<tr>
<td>Jacobs</td>
<td>Canada</td>
<td>private office</td>
<td>40</td>
<td>cultures clinical criteria</td>
<td>patient</td>
<td>diabetic patients with Wagner grade 1 or 2 ulcers</td>
</tr>
<tr>
<td>Jude</td>
<td>UK, Germany, nr Sweden, France</td>
<td>134</td>
<td>clinical criteria</td>
<td>patient</td>
<td>diabetic patients with Wagner grade 1 or 2 foot ulcers ≥ 1cm^2</td>
<td></td>
</tr>
<tr>
<td>Intervention (n) Comparison (n)</td>
<td>Outcomes</td>
<td>Follow-up</td>
<td>Drop-outs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------</td>
<td>-----------</td>
<td>-----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD (1%)/Bactigras (15) DuoDerm (15)</td>
<td>Infection rate; wound healing rate; pain; patient satisfaction</td>
<td>nr</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD (1%)/ chlorhexidine (32) Phenytoin (32)</td>
<td>Infection rate; wound healing rate; pain; length of hospital stay</td>
<td>till discharge</td>
<td>none</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD (40) Hydrofibre dressing containing ionic silver (42)</td>
<td>Infection rate; wound healing rate; adverse effects; use of systemic antibiotics; pain; costs</td>
<td>3 weeks</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD (1%) (30) SSD/ceurium-nitrate (30)</td>
<td>Infection rate; wound healing rate; adverse effects; use of systemic antibiotics; length of hospital stay</td>
<td>nr</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD (99) Bacitracin (109) Neomycin sulfate (110) Petrolatum (108)</td>
<td>Infection rate</td>
<td>nr</td>
<td>39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD (1%) (27) Hydron-SSD (27)</td>
<td>Infection rate; wound healing rate</td>
<td>nr</td>
<td>nr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD (1%) (23) Biosynthetic dressing (27) Paired Controls (7)</td>
<td>Infection rate; wound healing rate; pain; costs</td>
<td>nr</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD (1%) (26) Biosynthetic dressing (30)</td>
<td>Infection rate; wound healing rate; pain; costs</td>
<td>nr</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD (1%) (79) Collagenase ointment with polymyxin B sulfate/bacitracin (79)</td>
<td>Infection rate; wound healing rate; pain</td>
<td>nr</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD (43) Liposome hydrogel with PVP-I (43)</td>
<td>Infection rate; wound healing rate; adverse effects; pain</td>
<td>till healing</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD (1%)/ Hydrocolloid (58) Hydrocolloid (108) Non-occlusive paraffin gauze (126)</td>
<td>Infection rate</td>
<td>3 weeks</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD (1%) (54) SSD/chlorhexidine digluconate 0.2% (67)</td>
<td>Infection rate; use of systemic antibiotics; pain</td>
<td>nr</td>
<td>unclear</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nanocrystalline silver-coated (16) Hydrophilic polyurethane (16)</td>
<td>Infection rate; wound healing rate; costs</td>
<td>&gt;3 months</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSD (20) Benzoic acid-6%, salicylic acid-3%, Quercus rubra extract-3% (20)</td>
<td>Infection rate; wound healing rate</td>
<td>6 weeks</td>
<td>none</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydrofibre dressing containing ionic silver (67) Calcium alginate dressing (67)</td>
<td>Infection rate; wound healing rate; adverse effects</td>
<td>8 weeks</td>
<td>21 but included in analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Setting</td>
<td>N (n)</td>
<td>Definition Infection</td>
<td>Unit of allocation</td>
<td>Wound</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------</td>
<td>---------------</td>
<td>-------</td>
<td>----------------------</td>
<td>--------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Jurczak 2007</td>
<td>Great-Britain, Germany, France</td>
<td>nr</td>
<td>67</td>
<td>clinical criteria</td>
<td>patient</td>
<td>open surgical or traumatic wounds left to heal by secondary intent</td>
</tr>
<tr>
<td>Livingston 1990</td>
<td>USA</td>
<td>Burn unit</td>
<td>52</td>
<td>&gt;10^5 organisms gram/tissue</td>
<td>patient</td>
<td>thermal injury requiring skin grafting</td>
</tr>
<tr>
<td>Mashhood  2006</td>
<td>Pakistan</td>
<td>outpatient</td>
<td>50</td>
<td>cultures clinical criteria</td>
<td>patient</td>
<td>superficial and partial-thickness burns &lt;15% TBSA</td>
</tr>
<tr>
<td>Miller    1990</td>
<td>USA</td>
<td>Burn centre</td>
<td>59</td>
<td>&gt;10^5 organisms gram/tissue</td>
<td>wounds</td>
<td>full-thickness burns &lt; 40% TBSA</td>
</tr>
<tr>
<td>Muangman  2006</td>
<td>Thailand</td>
<td>Burn unit</td>
<td>50</td>
<td>swabs clinical criteria</td>
<td>patient</td>
<td>partial-thickness burns &lt;25% TBSA</td>
</tr>
<tr>
<td>Noordenbos 1999</td>
<td>USA</td>
<td>admitted</td>
<td>14</td>
<td>nr</td>
<td>wounds</td>
<td>partial-thickness burns 2-30% TBSA</td>
</tr>
<tr>
<td>Soroff    1994</td>
<td>USA</td>
<td></td>
<td>15</td>
<td>nr</td>
<td>wounds</td>
<td>partial-thickness burns &lt;25% TBSA</td>
</tr>
<tr>
<td>Subrahmanyan 1998</td>
<td>India</td>
<td>Inpatient</td>
<td>50</td>
<td>clinical criteria</td>
<td>patient</td>
<td>superficial thermal burns &lt; 40% TBSA</td>
</tr>
<tr>
<td>Tredget  1998</td>
<td>Canada</td>
<td>Burn unit</td>
<td>30 (60)</td>
<td>&gt;10^5 organisms gram/tissue</td>
<td>wounds</td>
<td>deep partial and full-thickness burns</td>
</tr>
<tr>
<td>Wunderlich 1991</td>
<td>Germany</td>
<td>nr</td>
<td>40</td>
<td>swabs</td>
<td>patient</td>
<td>venous leg ulcers</td>
</tr>
<tr>
<td>Wyatt     1990</td>
<td>USA</td>
<td>ED</td>
<td>50</td>
<td>clinical criteria</td>
<td>patient</td>
<td>minor 2nd degree burns</td>
</tr>
</tbody>
</table>

N = Number of patients included; n = Number of patients or number of wounds in each treatment group; ED: Emergency Department; nr: Not reported; SSD = Silver sulfadiazine cream
### Table 1: Cont.

<table>
<thead>
<tr>
<th>Study</th>
<th>Setting</th>
<th>N (n)</th>
<th>Patient (wound)</th>
<th>Definition in infection Unit of allocation wound</th>
<th>Intervention (n)</th>
<th>Comparison (n)</th>
<th>Outcomes</th>
<th>Follow-up</th>
<th>Drop-outs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jurczak 2007</td>
<td>Great-Britain, Germany, France</td>
<td>nr 67</td>
<td>clinical criteria</td>
<td>open surgical or traumatic wounds left to heal by secondary intent</td>
<td>Hydrofibre dressing containing ionic silver (35)</td>
<td>Povidone iodine dressing (32)</td>
<td>Infection rate; wound healing rate; adverse effects; pain</td>
<td>2 weeks</td>
<td>5</td>
</tr>
<tr>
<td>Livingstone 1990</td>
<td>USA Burn unit</td>
<td>52 &gt;10</td>
<td>5 organisms gram/tissue</td>
<td>patient thermal injury requiring skin grafting</td>
<td>Silver nitrate 0.5% (19)</td>
<td>Ringer’s lactate (15) Neomycin with bacitracin (18)</td>
<td>Infection rate; length of hospital stay</td>
<td>nr</td>
<td>none</td>
</tr>
<tr>
<td>Mashhood 2006</td>
<td>Pakistan outpatient</td>
<td>50 cultures</td>
<td>clinical criteria</td>
<td>patient superficial and partial-thickness burns &lt;15% TBSA</td>
<td>SSD (25)</td>
<td>Honey (25)</td>
<td>Wound healing rate; pain; costs</td>
<td>6 months</td>
<td>none</td>
</tr>
<tr>
<td>Miller 1990</td>
<td>USA Burn centre</td>
<td>59 &gt;10</td>
<td>5 organisms gram/tissue</td>
<td>wounds full-thickness burns &lt; 40% TBSA</td>
<td>SSD (51) Polyethylene glycol 400, poly-2-hydroxyethyl methacrylate, and dimethylsulfoxide: Dimac-containing SSD (51)</td>
<td>Infection rate; adverse effects</td>
<td>2 weeks</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Muangman 2006</td>
<td>Thailand Burn unit</td>
<td>50 swabs</td>
<td>clinical criteria</td>
<td>patient partial-thickness burns &lt;25% TBSA</td>
<td>SSD (25)</td>
<td>Nanocrystalline silver-coated dressing (25)</td>
<td>Infection rate; pain; length of hospital stay</td>
<td>nr</td>
<td>none</td>
</tr>
<tr>
<td>Noordenbos 1999</td>
<td>USA admitted</td>
<td>14</td>
<td>wounds partial-thickness burns 2-30% TBSA</td>
<td></td>
<td>SSD (14)</td>
<td>Biosynthetic dressing with skin substitute (14)</td>
<td>Infection rate; wound healing rate</td>
<td>12 months</td>
<td>none</td>
</tr>
<tr>
<td>Soroff 1994</td>
<td>USA</td>
<td>15</td>
<td>wounds partial-thickness burns &lt;25% TBSA</td>
<td></td>
<td>SSD (15)</td>
<td>Collagenase ointment with polymyxin B sulfate/bacitracin (15)</td>
<td>Wound healing rate</td>
<td>till healing</td>
<td>2</td>
</tr>
<tr>
<td>Subrahmanyan 1998</td>
<td>India Inpatient</td>
<td>50</td>
<td>clinical criteria</td>
<td>patient superficial thermal burns &lt; 40% TBSA</td>
<td>SSD (25)</td>
<td>Honey (25)</td>
<td>Infection rate; wound healing rate</td>
<td>&gt;3 weeks</td>
<td>none</td>
</tr>
<tr>
<td>Tredget 1998</td>
<td>Canada Burn unit</td>
<td>30 (60)</td>
<td>&gt;10 organisms gram/tissue</td>
<td>wounds deep partial and full-thickness burns</td>
<td>SSD (30)</td>
<td>Activated charcoal xerodressing silver-impregnated (SIAX) (19)</td>
<td>Infection rate; wound healing rate; pain</td>
<td>till first operative procedure</td>
<td>1</td>
</tr>
<tr>
<td>Wyatt 1990</td>
<td>USA Emergency Department</td>
<td>50</td>
<td>clinical criteria</td>
<td>patient minor 2nd degree burns</td>
<td>SSD (20)</td>
<td>Hydrocolloid (22)</td>
<td>Infection rate; wound healing rate; pain</td>
<td>till healing</td>
<td>8</td>
</tr>
</tbody>
</table>

N = Number of patients included; n = Number of patients or number of wounds in each treatment group; ED: Emergency Department; nr: Not reported; SSD = Silver sulfadiazine cream
### Methodological Quality Summary

<table>
<thead>
<tr>
<th>Study</th>
<th>Random sequence generation (selection bias)</th>
<th>Allocation concealment (selection bias)</th>
<th>Blinding (performance bias and detection bias)</th>
<th>Incomplete outcome data (attrition bias)</th>
<th>ITT analysis</th>
<th>Financial support for trial or trialists?</th>
<th>Groups similar at baseline?</th>
<th>Co-interventions avoided or similar?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alfaro 1992</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carneiro 2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canuso 2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>De Gracia 2001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dire 1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fang 1987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gerding 1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gerding 1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hansbrough 1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homann 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hutchinson 1993</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inman 1984</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imes 2001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jacobs 2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jude 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jurczak 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Livingston 1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mashhood 2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miller 1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muangman 2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noordenbos 1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soroff 1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subrahmanyam 1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tredget 1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wunderlich 1991</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyatt 1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2.** Methodological quality summary: review authors’ judgements about each methodological quality item for each included study.
1 topical silver-containing agents compared with topical agents or dressings without silver (SSD versus no silver);
2 dressings containing silver compared with any dressings without silver (silver versus no silver);
3 comparisons between alternative topical preparations of silver, e.g. SSD cream (SSD versus silver);
4 Comparisons between alternative silver-containing dressings including dose comparisons (silver versus silver).

For each outcome and comparison the results are presented below. Trial details are summarised in Table 1. We were only able to assess the possibility of publication bias for one comparison, SSD/silver versus no silver, where we performed a funnel plot for the outcome of infection rate. The funnel plot included 10 trials with 11 comparisons demonstrating symmetry, indicating no publication bias.

1. Acute wounds: burns

1.1 Topical silver-containing agents compared with topical agents without silver (SSD versus no silver)

Eleven trials compared a topical application containing silver (1% SSD) with another topical agent or dressing not containing silver. Only two trials compared similar interventions and were pooled,\(^5\)\(^5\)\(^5\) while the remainder were considered separately.

1.1.1 SSD cream compared with biosynthetic dressing (Biobrane\textregistered) (two trials)

Gerding\(^5\)\(^6\) enrolled 43 patients with 50 acute partial-thickness burns, and Gerding\(^5\)\(^6\) enrolled 52 patients with 56 acute partial-thickness thermal wounds in two trials comparing 1% SSD cream with a biosynthetic dressing.
Primary outcome: infection rate

Gerding\textsuperscript{55} defined wound infection on clinical grounds in conjunction with semi-quantitative surface swab cultures. In this trial a mixture of paired and unpaired data were presented; seven patients were used as matched controls by randomising the paired wounds to treatment with opposite modalities. Gerding\textsuperscript{56} defined wound infection on clinical grounds, but did not give a detailed description. In Gerding\textsuperscript{55} 4/23 wounds in the SSD group, and 4/27 in the biosynthetic dressing group were judged to be infected. While in Gerding,\textsuperscript{56} 2/26 wounds in the SSD group and 3/30 in the biosynthetic dressing group were infected.

Analysis 1.1: Number of patients that developed wound infection

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>SSD/Silver</th>
<th>NO Silver</th>
<th>Risk Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Events</td>
<td>Total</td>
<td>Events</td>
</tr>
<tr>
<td>1.1.1 Burns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afilalo 1992</td>
<td>1</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>Carneiro 2002</td>
<td>15</td>
<td>32</td>
<td>3</td>
</tr>
<tr>
<td>Gerding 1988</td>
<td>4</td>
<td>23</td>
<td>4</td>
</tr>
<tr>
<td>Gerding 1990</td>
<td>2</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>Hansbrough 1995</td>
<td>11</td>
<td>79</td>
<td>12</td>
</tr>
<tr>
<td>Homann 2007</td>
<td>0</td>
<td>43</td>
<td>0</td>
</tr>
<tr>
<td>Innes 2001</td>
<td>0</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>Livingston 1990</td>
<td>2</td>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>Livingston 1990</td>
<td>2</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td>Noordenbos 1999</td>
<td>6</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>Subrahmanyam 1998</td>
<td>5</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>Wyatt 1990</td>
<td>0</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>1.1.2 Acute wounds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dire 1995</td>
<td>12</td>
<td>99</td>
<td>6</td>
</tr>
<tr>
<td>Dire 1995</td>
<td>12</td>
<td>99</td>
<td>5</td>
</tr>
<tr>
<td>Dire 1995</td>
<td>12</td>
<td>99</td>
<td>19</td>
</tr>
<tr>
<td>Jurczak 2007</td>
<td>4</td>
<td>35</td>
<td>4</td>
</tr>
<tr>
<td>1.1.3 Chronic wounds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jacobs 2008</td>
<td>0</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>Jude 2007</td>
<td>11</td>
<td>67</td>
<td>8</td>
</tr>
<tr>
<td>1.1.4 Mixed wounds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hutchinson 1993</td>
<td>0</td>
<td>58</td>
<td>7</td>
</tr>
<tr>
<td>Hutchinson 1993</td>
<td>0</td>
<td>58</td>
<td>2</td>
</tr>
</tbody>
</table>

SSD: Silver sulfadiazine

Pooling these two trials ($I^2 = 0\%$) using a fixed effect model showed no statistically significant difference between groups (RD 0.00; 95% CI -0.12 to 0.12) (Analysis 1.1).
Primary outcome: wound healing rate

Both trials reported the standard error of the mean; the standard deviation (SD) was calculated for our analysis. In both trials, healing was defined as complete re-epithelialisation. Gerding\(^55\) reported the mean time to complete healing as 21.3 days (SD 11.03) in the SSD group, and 13.7 days (SD 6.75) in the biosynthetic dressing group, while Gerding\(^56\) reported the mean time to complete healing as 15.0 days (SD 6.12) in the SSD group and 10.6 days (SD 4.38) in the biosynthetic dressing group (Analysis 1.2). Both trials reported a statistically significant difference in favour of the biosynthetic dressing, however, these original trials analysed time to healing (a time-to-event outcome) as a continuous variable, which is inappropriate and potentially misleading (since it cannot take account of people who did not heal). We did not have access to the original data and therefore could not re-analyse it.

Secondary outcome: pain

Both trials measured pain on a scale from one (none) to five (severe). The pain score was statistically significantly lower in the biosynthetic dressing groups (pooled, fixed-effect, MD 1.41; 95% CI 0.99 to 1.83). Both trials reported standard error of the mean, we calculated the SD for the purposes of analysis.

Secondary outcome: costs

Gerding\(^55\) reported no statistically significant differences in the mean material costs, based on the total cost of topical cream, dressing materials and medications used in each case.
Nursing costs were $238 in the SSD group and $71 in the biosynthetic dressing group (P value < 0.001). No SDs were reported; therefore no mean difference could be calculated. Gerding\textsuperscript{56} reported that mean costs, based on hospital charges, were significantly lower in the biosynthetic dressing group (MD 70; 95% CI 15.5 to 124.5).

1.1.2 SSD cream compared with biosynthetic dressing with human fibroblast skin substitute (Transcyte on Biobrane mesh) (one trial)

Noordenbos\textsuperscript{69} enrolled 14 patients, each with two partial-thickness burns of similar size, and compared SSD cream on one burn with a biosynthetic dressing combined with human fibroblasts on the other.

\textit{Primary outcome: infection rate}

The trial report defined wound infection as cellulitis. Six of the 14 burns in the SSD group, and none in the biosynthetic dressing group developed cellulitis. The number of burns that developed cellulitis was significantly lower in the biosynthetic dressing group (RD 0.43; 95% CI 0.16 to 0.70) (Analysis 1.1). The number needed to treat (NNT) with biosynthetic dressings was two, in order to prevent one additional patient developing cellulitis.

\textit{Primary outcome: wound healing rate}

The report defined healing as 90% re-epithelialisation. The mean time to 90% healing in the SSD group was 18.14 days, compared to 11.14 days in the biosynthetic dressing group (Analysis 1.2). The mean time to healing was significantly shorter in the biosynthetic dressing group. Time to healing is a time-to-event outcome, however, the trialists did not analyse it as such and, therefore, this effect estimate may be inaccurate.

1.1.3 SSD cream with chlorhexidine-impregnated gauze (Bactigras\textregistered) compared with hydrocolloid dressing (Duoderm\textsuperscript{®} Hydroactive) (one trial)

Afilalo\textsuperscript{49} enrolled 48 patients with partial-thickness burns and compared a layer of SSD cream covered by chlorhexidine-impregnated gauze (Bactigras) with a hydrocolloid dressing (Duoderm Hydroactive).

\textit{Primary outcome: infection rate}

Wound infection was not defined in this trial, but was based on the unblinded, subjective opinion of the investigator or the plastic surgeon, and, therefore, was subject to bias. One participant out of 24 in the SSD with chlorhexidine-impregnated gauze group developed an infection, and 2/24 in the hydrocolloid dressing group. There was no statistically significant difference in the number of patients that developed a wound infection (RD: -0.04, 95% CI -0.18 to 0.09) (Analysis 1.1).

\textit{Primary outcome: wound healing rate}

The trialists defined wound healing as complete re-epithelialisation. The mean time to complete healing was 11.2 days in the SSD with chlorhexidine-impregnated gauze group, and
10.7 days in the hydrocolloid group (Analysis 1.2). There was no statistically significant difference in the mean time to complete healing. Again, this time-to-event outcome had been inappropriately analysed as a continuous variable rather than by survival analysis, and, therefore, was inaccurate.

**Secondary outcome: pain**
The pain scores at baseline and the second visit (24 hours after the initial visit) were assessed. Pain was measured on a scale from 1 to 10. There was no statistically significant difference in the groups for the median pain score at baseline or at the second visit.

**Secondary outcome: patient satisfaction**
Overall satisfaction was reported as excellent or satisfactory for all patients, and there was no statistically significant difference between the groups.

### 1.1.4 SSD cream compared with hydrocolloid dressing (DuoDerm® Hydroactive) (one trial)
Wyatt\textsuperscript{74} enrolled 50 patients with minor, second-degree burn injuries in order to compare the effects of SSD cream with hydrocolloid dressings.

**Primary outcome: infection rate**
Wound infection was defined on clinical grounds, but how exactly was unclear. None of the patients developed a wound infection (RD 0.00; 95% CI -0.09 to 0.09) (Analysis 1.1).

**Primary outcome: wound healing rate**
Healing was defined as complete healing. The mean time to complete healing was 15.59 days in the SSD group, and 10.23 days in the hydrocolloid dressing group. The mean time to complete healing was significantly shorter in the hydrocolloid group (Analysis 1.2). Again, this time-to-event outcome was inappropriately analysed as a continuous variable, and is, therefore, inaccurate.

**Secondary outcome: pain**
Pain was measured on a scale from one (no pain) to 10 (maximum pain). The mean pain score was 2.28 in the SSD group, and 1.09 in the hydrocolloid dressing group. The mean reported pain score was significantly lower in the hydrocolloid group (MD 1.19; 95% CI 0.56 to 1.82).

### 1.1.5 SSD cream compared with honey (two trials)
Mashhood\textsuperscript{66} enrolled 50 patients with superficial and partial-thickness burns. Subrahmanyam\textsuperscript{71} enrolled 50 patients with superficial thermal burns. Both compared the effects of SSD cream with pure, unprocessed, undiluted honey. Mashhood described it as ‘traditional medicine honey’ and Subrahmanyam stated only that the honey was obtained from hives.
Primary outcome: infection rate

While Mashhood defined wound infection on clinical grounds, and via swabs for bacterial density and culture, infection rate was not reported. For Subrahmanyam wound infection was defined clinically (presence of pus or slough), and by means of bacterial cultures. There was no statistically significant difference between the two groups in this trial with respect to clinical evidence of wound infection in the short term (day 7), but in the longer term (day 21), the honey group demonstrated significantly fewer infections (RD 0.20; 95% CI 0.03 to 0.37) (Analysis 1.1). The NNT with honey was five, in order to prevent one wound infection.

Primary outcome: wound healing rate

In the Mashhood trial healing was defined as 100% epithelialisation. The number of wounds completely healed was reported after two, four and six weeks’ treatment. At the two and four weeks’ treatment time-points, the honey group did significantly better. The number of wounds completely healed after two weeks was 25/25 in the SSD group and 13/25 in the honey treated group (RD -0.32; 95% CI -0.57 to -0.07) (Analysis 1.3). The number of wounds completely healed after four weeks was 15/25 in the SSD group and 25/25 in the honey treated group (RD -0.40; 95% CI -0.60 to -0.20). The NNT with honey was three, in order to promote the healing of one extra wound. All wounds were completely healed after six weeks.

In the Subrahmanyam trial healing was defined as “patients with clinical and histological evidence of epithelialisation”. The number of patients with clinical evidence of wound healing was reported on days 21 and 30, with histological evidence of wound healing reported for days 7 and 21. There was no statistically significant difference between the two groups for the clinical evidence on day 30. For the other time points, the honey group
performed significantly better than the SSD group. The number of patients with clinical evidence of wound healing on day 21 was 21/25 in the SSD group and 25/25 in the honey group (RD -0.16; 95% CI -0.31 to -0.01) (Analysis 1.3). The NNT with honey was six, in order to promote the healing of one extra wound.

**Secondary outcome: pain**
Mashhood reported pain on the basis of the number of participants who were free of pain after one, two, three and four weeks of treatment. While there was no statistically significant difference between the two groups at the start and end of the trial (i.e. weeks 1 and 4), there was a statistically significant difference between groups in the middle (i.e. weeks 2 and 3), with more patients free of pain in the honey group (RD -0.36; 95% CI -0.61 to -0.11). We calculated the Mann-Whitney U test: $z = -2.823$, $P$ value = 0.005.

**Secondary outcome: costs**
Mashhood reported the cost of dressing material for one percent of body surface area burnt. The cost of dressing material for each percent of body surface area burnt was PKR 0.10/2 g for SSD, and PKR 0.75/5 ml for honey. No SDs were reported, so no mean difference could be calculated.

### 1.1.6 SSD cream compared with liposome hydrogel containing polyvinyl-pyrrolidone iodine (PVP-I) (one trial)
Homann\textsuperscript{58} enrolled 47 patients with 94 partial-thickness burns (degree IIa).

**Primary outcome: infection rate**
Wound infection was defined using clinical criteria such as inflammation. When wound infection was suspected, wound swabs were taken for microbiological investigation. None of the patients developed a wound infection (RD 0.00; 95% CI -0.04 to 0.04) (Analysis 1.1).

**Primary outcome: wound healing rate**
Healing was defined as 95% to 100% re-epithelialisation. There was no statistically significant difference in the mean time to complete healing (11.3 days for the SSD group, 9.9 days for the liposome hydrogel containing polyvinyl-pyrrolidone iodine (PVPI) group) (Analysis 1.2). Again, this time-to-event outcome was inappropriately analysed as a continuous variable rather than by means of survival analysis.

**Secondary outcome: adverse events**
There was no statistically significant difference between the groups with respect to wound necrosis and wound itching (RD 0.02; 95% CI -0.05 to 0.10).

**Secondary outcome: pain**
Pain was measured, but the method the trialists used was not reported. There was no statistically significant difference in the number of patients reporting wound pain (RD -0.02; 95% CI -0.16 to 0.12).
1.1.7 SSD cream compared with collagenase ointment applied with polymyxin B sulfate/bacitrin (Santyl®) (two trials)
Soroff\(^7\) enrolled 15 patients with 30 partial-thickness burns. Hansbrough\(^5\) enrolled 79 patients with 158 partial-thickness burns.

**Primary outcome: infection rate**
Soroff did not report infection rate. Hansbrough did not define wound infection, but the number of patients with cellulitis was reported. There was no statistically significant difference in the number of patients who developed cellulitis between the groups (11/79 in the SSD group; 12/79 in the collagenase ointment applied with polymyxin B sulfate/bacitrin (Santyl®) group), (RD: -0.01; 95% CI: -0.12 to 0.10) (Analysis 1.1).

**Primary outcome: wound healing rate**
Soroff defined healing as complete re-epithelialisation and time to a clean wound bed (determined by the disappearance of injured dermis), while Hansbrough defined healing as complete re-epithelialisation and time to a clean wound bed (determined by the absence of retained dermis). In both trials, healing was significantly better in the Santyl® group. In Soroff the median time to complete epithelialisation was 15 days in the SSD group and 10 days in the Santyl® group (P value: 0.00007). In the Hansbrough trial, the mean time to epithelial closure was 22.1 days in the SSD group, and 19.0 days in the Santyl® group (no SD was reported) (P value: < 0.001). Again, this time-to-event outcome was inappropriately analysed as a continuous variable.

**Secondary outcome: pain and adverse events**
Hansbrough reported pain as an adverse event and described it as burning or stinging. The number of patients reporting pain was significantly lower in the SSD group (RD: -0.19; 95% CI: -0.31 to -0.07). The NNT with SSD was five, in order to prevent one patient from experiencing pain. Soroff reported three patients who described a burning sensation at the wound site in the Santyl® group.

1.1.8 SSD cream/chlorhexidine (Silverex) compared with diphenylhydantoin (Phenytoin) (one trial)
Carneiro\(^5\) enrolled 64 patients with second degree burns.

**Primary outcome: infection rate**
Bacterial cultures were obtained on days 5 and 10. Negative cultures were defined as the absence of pathogens. The number of positive bacterial cultures on both days was significantly lower in the diphenylhydantoin group. At day 10 15/32 cultures were positive in the SSD/chlorhexidine group compared with 3/32 in the diphenylhydantoin group (RD: 0.38; 95% CI: 0.17 to 0.58) (Analysis 1.1). The NNT with diphenylhydantoin was three, in order to prevent one additional positive culture.
Primary outcome: wound healing rate
Wound healing was defined as complete healing. There was no statistically significant difference between the groups in the rate of complete healing; 24/32 wounds in the SSD/chlorhexidine group were completely healed, and 29/32 in the diphenylhydantoin group (RD -0.16; 95% CI -0.34 to 0.02) (Analysis 1.3).

Secondary outcome: pain
Pain was measured in categories: moderate to severe pain or discomfort; mild; or no pain or discomfort. Statistically significantly more patients reported moderate to severe pain or discomfort in the SSD/chlorhexidine group (17/32), than in the diphenylhydantoin group (7/32) (RD 0.31; 95% CI 0.09 to 0.54).

Secondary outcome: length of hospital stay
The mean length of hospital stay was 16.3 days in the SSD/chlorhexidine group and 14.2 days in the diphenylhydantoin group (not statistically significant). No SDs were reported; therefore no mean difference could be calculated.

Summary for burns: SSD versus no silver
Eleven trials compared SSD with a range of non-silver comparators in participants with superficial or partial-thickness burns. Only four of the eleven trials reported adequate sequence generation, and only two described allocation concealment, therefore these trials were generally of at least moderate, (or unknown), risk of bias and the findings should be interpreted with this in mind.

Infection rate was reported in nine trials. Six trials found no statistically significant differences, and three trials found a statistically significant increase in infection with SSD compared with the non-silver comparators.

Time to complete healing was reported in eight trials, though in each trial this had been inappropriately analysed as a continuous variable ("mean time") rather than as a time-to-event outcome. Six trials showed a statistically significant difference in favour of non-silver dressings, and two trials showed no differences. However, these data would be inaccurate if not all the participants were followed to complete healing.

The proportions of wounds healed and unhealed at specific time points were reported in three trials. Two trials showed a statistically significant difference in favour of non-silver dressings, and one trial showed no difference.

Pain was reported in eight trials. While one trial showed a statistically significant difference in favour of SSD, five trials showed a statistically significant difference is favour of non-silver dressings, and two trials showed no difference.
1.2 Dressings containing silver compared with any dressings without silver (silver versus no silver)

1.2.1 Nanocrystalline silver coated dressing (Acticoat®) compared with hydrophilic polyurethane dressing (Allevyn®) (one trial)

Innes\(^{61}\) enrolled 17 patients, with 18 paired adjacent burn sites, who required a split-thickness skin graft.

*Primary outcome: infection rate*

Wound infection was defined clinically by criteria such as erythema, induration, purulent discharge, and malodour. Every third day, swabs were taken and were rated as 1 (light growth), 2 (medium growth), or 3 (heavy bacterial growth). There was no statistically significant difference in the number of patients who developed an infection, or in the number of positive cultures at any time point. None of the patients developed a wound infection (RD 0.00; 95% CI -0.11 to 0.11) (Analysis 1.1).

*Primary outcome: wound healing rate*

Healing was defined as 90% or more re-epithelialisation. Healing was significantly faster in the hydrophilic polyurethane dressing group (Allevyn®) (14.5 days for the nanocrystalline silver coated dressing (Acticoat®) group, and 9.1 days for the Allevyn® group) (Analysis 1.2). Again, this time-to-event outcome was inappropriately analysed as a continuous variable. The number of wounds healed by day of discharge showed a statistical significance in favour of Allevyn® (RD -0.69; 95% CI -0.92 to -0.45) (Analysis 1.3). The NNT with Allevyn® was six, in order to promote one additional wound to heal.

*Secondary outcome: cost*

The mean cost per cm\(^2\) was USD 0.088 in the Acticoat® group and USD 0.059 in the Allevyn® group. No SDs were reported, so no mean difference could be calculated.

1.2.2 Silver nitrate (0.5%) compared with Ringer’s lactate (one trial)

Livingston\(^{65}\) enrolled 52 patients with burns who required skin grafting. The trial had three treatment groups; silver nitrate (0.5%) (19 participants), Ringer’s lactate (15 participants), and neomycin with bacitracin (18 participants).

*Primary outcome: infection rate*

Wound infection was defined as present when there were more than \(10^5\) organisms per gram of tissue. The silver nitrate group showed significantly fewer infections (2/19 infections in the silver nitrate group; 8/15 in the Ringer’s lactate group) (RD -0.43; 95% CI -0.72 to -0.14) (Analysis 1.1). The NNT with silver nitrate was two, in order to prevent one wound infection. Mean time to development of wound infection was significantly shorter in the Ringer’s lactate group (13.7 days in the silver nitrate group, versus 5.5 days in the Ringer’s lactate group). Again, the outcome was inappropriately analysed as a continuous variable.
Secondary outcome: length of hospital stay
Length of hospital stay was only reported for subgroups, and was reported as being significantly shorter for patients in the silver nitrate group with wounds covering 20% to 40% TBSA.

1.2.3 Silver nitrate (0.5%) compared with neomycin with bacitracin (one trial)
In the same trial, the comparison arm of silver nitrate (0.5%) (19 participants) was compared with the neomycin with bacitracin arm (18 participants).

Primary outcome: infection rate
Wound infection was defined as present when there were more than $10^5$ organisms per gram of tissue. There was no statistically significant difference in the number of patients who developed an infection (2/19 in the silver nitrate group and 6/18 in the neomycin with bacitracin group) (RD -0.23; 95% CI -0.49 to 0.03) (Analysis 1.1). Mean time to development of wound infection was significantly shorter in the neomycin with bacitracin group (13.7 days in the silver nitrate group versus 5.5 days in the neomycin with bacitracin group). Again, this time-to-event outcome was inappropriately analysed as a continuous variable.

Secondary outcome: length of hospital stay
Length of hospital stay was only reported for subgroups, and there were no statistically significant differences between them.

Summary for burns: silver versus no silver
Both trials investigated burns requiring skin grafting. Only one of the trials reported adequate sequence generation, and neither trial reported adequate allocation concealment.

Infection rate was reported in both trials with a total of three dressing comparisons. Two comparisons showed no differences, and one comparison showed a statistically significant difference in favour of silver nitrate.

Time to complete healing was reported in one trial, which showed a statistically significant difference in favour of non-silver dressings however it had been wrongly analysed as a continuous variable (with mean healing time calculated) whereas time to healing is a time-to-event outcome which should be subject to analysis by survival methods.

The number of wounds healed was reported in one trial, which showed a statistically significant difference in favour of non-silver dressings. An overview of the number of patients who developed a wound infection for all trials comparing SSD/silver versus no silver is given in Analysis 1.1. A funnel plot revealed no evidence of publication bias for wound infection.
1.3 Comparisons between alternative topical preparations of silver, e.g. SSD cream (SSD versus silver)

1.3.1 SSD cream compared with nanocrystalline silver-coated dressing (Acticoat®) (one trial)
Muangman\(^6\) enrolled 50 patients, with partial-thickness burns.

*Primary outcome: infection rate*
Wound infection was defined as the presence of erythema, induration, purulent discharge and malodour. There was no statistically significant difference in the number of patients who developed an infection (4/25 in the SSD group; 3/25 in the nanocrystalline silver-coated dressing (Acticoat®) group) (RD 0.04; 95% CI -0.15 to 0.23).

*Secondary outcome: pain*
Pain was measured on a visual analogue pain scale from 1 (no pain) to 10 (extreme pain). Background pain, between dressings, was significantly lower in the Acticoat® group (5 in the SSD group, 4 in the Acticoat® group) (MD 1.00; 95% CI 0.64 to 1.36).

*Secondary outcome: length of hospital stay*
The mean length of hospital stay was 21 days in both groups (MD 0.00; 95% CI -6.43 to 6.43).

1.3.2 SSD cream compared with hydrofibre dressing containing ionic silver (Aquacel® Ag) (one trial)
Caruso\(^5\) enrolled 82 patients, with superficial, mid-dermal or mixed partial-thickness burns.

*Primary outcome: infection rate*
Wound infection was not defined. There was no statistically significant difference in the number of patients who developed an infection (6/40 in the SSD group; 8/42 in the hydrofibre dressing containing ionic silver (Aquacel® Ag) group) (RD -0.04; 95% CI -0.20 to 0.12).

*Primary outcome: wound healing rate*
Healing was defined as either 100% re-epithelialisation, including open areas; less than 1 cm fully re-epithelialized area; or re-epithelialisation less than 100% but to the extent that surgical interventions were not required. There were no differences in healing within 21 days (24/40 in the SSD group; 31/42 in the Aquacel® Ag group) (RD -0.14; 95% CI -0.34 to 0.06). For the time to complete re-epithelialisation only median values were given: 17 days in the SSD group and 16 days in the Aquacel® Ag group (P value 0.517). No MD could be calculated. The time to complete re-epithelialisation was analyzed using life table methods. Kaplan Meier survival curves for each treatment group were plotted.
Secondary outcome: adverse events
Adverse events were defined as any untoward medical occurrence that was new or worsened during the trial. There were no statistically significant differences between SSD and Aquacel® Ag for adverse events (RD -0.03; 95% CI -0.24 to 0.19).

Secondary outcome: use of systemic antibiotics
There was no statistically significant difference between groups in the number of patients that used antibiotics (RD -0.04; 95% CI -0.20 to 0.12).

Secondary outcome: pain
Pain was measured on a visual analogue scale from 1 (no pain) to 10 (extreme pain). The mean pain score per week was 4.77 in the SSD group and 3.63 in the Aquacel® Ag group (P value 0.003). No SDs were reported, so no mean difference could be calculated. Pain was also measured on an observational scale. Patients were able to grade the extent to which the dressings reduced pain from “extremely well” to “not very well at all”. Patients reported statistically significantly less pain associated with the Aquacel® Ag dressing (P value 0.002).

Secondary outcome: costs
Different components of costs were measured and combined later to be able to calculate cost effectiveness. For most components no SDs were reported, so no mean difference could be calculated. All costs were expressed as US dollars. There was no statistically significant difference in the mean total costs of clinical care ($1181 for the SSD group and $1040 for the Aquacel® Ag group) (MD $141; 95% CI -$216 to $498). The average cost effectiveness, calculated from the total cost of clinical care, divided by the proportion of patients with full epithelialisation, was $1968 (95% CI $1483 to $2690) in the SSD group and $1409 (95% CI $1050 to $1858) in the Aquacel® Ag group.

1.3.3 SSD cream compared with synthetic dressing containing silver (Hydron-AgSD)
(one trial)
Fang\textsuperscript{54} enrolled 27 patients with 54 second degree burns, with areas of similar size and injury matched.

Primary outcome: infection rate
Wound infection was determined by taking swabs for bacterial colonisation and reporting on the number of positive cultures. The time-point(s) at which the swabs were taken was not reported. The number of positive culture swabs was significantly higher in the SSD group (46/98 swabs in the SSD group; 32/98 in the synthetic dressing containing silver (Hydron-AgSD) group) (RD 0.14; 95% CI 0.01 to 0.28). The NNT with Hydron-AgSD was seven, in order to prevent one positive culture.
Primary outcome: wound healing rate
No definition of healing was reported. Fang stated that wounds healed equally in both groups, no data were reported to support this statement.

1.3.4 SSD cream (Flamazine®) compared with 1% SSD plus 0.2% chlorhexidine digluconate cream (Silvazine®) (one trial)
Inman enrolled 121 patients with fresh, full-thickness burns.

Primary outcome: infection rate
Wound infection was defined by clinical criteria such as softening of eschar, erythema, or colour change accompanied with a quantitative culture with $10^5$ or more organisms per gram of burn tissue. There was no statistically significant difference between the groups in the number of patients that developed an infection (12/67 in the SSD group; 10/54 in the SSD with chlorhexidine digluconate cream group) (RD -0.01; 95% CI -0.14 to 0.13).

Secondary outcome: use of systemic antibiotics
There was no statistically significant difference between groups in the use of antibiotics during the in-hospital period (RD 0.10; 95% CI -0.03 to 0.24).

Secondary outcome: pain
Pain was not defined. There was no statistically significant difference between groups in the number of patients who experienced extreme pain at the time when cream was being applied (RD -0.02; 95% CI -0.07 to 0.03).

1.3.5 SSD cream compared with SSD cream containing cerium nitrate (SSD-CN) (one trial)
De Gracia enrolled 60 patients with moderate and severe burns.

Primary outcome: infection rate
In the De Gracia trial, wound sepsis was defined as wound deterioration with severe inflammation. Wound biopsies were taken and bacterial growth on culture media was reported. De Gracia found no statistically significant difference between the groups for any infection outcome. The number of patients developing sepsis after ten days was 3/30 in the SSD group and 0/30 in the SSD-cerium nitrate (SSD-CN) group (RD 0.10; 95% CI -0.02 to 0.22).

Primary outcome: wound healing rate
The De Gracia trial defined healing as complete re-epithelialisation, or wounds being ready for skin grafting. Re-epithelialisation was categorised into four groups: ‘quick’ (0 to 14 days), ‘moderate’ (15 to 21 days), ‘slow’ (22 to 35 days), and ‘very slow’ (more than 35 days). We calculated the Chi$^2$ statistic as 5.233, and the P value as 0.155. There were no
statistically significant differences between the groups. The mean number of days until complete re-epithelialisation was significantly shorter in the SSD-CN nitrate group (25.1 days in the SSD group; 17.2 days in the SSD-CN group). The mean time to readiness to accept a skin graft was significantly shorter in the SSD-CN group (24.6 days in the SSD group (17 participants); 13.6 days in the SSD-CN group (nine participants). Once again, these were time-to-event outcomes that had been inappropriately analysed as continuous data.

Secondary outcome: adverse events
In the De Gracia trial skin rashes were observed in both groups, but did not differ significantly between the groups. A subjective stinging effect was significantly higher in the SSD-CN group (RD -0.37; 95% CI -0.58 to -0.15). The NNT with SSD was three, in order to prevent one participant experiencing a stinging effect.

Secondary outcome: use of systemic antibiotics
There was no statistically significant difference between groups in the number of patients who received oral antibiotics for at least seven days (RD -0.03; 95% CI -0.20 to 0.13).

Secondary outcome: length of hospital stay
There was no statistically significant difference between groups in the mean length of hospital stay (MD 7.4; 95% CI -1.69 to 16.49).

1.3.6 SSD cream compared with Dimac containing SSD (Sildimac®) (one trial)
Miller enrolled 59 patients with two separate, comparable, sustained full-thickness burns.

Primary outcome: infection rate
Wound infection was defined as present when there were more than $10^5$ organisms per gram of tissue. Wound biopsies were obtained before treatment, and every seven days thereafter until the last day of treatment. Positive cultures were defined as any growth of any organism. Wound infection was based on clinical judgement. There was no statistically significant difference between the groups in the number of patients who developed an infection at any time point. Clinical wound infection occurred in 2/51 patients in the SSD group and 1/51 patients in the Dimac SSD group (RD 0.02; 95% CI -0.05 to 0.09).

Secondary outcome: adverse events
There was no statistically significant difference between groups in the number of patients reporting local adverse effects (such as burning and stinging) (RD 0.03; 95% CI -0.10 to 0.16). Six patients reported adverse effects at both the SSD site and the Dimac SSD site.

Summary for burns: SSD versus silver
Two trials investigated partial-thickness burns and four trials full thickness or severe burns. Only two out of six trial reports described adequate sequence generation, and none described adequate allocation concealment.
Infection rates were reported in six trials. No statistically significant differences were found in five trials,\(^5\)_1,\(^6\)_0,\(^6\)_7,\(^6\)_8 though one trial showed a statistically significant difference for the number of positive-culture swabs in favour of the synthetic silver dressings.\(^5\)_4

Time to complete healing was reported in two trials. One trial used appropriate analysis methods and showed no statistically significant differences,\(^5\)_1 and the second trial analysed a time-to-event outcome (time to complete healing) inappropriately as a continuous variable\(^5\)_2 and showed a statistical significance in favour of the SSD cerium nitrate group.

The number of wounds healed was reported in three trials. None of the trials showed statistically significant differences.\(^5\)_1,\(^5\)_2,\(^5\)_4

Pain was reported in three trials. One trial showed no statistically significant differences,\(^6\)_0 while two trials showed a statistically significant difference in favour of the silver-containing dressings Acticoat\(^\circledR\) and Aquacel\(^\circledR\) Ag.\(^5\)_1,\(^6\)_8

1.4 Comparisons between alternative silver-containing dressings including dose comparisons (silver versus silver)

1.4.1 Nanocrystalline silver-coated dressing (Acticoat\(^\circledR\)) compared with fine-mesh gauze with silver nitrate (0.5%) (one trial)

Tredget\(^7\)_2 enrolled 30 patients with 60 deep partial- and full thickness burns.

**Primary outcome: infection rate**

Wound infection was defined as present when there were more than 10\(^5\) organisms per gram of tissue present. Bacteraemia was defined as the presence of the same bacterium isolated from the blood and the burn wound at concentrations of more than 10\(^5\) organisms per gram of tissue. Significantly fewer patients developed a wound infection in the nanocrystalline silver-coated (Acticoat\(^\circledR\)) group (5/17 in the Acticoat\(^\circledR\) group; 16/17 in the fine-mesh gauze with silver nitrate group) (RD -0.65; 95% CI -0.89 to -0.40). The NNT with nanocrystalline silver was two, in order to prevent one infection. There was no statistically significant difference between groups in the number of patients who developed bacteraemia (1/17 in the Acticoat\(^\circledR\) group and 5/17 fine-mesh gauze with silver nitrate group) (RD -0.24; 95% CI -0.48 to 0.01).

**Primary outcome: wound healing rate**

Healing was defined as complete re-epithelialisation; the authors reported there was no difference between the treatments, but no data were reported to support this statement.

**Secondary outcome: pain**

Pain was measured on a visual analogue scale from 1 (not painful) to 5 (very painful). Only the mean pain score on dressing removal was significantly lower in the Acticoat\(^\circledR\) group, but not the mean overall pain score (MD -0.28; 95% CI -0.93 to 0.37).
2. Acute wounds: other wounds

2.1 Topical silver-containing agents compared with topical agents without silver (SSD versus no silver)

Diren enrolled 465 patients with minor, uncomplicated, soft tissue wounds requiring sutures into a study that compared three antimicrobial regimens with paraffin-impregnated gauze. Data from 39 enrolled participants were excluded for protocol violations, so only 426 participants were included in the analysis (i.e. not analysed by intention-to-treat). The trial had four treatment groups in which the following numbers of participants completed the trial: SSD cream (99 participants), bacitracin zinc ointment (109 participants), neomycin sulfate (110 participants), and petrolatum (108 participants). We compared each of these antimicrobial alternatives with SSD cream.

Wound infection was defined as any subjective or objective sign or symptom of infection, e.g. fever, erythema, oedema, induration, tenderness, heat, exudate, adenopathy, and lymphangitis. Wounds were classified into one of five categories based upon clinical assessment, ranging from no signs of infection (384 participants), simple stitch abscess (25 participants), surrounding cellulitis (14 participants), accompanying lymphangitis (three participants), and systemic symptoms (no participants).

2.1.1 SSD cream compared with bacitracin zinc ointment

*Primary outcome: infection rate*

There was no statistically significant difference between groups in the number of patients who developed wound infections (12/99 in the SSD group; 6/109 in the bacitracin zinc group) (RD 0.07; 95% CI -0.01 to 0.14) (Analysis 1.1).

2.1.2 SSD cream compared with neomycin sulfate

*Primary outcome: infection rate*

Significantly fewer patients developed wound infections in the neomycin sulfate group (12/99 in the SSD group; 5/110 in the neomycin sulfate group) (RD 0.08; 95% CI 0.00 to 0.15) (Analysis 1.1). The NNT with neomycin sulfate was 13, in order to prevent one infection.

2.1.3 SSD cream compared with petrolatum

*Primary outcome: infection rate*

There was no statistically significant difference between groups in the number of patients who developed wound infections (12/99 in the SSD group; 19/108 in the petrolatum group) (RD -0.05; 95% CI -0.15 to 0.04) (Analysis 1.1).
2.2 Dressings containing silver compared with dressings without silver (silver versus no silver)

2.2.1 Hydrofibre dressing containing ionic silver (Aquacel® Ag) compared with povidone iodine gauze (one trial)
Jurczak\textsuperscript{64} enrolled 67 patients with open surgical wounds or open traumatic wounds all healing by secondary intention to a randomised controlled trial comparing silver-containing hydrofibre (hydrofibre-Ag) with povidone iodine gauze.

Primary outcome: infection rate
Wound infection was defined on clinical criteria such as warmth, redness, increased tenderness, swelling, increased exudate or purulent discharge, and malodour. There was no statistically significant difference in the number of patients who developed a wound infection during the trial period (4/35 in the Aquacel®Ag group; 4/32 in the povidone iodine group) (RD -0.01; 95% CI -0.17 to 0.14) (Analysis 1.1).

Primary outcome: wound healing rate
Healing was defined as epithelialisation, but also reduction in wound area in mm\textsuperscript{2}, and reduction in wound depth in mm were reported. The mean time to complete healing was 14.1 days in the Aquacel® Ag group and 13.9 days in the povidone iodine group (log-rank test: not statistically significant). There was no statistically significant difference in the number of patients with complete wound healing at two weeks (8/35 in the Aquacel® Ag group; 3/32 in the povidone iodine group) (RD 0.13; 95% CI -0.04 to 0.31) (Analysis 1.2).

The authors stated that the adjusted mean reduction in wound area was 551 mm\textsuperscript{2} in the Aquacel®Ag group and 401 mm\textsuperscript{2} in the povidone iodine group. The adjusted mean reduction in wound depth was 9 mm in the Aquacel® Ag group and 10 mm in the povidone iodine group. How, and why, the adjustment was made was not reported. The authors stated that both reductions were statistically significant when compared with baseline, but, when compared with each other, no statistically significant difference was found. No SDs were reported; therefore the mean difference could not be calculated.

Secondary outcome: adverse events
Adverse events were defined as any event that occurred during the trial period, e.g. allergy, skin burn, haemorrhage. There was no statistically significant difference between the groups (RD -0.09; 95% CI -0.21 to 0.02).

Secondary outcome: pain
Pain was measured on a visual analogue scale from 1 (no pain) to 10 (worst pain imaginable). Although no statistically significant differences were found for the pain score at dressing removal and application, the decrease in mean pain score from baseline when the dressings were in place was -0.7 for Aquacel® Ag versus 0 for povidone iodine gauze, though no SD was given. The overall ability to manage pain could be scored as excellent, good, fair
or poor. The pain management was evaluated at the final visit (i.e. when the wound was completely healed or at week 2). Overall 70.6% of participants rated pain management as excellent in the Aquacel®Ag group compared with 22.6% in the povidone iodine gauze group. There was a statistically significant difference in the ability to manage pain in favour of the Aquacel® Ag group; P value < 0.001.

**Summary for acute wounds: SSD/silver versus no silver**

One of the two trials reported adequate sequence generation and adequate allocation concealment.

Infection rate was reported in both trials with a total of four different dressing comparisons. Three comparisons were not statistically significantly different, and one comparison showed a statistically significant difference in favour of neomycin sulfate.

Time to complete healing was reported in one trial, and was not statistically significant.

The number of wounds healed was reported in one trial, and was not statistically significant.

Pain was reported in one trial, and showed a statistically significant difference in favour of hydrofibre dressing containing ionic silver.

### 3. Chronic wounds

#### 3.1 Topical silver-containing agents compared with topical agents without silver (SSD versus no silver)

##### 3.1.1 SSD cream compared with Bensal HP with QRB7 (one trial)

Jacobs enrolled 40 patients with Wagner grade 1 or 2 diabetic foot ulcers in a trial comparing SSD with Bensal HP with QRB7, which is a mixture of 6% benzoic acid, 3% salicylic acid and 3% extract of Q rubra (an extract of oak (Quercus rubra) bark).

**Primary outcome: infection rate**

Wound infection was defined on the basis of clinical signs (foul odour, exudation, or erythema) and bacterial cultures. None of the treated wounds demonstrated growth of pathogenic bacteria at six weeks (Analysis 1.1).

**Primary outcome: wound healing rate**

Healing was defined as the percentage reduction in total wound size (derived by adding the individual wound areas for each participant in each group) at two, four and six weeks. Complete healing was not defined. The “collective” wound diameter of the Bensal HP-treated patients had decreased by 72.5%, whereas the collective diameter of the SSD group had reduced by 54.7% (Student t test: P value 0.059). There was no statistically significant difference in the number of patients with complete wound healing within six
weeks (6/20 in the SSD group; 8/20 in the Bensal HP group) (RD -0.10; 95% CI -0.39 to 0.19) (Analysis 1.3).

Secondary outcome: adverse events
None of the patients experienced adverse effects.

3.2 Dressings containing silver compared with non-silver dressings (silver versus no silver)

3.2.1 Activated-charcoal dressing containing silver (Actisorb Plus®) compared with conventional phase-adapted therapy using diverse topical modalities (one trial)
Wunderlich\(^ 73\) enrolled 40 patients with venous leg ulcers of whom 38 were followed to study completion.

Primary outcome: infection rate
Every two weeks swabs were taken and were rated as 0 (no bacterial growth), 1 (light bacterial growth), 2 (medium bacterial growth), or 3 (heavy bacterial growth). The authors reported no differences in infection rates, but no actual data were reported.

Primary outcome: wound healing rate
Healing was defined as granulation (on an ordinal scale from 0 to 3), epithelialisation (on an ordinal scale from 0 to 3), and also as the reduction of the mean ulcer area in cm\(^ 2\). There was no statistically significant difference in the number of patients healed after six weeks of treatment (6/19 patients in the charcoal-silver group; 2/19 patients in the conventional phase-adapted therapy using diverse topical modalities group) (RD 0.21; 95% CI -0.04 to 0.46) (Analysis 1.3).

3.2.2 Hydrofibre dressing containing ionic silver (Aquacel® Ag) compared with calcium alginate dressing (Algosteril®) (one trial)
Jude\(^ 63\) enrolled 434 patients with diabetic foot ulcers (Wagner grade 1 and 2). Although, at baseline, the calcium alginate dressing group (Algosteril®) seemed to have larger ulcers, and more patients in the hydrofibre dressing containing ionic silver (Aquacel® Ag) group were receiving antibiotics, the authors stated that the groups were comparable.

Primary outcome: infection rate
Wound infection was defined on the basis of clinical signs and/or bacterial cultures. There was no statistically significant difference in the number of patients who developed wound infection (11/67 in the Aquacel® Ag group; 8/67 in the Algosteril® group) (RD 0.04; 95% CI -0.07 to 0.16) (Analysis 1.1).

Primary outcome: wound healing rate
Healing was defined as complete re-epithelialisation, and as the reduction of the mean ulcer area in percentage and ulcer depth. Healing speed was defined as a weekly reduction in
absolute and percentage ulcer area. Only the mean time to complete healing was significantly lower in the Aquacel® Ag group (52.6 days +/-1.8 days (SD) Aquacel® Ag group; 57.7 days +/- 1.7 days (SD) Algosteril® group) (MD -5.1; 95% CI -5.69 to -4.51) (Analysis 1.2). Time in days to 100% healing was estimated by Kaplan-Meier survival analysis. The number of patients with complete wound healing within eight weeks was 21/67 in the Aquacel® Ag group and 15/67 in the Algosteril® group (RD 0.09; 95% CI -0.06 to 0.24) (Analysis 1.3). The mean percentage ulcer area reduction by eight weeks was 58.1% in the Aquacel® Ag group and 60.5% in the Algosteril® group (MD -2.4; 95% CI -18.72 to 13.92). The reduction in mean ulcer depth at eight weeks was 0.25 cm in the Aquacel® Ag group and 0.13 cm in the Algosteril® group (MD 0.12; 95% CI -0.05 to 0.29).

Secondary outcome: adverse events
Adverse events were not clearly defined. One of the events mentioned was infection. There was no statistically significant difference in the number of patients who experienced adverse effects (25/67 in the Aquacel® Ag group; 26/67 in the Algosteril® group) (RD -0.01; 95% CI -0.18 to 0.15).

Summary for chronic wounds: SSD/silver versus no silver
Two of the three trials reported adequate sequence generation, and none adequate allocation concealment.

Infection rate was reported in three trials, and showed no statistically significant differences.

Time to complete healing was reported in one trial, and was significantly faster with the silver hydrofibre (Aquacel® Ag) dressing. Time to healing was appropriately analysed using survival analysis.

The number of wounds healed was reported in all three trials, and showed no statistically significant difference.

4. Mixed wounds

4.1 Topical silver-containing agents compared with topical agents without silver (SSD versus no silver)
Hutchinson enrolled 292 patients with venous leg ulcers, partial-thickness burns or partial-thickness donor sites. The trial had three treatment groups; SSD cream/hydrocolloid (58 participants), hydrocolloid alone (108 participants), and non-occlusive paraffin impregnated gauze (126 participants). The results are presented comparing SSD cream to each of the comparators. Wound infection was defined using clinical criteria such as erythema, oedema, pain and purulent discharge.
4.1.1 SSD cream/hydrocolloid compared with hydrocolloid alone (one trial)

*Primary outcome: infection rate*

There was no statistically significant difference in the number of patients who developed a wound infection (0/58 in the SSD/hydrocolloid group, and 2/108 in the hydrocolloid group) (RD -0.02; 95% CI -0.06 to 0.02) (Analysis 1.1).

4.1.2 SSD cream/hydrocolloid compared with non-occlusive paraffin impregnated gauze

*Primary outcome: infection rate*

Significantly fewer patients in the SSD/hydrocolloid group developed a wound infection when compared with the non-occlusive paraffin impregnated gauze group (0/58 in the SSD/hydrocolloid group; 7/126 in the non-occlusive paraffin impregnated gauze group) (RD -0.06; 95% CI -0.10 to -0.01) (Analysis 1.1). The NNT with SSD/hydrocolloid was 18, in order to prevent one infection.

**Summary for mixed wounds: SSD versus no silver**

This trial did not report adequate sequence generation or adequate allocation concealment, therefore effect estimates may be biased.

Infection rate was reported in this trial with a total of two different dressing comparisons. One comparison showed a statistically significant difference in favour of SSD/hydrocolloid, and the other showed no differences.

**Summary for all wounds: SSD/silver versus no silver**

*Infection rate*

Infection rates were reported in 17 trials with a total of 21 different dressing comparisons. One comparison showed a statistically significant difference in favour of silver nitrate dressings,\(^{65}\) 15 comparisons showed no differences, and five comparisons using SSD showed a statistically significant difference in favour of non-silver dressings.\(^{50,53,59,69,71}\)

*Wound healing rate*

Time to complete wound healing was reported in eleven trials. One trial showed a statistically significant difference in favour of hydrofibre dressing with ionic silver,\(^{63}\) three trials showed no differences,\(^{49,58,64}\) and seven trials showed a statistically significant difference in favour of non-silver dressings.\(^{55-57,61,69,70,74}\) In most cases, time to complete wound healing was inappropriately regarded as a continuous outcome and the analysis of these outcomes was, therefore, flawed, leading to potentially misleading results. Eight trials reported the number of wounds completely healed. Five trials showed no differences,\(^{50,62-64,73}\) and three trials showed a statistically significant difference in favour of non-silver dressings.\(^{61,66,71}\)
Adverse events
Adverse events were reported in four trials. None of them showed statistically significant differences.\textsuperscript{58,62-64}

Pain
Pain was reported in nine trials, but was expressed in different ways, e.g. the need for analgesia, or on a visual analogue scale (VAS). Overall, the reported pain scores were low in the majority of these trials, and the absolute differences in pain scores between the studied interventions were minimal. Two trials showed a statistically significant difference in favour of silver-containing dressings,\textsuperscript{57,64} two trials found no differences,\textsuperscript{49,58} and five trials showed a statistically significant difference in favour of non-silver dressings.\textsuperscript{50,55,56,66,74}

Patient satisfaction
Patient satisfaction was reported in one trial,\textsuperscript{49} and showed no statistically significant differences.

Length of hospital stay
Length of hospital stay was reported in two trials, with a total of three dressing comparisons. Only one patient group treated with silver nitrate for burns covering 20% to 40% of the total body surface area experienced significantly shorter hospital stay compared with participants who received Ringer’s lactate.\textsuperscript{65} No statistically significant differences were present for any of the other groups,\textsuperscript{65} or trial.\textsuperscript{50}

Costs
Costs were reported in four trials. One trial,\textsuperscript{61} found that the mean costs per cm\textsuperscript{2} of dressing - based on price lists supplied by the manufacturers - were lower in the non-silver dressings group, compared with the silver-containing dressing group. One trial reported costs of dressings per percent of body surface burnt,\textsuperscript{66} but differences were not reported. Both of the remaining two trials showed a statistically significant difference in favour of non-silver dressings.\textsuperscript{55,56}

DISCUSSION
This review highlights the lack of conclusive evidence on the effects of silver-containing dressings or agents to prevent wound infection and to promote wound healing. In particular, there was no evidence to support the use of SSD for prevention of wound infection in patients with partial-thickness burns. None of the trials indicated a beneficial effect for SSD for other outcomes when compared with other silver-containing or non-silver dressings. Furthermore, there was evidence that SSD may delay wound healing, may be more expensive, and may be more painful when applied to burns. The few trials on full-thickness burns and acute, chronic, or mixed wounds showed insufficient evidence for a beneficial
effect of silver-containing dressings to decrease infection rates and to aid wound healing. Only one trial showed significantly better results in terms of infection rates when another agent was added to the silver-containing dressing: infection rates were significantly lower than with SSD cream alone when a synthetic dressing was added to SSD cream (Hydron-SSD). The nanocrystalline form of silver present in the Hydron-SSD dressing, which releases silver ions faster, might explain the better results in burns. Furthermore, most trials used 1% SSD cream, but its effect might be dose-related. On the other hand, higher doses could also result in higher toxicity and more adverse effects.

Recently published literature had already suggested the lack of evidence of effectiveness for silver-containing dressings and topical agents in burns. Hussain published a Best Evidence Topic report on burns, including evidence from RCTs and CCTs. The authors concluded that there was little evidence for using silver-containing dressings to prevent wound infection, and that such products tend to delay wound healing. Furthermore, silver may have serious cytotoxic activity on various host cells.

In minor thermal burns (less than 15% TBSA) SSD cream was found to delay healing time and increase pain when compared with other treatments. Wasiak also evaluated different dressings for burn wounds and found evidence for a delayed healing time for SSD. Similarly, Bergin found no RCTs that evaluated the effects of silver-containing dressings for the treatment of diabetic foot ulcers, and Vermeulen found three RCTs and concluded that there was insufficient evidence of effectiveness for silver-containing dressings as a treatment for infected wounds.

**The following limitations of this review should be noted**

Firstly, the methodological quality of the 26 included trials was relatively low, and a large proportion of the evidence presented here is accrued from trials which demonstrate a high or uncertain risk of bias. Most of the studies had small sample sizes and were, therefore, at risk of not detecting any existing differences, and of incurring chance baseline imbalances for important prognostic factors. Only one-third of the trials reported adequate sequence generation, and even fewer reported allocation concealment.

Blinding of participants and care providers was not really possible, but outcome assessors could have been blinded, or healing confirmed by blinded assessment of photographs. This was almost never achieved or reported. Similarly the drop-out rate or reasons for drop-out were not always described.

The duration of follow-up of the included studies ranged from a few days to more than three months, whilst in only five studies was follow-up continued until complete wound re-epithelialisation was achieved. In some trials the length of follow-up was unclear, or too short, and almost half of the trials were supported financially by a single manufacturer. If this caused publication bias - which was shown to be present in studies on negative pressure wound therapy - the real effect is likely to be even less favourable.
Secondly, one of the strengths of a systematic review is the ability to pool data from several - often small - trials to achieve greater statistical power and a more precise overall effect size estimate. In this review few data could be pooled because the trials did not compare similar interventions, and there was considerable heterogeneity in the wounds being compared. Therefore, the lack of conclusive evidence for the effects of silver-containing dressings remains.

Thirdly, some trials used repeated measurements, for example, healing rate or swabs taken (e.g. at three, six, or nine days for one endpoint). This may illustrate the eagerness of the investigators (or the sponsors) to identify any sign of a treatment difference, at the cost of an increased chance of false positive results, while the shorter intervals are not relevant to patients. Furthermore, outcome parameters were measured in different ways and on different scales. Many secondary outcomes were based on subjective concepts such as “ease of use”, “comfortable to wear”. These subjective findings can hardly help in clinical practice and should be measured with standardised objective measurements whenever possible. Also, some trials measured “time per dressing”, or “costs per cm²”. These measures alone are meaningless and should be reported in combination with other aspects of costs.

Fourthly, the majority of studies that reported outcomes such as time to healing or time to skin grafting, incorrectly reported and analysed these outcomes as continuous - rather than time-to-event - variables. The problem with this approach is that the time to the event is only known for those people who actually experienced it (in this case healing, or grafting), and no information is obtained from those who were observed, but did not experience the event. This approach may introduce bias. Time-to-event data, such as time to wound healing, should be analysed using survival analysis in which the treatment effect is expressed as a hazard ratio.

Finally, eight trials did not attempt to define infection. Some trials defined infection only on clinical grounds and others merely on the presence of bacterial cultures. It is clearly difficult to interpret the results of studies that do not define their main outcomes. We reported the definition of infection and healing as used by the study authors and were unable to conduct any pooling due to heterogeneity.

Apart from the definition used, Sibbald stated that chronic wounds always contain bacteria and a diagnosis of infection should be based on clinical signs and not solely on bacterial cultures.

**AUTHORS’ CONCLUSIONS**

**Implications for practice**
There is currently insufficient evidence that silver-containing dressings prevent wound infection or promote wound healing; the available evidence is low both in volume and
quality. There is some evidence from small, poor-quality trials, that SSD does not reduce wound infection and slows down wound healing in people with partial-thickness burns.

Implications for research
More studies, and particularly studies with a low risk of bias, are needed to confirm any effect of silver-containing dressings in full thickness burns and other wound groups. Future research must develop clear, valid, and reliable measures of wound infection. The use of common, quantifiable, and clinically-relevant endpoints (time to complete wound healing, number and time to wound infection, pain, adverse events, costs, and, preferably, a validated scale for patient satisfaction) should always be used. Whilst it is very difficult to blind patients and medical professionals with regard to the intervention, it is possible to blind outcome assessors, or to use computer programmes to measure wound size. Future research must adopt a survival approach for the analysis of time-to-event data, such as time to healing. Finally, a sufficiently long follow-up period of at least six months is essential if treatment effects in chronic wounds are to be detected. Interventions under evaluation should be thoroughly, and clearly, described. For this purpose use of the revised CONSORT statement is recommended in order to report these trials adequately.
REFERENCES


13. Moralejo D, Jull A. Handrubbing with an aqueous alcohol solution was as effective as handscrubbing with an antiseptic soap for preventing surgical site infections. EBN 2003;6(2):54–5.


59. Hutchinson JJ. A prospective clinical trial of wound dressings to investigate the rate of infection under occlusion. 3rd European Conf on Advances in Wound Management; 19-22 October; Harrogate, UK. 1993.
60. Inman RJ, Snelling CF, Roberts FJ, et al. Prospective comparison of silver sulfadiazine 1 per cent plus chlorhexidine digluconate 0.2 per cent (Silvazine) and silver sulfadiazine 1 per cent (Flamazine) as prophylaxis against burn wound infection. Burns 1984;11(1):35–40.


PART 2

EFFECT OF EVIDENCE-BASED DECISIONS ON QUALITY OF CARE
Influence of the Manchester Triage System on waiting time, treatment time, length of stay and patient satisfaction; a before and after study
ABSTRACT

**Objectives:** To compare waiting time, treatment time, length of stay (LOS), patient satisfaction, and distribution of waiting times over levels of urgency before and after the implementation of the Manchester Triage system (MTS) at an Emergency Department (ED).

**Methods:** Before and after study, by means of timeline measurements and questionnaires on satisfaction in two consecutive patient series (n=1808). Questionnaires covered aspects of provision of information, opportunity given to explain problems, waiting time, and sorting out the problem. After implementation of MTS, patients were triaged between 12 noon and 10 pm. Sub-analysis was performed on triaging and non-triaging; and between urgency levels.

**Results:** Waiting time did not decrease after implementation of the MTS, however treatment time and LOS were significantly longer. Total LOS did not differ. After implementation, waiting time was better distributed over urgency levels. Furthermore, after implementation, patient satisfaction scored significantly lower on the provision of information and opportunity to explain their problems, however waiting time and the feeling that their problem had been sorted out scored better. No significant differences were found between triaged and non-triaged patients. Although not significant, patients in the lower urgency levels seemed more satisfied than patients in the higher urgency levels.

**Conclusion:** Implementing MTS on its own is not sufficient to improve efficiency and quality of EDs. More complex interventions including process redesigning that targets various groups of ED patients should be evaluated in the future by using rigorous research designs for quality improvement of EDs.
INTRODUCTION

Annually more than two million people attend Dutch hospital Emergency Departments (EDs). An increasing number of patients are bypassing the general practitioner (GP) and attending with non-urgent problems. This has caused overcrowding at the ED\(^1\) which has become a serious problem leading to long waiting times, patient dissatisfaction\(^2\) and putting patients at risk for poor outcome.\(^3,4\) For these reasons the implementation of accurate triage systems at EDs is advocated.\(^5\)

Although actual waiting time is important, it is perceived waiting time which is strongly correlated with patient satisfaction.\(^6\) Patient satisfaction may seem to be a secondary outcome, but it is important as satisfied patients are more likely to comply with treatment and therefore achieve better clinical outcomes than dissatisfied patients.\(^7\) Consequently, in recent decades patient perception of quality of care at an ED has become an increasingly important concept in continuous quality improvement.\(^6,8\)

Patient satisfaction, waiting time and length of stay (LOS) relating to the Emergency Severity Index (ESI) and the Canadian Emergency Department Triage and Acuity Scale,\(^9,10\) have all been studied but to our knowledge no studies on the Manchester Triage System (MTS) have been published on patient satisfaction. In previous research, we reported on validity and accuracy, e.g. the relationship of triage ratings with different patient outcomes.\(^11,12\) Following these results, the MTS was implemented at our hospital in October 2007. The MTS uses flowcharts, which are based on a five-step decision process, to assign patients to one of five triage categories.\(^13\) A color indicates the level of urgency and its associated maximum waiting time: red - immediate care by a physician; orange - 10 minutes; yellow - 60 minutes; green - 2 hours; and blue - 4 hours.

The objective of this study was to compare waiting time, treatment time, LOS, patient satisfaction, and distribution of waiting times over the levels of urgency, before and after the implementation of the MTS.

METHODS

Study design, study setting and population

In this prospective, single-center, before and after study, we compared waiting time, treatment time, LOS, and patient satisfaction. The study was carried out in the Netherlands at a university teaching hospital with a Level 1 trauma center. Annually more than 31,000 patients attend this hospital’s ED, 750 of whom (2%) are seen in the trauma resuscitation room. Approximately two years before the implementation of the MTS, 906 consecutive patients were included over a 10-day period. Six months after the implementation of the MTS, a further 900 consecutive patients were included. The local institutional review board waived the requirement for written informed consent from the patients.
**Implementation of the Manchester Triage System**

Before implementation of the MTS in October 2007, an informally-structured triage system was used. This system was based on clinical expertise, but not on explicit criteria and information. When registered at the ED, based on patients’ appearance and presenting complaints, patients were (implicitly) judged by the ED nurse or an experienced receptionist if they could wait safely or have to be seen immediately. All registered ED nurses received a combination of didactic and practical training in the MTS, in accordance with national standards. After implementation of the MTS, patients arriving at the ED between 12 noon and 10 pm were triaged. This group comprised 70% of all patients attending the ED over a 24-hour period. Outside these times, patients were registered and implicitly judged by an ED nurse, when necessary the nurse triaged patients according to the MTS. Patients already triaged before hospital arrival by ambulance staff, were not triaged again, but placed in the treating room according local guidelines.

**Study protocol**

*Data collection and definitions*

We collected the following patient data: patient characteristics, mode of arrival, mode of referral, triage ratings by a triage nurse, and admission to hospital. Timeline information was extracted from ED forms. Times recorded routinely were: time of registration, time of start of triage, time to treatment and time of leaving ED. These were supplemented with time from arrival at the ED, and time leaving the triage room. Patients entered this additional information onto an extended time measurement form. Arrival time was defined as the time from arrival at ED to registration; waiting time as the time in minutes from registration to entering the treatment room; time to triage as the time from registration to triage; triage time as the time from start triage to end triage; triage to treatment as the time from end triage to entrance the treatment room; treatment time as the time from entering the treatment room to discharge or admission; and the LOS as the time from registration to discharge or admission. Following local guidelines, diagnostic procedures (for example blood samples) started before seeing a doctor. Therefore, ‘time entering the treatment room’ seems more accurate and closer to real life practice then ‘time seeing a doctor’.

*Informed consent and patient instruction*

In both measurements, on arrival at the department, patients were given an information letter and asked to participate. After giving oral consent, they received a numbered patient form, a watch and a pencil and instructions about how to participate. To prevent patients becoming confused between triage and treatment, the patient form included a clear explanation and photographs of the triage room. The patient was asked to return the completed form to the receptionist on leaving the ED. If the patient refused or was unable to fill in the form, only the routinely recorded time points from the ED forms was used.
Patient satisfaction
To measure patient satisfaction, four questions were asked that reflected identified service-related factors during their ED stay.\textsuperscript{14,15} Patients were asked to rate: (1) the spoken information they received before they were seen or treated by a doctor, (2) the opportunity they were given to explain their problems, (3) the length of waiting time, and (4) the feeling that their complaint or problem had been sorted out. This was done on a five-point Likert-type scale; very satisfied, satisfied, reasonably satisfied, fairly satisfied and not satisfied. In the Before Implementation group, patients answered these four questions at the ED. In the After Implementation group, these four questions made up part of a more extensive questionnaire on patient satisfaction. This questionnaire was not filled out at the ED but handed over to be completed at a later time. Patients were asked to fill out this questionnaire and return it within two weeks. If patients were not given the questionnaire, it was sent by post. Reminders were sent within a month.

Data analysis
Data were analyzed using SPSS version 14.0, (Chicago, IL, USA). Patient characteristics were reported descriptively. The chi-squared test was used for categorical data, the independent sample two-tailed t-test for normal, continuous data, the Mann-Whitney test or the Kruskal Wallis test for non-normally distributed data, to compare the before and after implementation data. The level of significance was 0.05.

Sub-analysis of timeline data was carried out on mode of referral, mode of arrival, triaging and non-triaging of patients; and between the urgency levels of the patients triaged. In triaged patients, we also determined the number of patients who after being triaged entered the treatment room within the maximum waiting time defined by the MTS.\textsuperscript{13}

These comparisons were defined a priori. Exploring differences in timeline data, type 2 errors are not warranted, therefore the Bonferroni test was not considered to correct for multiple comparisons.\textsuperscript{16}

To compare patient satisfaction scores, we re-expressed the Likert scale into three levels: the ratings very satisfied and satisfied became satisfied; the ratings reasonably satisfied and fairly satisfied became fairly satisfied; and the rating not satisfied remained the same. Sub-analysis of satisfaction data was performed on triaging and not triaging, and between the urgency levels of the patients triaged.

RESULTS

Patient characteristics and response rate
Before implementation 907 patients were included, and after implementation 901. Five ED forms were incomplete and thus excluded: one in the Before Implementation group and
Table 1: Patient characteristics

<table>
<thead>
<tr>
<th></th>
<th>Before (N=906)</th>
<th>After (N=897)</th>
<th>After Triaged patients (N=310)</th>
<th>After Non-triaged patients (N=587)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Male</td>
<td>494</td>
<td>55</td>
<td>453</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>168</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>285</td>
<td>46</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Median (yr)</td>
<td>36</td>
<td>36</td>
<td>32</td>
<td>38</td>
</tr>
<tr>
<td>Range (yr)</td>
<td>0.94</td>
<td>0.97</td>
<td>0.91</td>
<td>0.97</td>
</tr>
<tr>
<td>IQR (25-75)</td>
<td>21.54</td>
<td>19.54</td>
<td>17.48</td>
<td>21.57</td>
</tr>
<tr>
<td>Age distribution</td>
<td></td>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>&lt;15 years</td>
<td>138</td>
<td>15</td>
<td>67</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>103</td>
<td>18</td>
</tr>
<tr>
<td>15-30 years</td>
<td>226</td>
<td>25</td>
<td>80</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>134</td>
<td>23</td>
</tr>
<tr>
<td>30-45 years</td>
<td>234</td>
<td>26</td>
<td>73</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>118</td>
<td>20</td>
</tr>
<tr>
<td>45-60 years</td>
<td>154</td>
<td>17</td>
<td>57</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>110</td>
<td>19</td>
</tr>
<tr>
<td>60-75 years</td>
<td>82</td>
<td>9</td>
<td>25</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>75</td>
<td>13</td>
</tr>
<tr>
<td>&gt; 75 years</td>
<td>72</td>
<td>8</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>47</td>
<td>8</td>
</tr>
<tr>
<td>Mode of arrival</td>
<td></td>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Private vehicle</td>
<td>749</td>
<td>83</td>
<td>308</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>441</td>
<td>75</td>
</tr>
<tr>
<td>By ambulance</td>
<td>157</td>
<td>17</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>146</td>
<td>25</td>
</tr>
<tr>
<td>Mode of referral</td>
<td></td>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Self-referral</td>
<td>675</td>
<td>75</td>
<td>231</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>399</td>
<td>68</td>
</tr>
<tr>
<td>Referred by GP</td>
<td>231</td>
<td>25</td>
<td>79</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>188</td>
<td>32</td>
</tr>
<tr>
<td>Number of admissions</td>
<td>160</td>
<td>18</td>
<td>37</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>154</td>
<td>26</td>
</tr>
</tbody>
</table>

GP: General physician; IQR: Interquartile range; Before: Before Implementation group; After: After Implementation group

Table 2: Median waiting and throughput time: for Before and After Implementation groups.

<table>
<thead>
<tr>
<th></th>
<th>Entrance time (min-max)</th>
<th>Waiting time (min-max)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before N = 469</td>
<td>After N = 353</td>
</tr>
<tr>
<td>Total</td>
<td>0.05 (0-0.38)*</td>
<td>0.02 (0-0.27)*</td>
</tr>
<tr>
<td>Self-referral</td>
<td>0.05 (0-0.36)*</td>
<td>0.02 (0-0.20)*</td>
</tr>
<tr>
<td>Referred by GP</td>
<td>0.05 (0-0.38)*</td>
<td>0.02 (0-0.27)*</td>
</tr>
<tr>
<td>Private vehicle</td>
<td>0.06 (0-0.38)*‡</td>
<td>0.02 (0-0.27)*</td>
</tr>
<tr>
<td>Ambulance</td>
<td>0.00 (0-0.30)‡</td>
<td>0.02 (0-0.20)</td>
</tr>
</tbody>
</table>

Data is presented as hours and minutes. * significance level p<0.05, between before and after implementation group; † significance level p<0.05 between self-referral and patients referred by general physician (GP); ‡ significance level p<0.05 between private vehicles and ambulance; Arrival time: time from arrival to registration; Waiting time: time from registration to entrance of the treatment room; Treatment time: time from entrance of the treatment room to discharge; Length of stay: time from registration to discharge; Before: Before Implementation group; After: After Implementation group
four in the After Implementation group. Gender, age, mode of arrival, admission rate and patients already triaged by ambulance staff before arrival at hospital did not differ between groups, although age distribution did differ. Triaged patients only included those who arrived by private vehicle. In the After Implementation group significantly more patients were presented by their GP (Table 1).

In the Before Implementation group, 356 (39.3%) and in the After Implementation group 286 (31.9%) patients completed the patient satisfaction questionnaire. No significant differences were found in sex, admission and mode of arrival between responding and non-responding patients.

**Waiting time, treatment time and length of stay**

*Before versus after implementation*

Waiting time did not differ between before and after implementation of the MTS, but arrival time, treatment time and total LOS did (Table 2). If a distinction was made for mode of referral, arrival time in self-referred and in GP-referred patients remained significant, as did the LOS in self-referred patients. For mode of arrival, arrival time, treatment time and LOS remained significant in patients arriving by private vehicle, but not in patients arriving by ambulance.

Waiting time from registration to the treatment room was longer for triaged than for non-triaged patients, but treatment time was shorter. Total LOS did not differ (Table 3).

If the urgency level was higher, the triage time was longer. Also time from triage to the treatment room, waiting time, treatment time and LOS were associated with urgency level (Table 3). After being triaged, the maximum waiting time per urgency level was only exceeded in 3 patients triaged to the green level (Table 4).

<table>
<thead>
<tr>
<th>Treatment time (min-max)</th>
<th>Length of stay (min-max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before N = 844</td>
<td>After N = 690</td>
</tr>
<tr>
<td>1.06 (0.01-9.05)*</td>
<td>1.20 (0-12.30)*</td>
</tr>
<tr>
<td>0.50 (0.01-9.05)†</td>
<td>0.55 (0-12.30)†</td>
</tr>
<tr>
<td>2.33 (0.15-7.52)†</td>
<td>2.25 (0.15-10.45)†</td>
</tr>
<tr>
<td>0.57 (0.01-9.05)*‡</td>
<td>1.06 (0-12.30)*‡</td>
</tr>
<tr>
<td>2.25 (0.11-7.05)‡</td>
<td>2.15 (0.05-10.26)‡</td>
</tr>
</tbody>
</table>
Patient satisfaction

Based on the four service-related questions, in the Before Implementation group patient satisfaction concerning the provision of information and the opportunity to explain their problems was significantly higher. Satisfaction with length of waiting time was significantly higher in the After Implementation group as was the level of satisfaction about the way that their problem had been sorted out (Table 5).

No significant differences were found between triaged and non-triaged patients (Table 5). Also, no significant differences were found between urgency levels in the triaged patients, although overall, patients in the green level scored higher satisfaction rates on all questions than patients in the yellow level (Table 5).

Table 3: After implementation group: median waiting and throughput time, for triaged and non-triaged patients.

<table>
<thead>
<tr>
<th>Urgency level</th>
<th>Maximum waiting time †</th>
<th>Entrance time (min-max)</th>
<th>Time to triage (min-max)</th>
<th>Triage time (min-max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-triaged</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>587</td>
<td>0.02 (0-0.27)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Triaged</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>291†</td>
<td>0.02 (0-0.20)</td>
<td>0.04 (0-0.35)</td>
<td>0.04 (0.01-0.16)</td>
</tr>
<tr>
<td>Urgency levels</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orange</td>
<td>18</td>
<td>0.02 (0-0.10)</td>
<td>0.05 (0-0.10)</td>
<td>0.06 (0.02-0.10)*</td>
</tr>
<tr>
<td>Yellow</td>
<td>94</td>
<td>0.02 (0-0.20)</td>
<td>0.05 (0-0.35)</td>
<td>0.05 (0.02-0.14)*</td>
</tr>
<tr>
<td>Green</td>
<td>172</td>
<td>0.02 (0-0.15)</td>
<td>0.04 (0-0.35)</td>
<td>0.03 (0.01-0.16)*</td>
</tr>
<tr>
<td>Blue</td>
<td>7</td>
<td>0.03 (0.1-0.17)</td>
<td>0.05 (0.01-0.34)</td>
<td>0.03 (0.01-0.05)*</td>
</tr>
</tbody>
</table>

Data is presented as hours and minutes.* significance level p<0.05; † N differs for triage, because of missing time values; Arrival time: time from arrival to registration; Time to triage: time from registration to triage; Triage time: time start triage to end triage; Time to treatment: time from end triage to entrance of the treatment room; Waiting time: time from registration to entrance of the treatment room; Treatment time: time from entrance of the treatment room to discharge; Length of stay: time from registration to discharge

Table 4: Number of patients: waiting time and triage to treatment within maximum waiting time by urgency level

<table>
<thead>
<tr>
<th>Urgency level</th>
<th>Maximum waiting time †</th>
<th>Entrance time (min-max)</th>
<th>Time to triage (min-max)</th>
<th>Triage time (min-max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orange</td>
<td>10 minutes</td>
<td>0.02 (0-0.10)</td>
<td>0.05 (0-0.10)</td>
<td>0.06 (0.02-0.10)*</td>
</tr>
<tr>
<td>Yellow</td>
<td>60 minutes</td>
<td>0.02 (0-0.20)</td>
<td>0.05 (0-0.35)</td>
<td>0.05 (0.02-0.14)*</td>
</tr>
<tr>
<td>Green</td>
<td>120 minutes</td>
<td>0.02 (0-0.15)</td>
<td>0.04 (0-0.35)</td>
<td>0.03 (0.01-0.16)*</td>
</tr>
<tr>
<td>Blue</td>
<td>240 minutes</td>
<td>0.03 (0.1-0.17)</td>
<td>0.05 (0.01-0.34)</td>
<td>0.03 (0.01-0.05)*</td>
</tr>
</tbody>
</table>

* n differs, because of missing time values for end of triage time; † Maximum waiting time: following Manchester Triage System definitions 11
Waiting time: time from registration to entrance of the treatment room; Triage - treatment: time from end triage to entrance of the treatment room

Patient satisfaction

Based on the four service-related questions, in the Before Implementation group patient satisfaction concerning the provision of information and the opportunity to explain their problems was significantly higher. Satisfaction with length of waiting time was significantly higher in the After Implementation group as was the level of satisfaction about the way that their problem had been sorted out (Table 5).

No significant differences were found between triaged and non-triaged patients (Table 5). Also, no significant differences were found between urgency levels in the triaged patients, although overall, patients in the green level scored higher satisfaction rates on all questions than patients in the yellow level (Table 5).
Table 3: After implementation group: median waiting and throughput time, for triaged and non-triaged patients.

<table>
<thead>
<tr>
<th></th>
<th>Time to treatment (min-max)</th>
<th>Waiting time (min-max)</th>
<th>Treatment time (min-max)</th>
<th>Length of stay (min-max)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.07 (0.2.09)*</td>
<td>1.28 (0.10.26)*</td>
<td>1.48 (0.05-10.26)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.08 (0.2.39)</td>
<td>0.20 (0.3.30)*</td>
<td>1.08 (0.04-12.30)*</td>
<td>1.43 (0.12-12.55)</td>
</tr>
<tr>
<td></td>
<td>0.02 (0.0.07)*</td>
<td></td>
<td>0.12 (0.0.17)*</td>
<td>2.19 (0.15-3.47)*</td>
</tr>
<tr>
<td></td>
<td>0.08 (0.0.51)*</td>
<td>0.18 (0.1.51)*</td>
<td>2.00 (0.0.712.30)*</td>
<td>2.16 (0.12-12.55)*</td>
</tr>
<tr>
<td></td>
<td>0.11 (0.2.39)*</td>
<td>0.22 (0.3.30)*</td>
<td>0.51 (0.04-10.45)*</td>
<td>1.26 (0.14-10.55)*</td>
</tr>
<tr>
<td></td>
<td>0.11 (0.02-1.45)*</td>
<td>0.27 (0.09-1.47)*</td>
<td>0.38 (0.15-1.23)*</td>
<td>0.59 (0.40-1.57)*</td>
</tr>
</tbody>
</table>

Data is presented as hours and minutes.* significance level p<0.05; † N differs for triage, because of missing
arrival time: time from arrival to registration; time to triage: time from registration to triage; triage
time: time start triage to end triage; time to treatment: time from end triage to entrance of the
treatment room; waiting time: time from registration to entrance of the treatment room; treatment
time: time from entrance of the treatment room to discharge; length of stay: time from registration to
discharge.

Table 5: Patient satisfaction

<table>
<thead>
<tr>
<th></th>
<th>n Total</th>
<th>Spoken information</th>
<th>Opportunity to tell complaints</th>
<th>Time periods</th>
<th>Problem sorted out</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Implementation group</td>
<td>356</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfied</td>
<td>76.5%*</td>
<td>82.1%*</td>
<td>26.7%*</td>
<td>53.0%*</td>
<td></td>
</tr>
<tr>
<td>Fairly satisfied</td>
<td>22.1%</td>
<td>26.2%</td>
<td>57.0%</td>
<td>43.6%</td>
<td></td>
</tr>
<tr>
<td>Not satisfied</td>
<td>1.4%</td>
<td>1.7%</td>
<td>16.3%</td>
<td>3.4%</td>
<td></td>
</tr>
<tr>
<td>After Implementation group</td>
<td>286</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfied</td>
<td>61.7%*</td>
<td>74.6%*</td>
<td>36.6%*</td>
<td>71.2%*</td>
<td></td>
</tr>
<tr>
<td>Fairly satisfied</td>
<td>37.6%</td>
<td>23.6%</td>
<td>46.9%</td>
<td>24.4%</td>
<td></td>
</tr>
<tr>
<td>Not satisfied</td>
<td>0.7%</td>
<td>1.8%</td>
<td>16.5%</td>
<td>4.4%</td>
<td></td>
</tr>
<tr>
<td>After implementation group</td>
<td>188</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not triaged</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfied</td>
<td>64.2%</td>
<td>75.6%</td>
<td>37.1%</td>
<td>69.7%</td>
<td></td>
</tr>
<tr>
<td>Fairly satisfied</td>
<td>34.7%</td>
<td>22.2%</td>
<td>46.0%</td>
<td>25.2%</td>
<td></td>
</tr>
<tr>
<td>Not satisfied</td>
<td>1.1%</td>
<td>2.2%</td>
<td>16.9%</td>
<td>5.1%</td>
<td></td>
</tr>
<tr>
<td>Triaged</td>
<td>98</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfied</td>
<td>56.8%</td>
<td>72.8%</td>
<td>35.8%</td>
<td>74.2%</td>
<td></td>
</tr>
<tr>
<td>Fairly satisfied</td>
<td>43.2%</td>
<td>26.1%</td>
<td>48.4%</td>
<td>22.6%</td>
<td></td>
</tr>
<tr>
<td>Not satisfied</td>
<td>0%</td>
<td>1.1%</td>
<td>15.8%</td>
<td>3.2%</td>
<td></td>
</tr>
<tr>
<td>Urgency levels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Green</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfied</td>
<td>57.1%</td>
<td>73.2%</td>
<td>35.7%</td>
<td>73.2%</td>
<td></td>
</tr>
<tr>
<td>Fairly satisfied</td>
<td>42.9%</td>
<td>26.8%</td>
<td>48.2%</td>
<td>23.2%</td>
<td></td>
</tr>
<tr>
<td>Not satisfied</td>
<td>0%</td>
<td>0%</td>
<td>16.1%</td>
<td>3.6%</td>
<td></td>
</tr>
<tr>
<td>Yellow</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfied</td>
<td>48.3%</td>
<td>69.2%</td>
<td>20.7%</td>
<td>77.8%</td>
<td></td>
</tr>
<tr>
<td>Fairly satisfied</td>
<td>51.7%</td>
<td>30.8%</td>
<td>58.6%</td>
<td>18.5%</td>
<td></td>
</tr>
<tr>
<td>Not satisfied</td>
<td>0%</td>
<td>0%</td>
<td>20.7%</td>
<td>3.7%</td>
<td></td>
</tr>
</tbody>
</table>

Due to the low numbers in the other urgency levels, numbers are not presented; * Significance level p<0.05
DISCUSSION

After implementation of the MTS, waiting time was better distributed over urgency levels: patients in the highest level of urgency waited for a shorter time than patients in the lower levels. As expected, after implementation of the MTS, overall we found no important decrease in waiting time and LOS. Patient satisfaction was higher for waiting time and the feeling that their problem had been sorted out; but lower for the provision of information and the opportunity to explain their complaint. Also, we detected a positive trend in the satisfaction rates of low urgency patients. As this group is by far the largest group of patients attending EDs, this is relevant to modern ED practice. These findings underpin previous research on patients’ perceptions of waiting time and LOS, in which perception of waiting time was associated more with satisfaction than actual time did.\textsuperscript{6,15,17}

The finding that waiting time was longer for triaged patients than for not triaged patients is counter productive. The longer waiting time in triaged patients was obviously caused by the triage process itself, but triaging patients resulted in lower treatment time. This suggests that the information obtained from triage, results in a more efficient start of treatment. Furthermore, we quantified the triage time and showed that the higher the urgency level, the longer it took to triage. We speculate that the higher the urgency, it takes more time to rule ‘no urgency’ out.

The strength of this study is that we involved patients to provide additional timeline data, like time to triage as time being triaged. This novel approach to obtain not only routinely sampled data of patients was tested in the before measurement and shows to be accurate. More than 80\% of the differences range from -5 to 5 minutes between data provided by patients and routinely sampled data (data not shown).

The Institute of Medicine’s Committee on the Future of Emergency Care in the United States Health Systems recommends prioritizing improvement of hospital efficiency and patient flow.\textsuperscript{18} From our study it seems that just simply implementing a triage system will not impact this. This is in line with other reviews on effective interventions to improve patient flow.\textsuperscript{19,20} For example, recently two studies on improving EDs by applying the principles of Lean Thinking were published in which ED processes were changed and an improvement in efficiency was found, e.g. decreased waiting time and LOS.\textsuperscript{21,22} One of the key components of Lean Thinking is the added value to the client (in healthcare the patient). Overall patient satisfaction was measured in one study only, and was found to have improved.\textsuperscript{22} The principles of Lean Thinking were also applied at our ED in an advanced triage project, in which emergency nurses initiated diagnostic examinations from the triage. A small reduction in LOS for self-referred patients was achieved.\textsuperscript{23} Another recently published study describes the introduction of a new ED care method, in which triage is combined with involvement of a GP in the ED department. This increases patient satisfaction and led to a 13\% decrease in additional examinations.\textsuperscript{24} These studies suggest
that different architecture of the operating activities can improve the output of that process in terms of efficiency and satisfaction.

LIMITATIONS

Our study has several limitations. Firstly, by evaluating the implementation of the MTS we wanted to measure quality improvement. A weakness of this study is the use of a before and after test design and the relative long time interval between both measurements. An interrupted time-series design or a stepped wedge design would perhaps have been more appropriate and could be used in future research to develop robust evidence on improving the quality of EDs. Still, we were able to establish a reasonably similar control group.

Secondly, our study suffered from a mediocre response rate on patient satisfaction. This may have biased the results, but we detected no differences between responders and non-responders. Furthermore, our response rate is comparable with other studies performed at the ED. A diversity of survey instruments in assessing patient perceptions of hospital care exist, in which response rates are ranging from 17% to 92%. Therefore, we suppose that our results are still valid.

Thirdly, preferable a valid and reliable patient satisfaction questionnaire should be used, but at the performance of this study, no one existed. Therefore, we used the literature and previously used survey measurements to identify the most relevant questions. A Likert scale offers a range of choices from strongly positive to negative. For benchmark activities we used the same Likert scale as in the previous measurements. For study purposes, the Dutch answer options were translated back and forth in English by a native speaking physician. Generally patient responses are biased towards positive choices. To identify quality improvements aspects, we followed the same norm values as used for benchmarking in the core questionnaire for the assessment of patient satisfaction in academic hospitals in the Netherlands.

Fourthly, although we focused on and measured the four most relevant questions on patient satisfaction, a more detailed questionnaire would have given more insight into this aspect. In future research recently identified consumer expectations such as information on parking, how an ED works and how to identify staff could be incorporated as satisfaction outcomes. As a target group the self-referred and low urgency patients could be chosen as they are by far the largest groups attending EDs.

Fifthly, we used differing methods to distribute the patient satisfaction questionnaire, which may have influenced the results. Patients who fill in a questionnaire immediately after their visit, are often more positive than if they have time to criticize the effectiveness of their visit. Studies on this aspect show that satisfaction declines over time. Our results showed the same, overall the satisfaction scores were higher if the handout questionnaire was used, except for waiting time and the feeling that their problem had been sorted out. This suggests that satisfaction scores regarding waiting time after implementation of the MTS
are even higher than our results showed. To overcome the problem of different distribution methods and the time between the before and the after measurement, we performed sub analyses between triaged and non-triaged patients in the after measurement. Differences between these two groups show that, although the satisfied group is somewhat lower in the triaged group, the ‘not satisfied’ group is decreasing.

CONCLUSION

In conclusion, implementing MTS on its own is not sufficient to improve the efficiency and quality of EDs. More complex interventions including process redesigning that targets various groups of ED patients should be evaluated in the future by using rigorous research designs for quality improvement of EDs.
REFERENCES


The number of smokers needed to screen and treat in a smoking cessation programme

MN Storm-Versloot
H Vermeulen
LCW Wiggers
EMA Smets
JCJM de Haes
RJG Peters
DA Legemate
R de Vos

ABSTRACT

Objective: Smoking cessation is an important factor to reduce cardiovascular mortality, but considerable effort is needed to successfully persuade patients to quit smoking. We studied the efficiency of the Minimal Intervention Strategy (C-MIS) in addition to nicotine replacement therapy (NRT) for smoking cessation in cardiovascular outpatients in relation to the outcome of mortality.

Design: Prospective cohort data studying the C-MIS in three outpatient clinics: cardiology, vascular surgery and vascular medicine.

Methods: 2725 consecutive patients attending the clinics for first or routine follow-up visits were screened for atheroscleroses and smoking. The efficiency of the C-MIS was expressed as the number of smokers needed to screen (NNS) and needed to treat (NNT) in relation to the number of deaths prevented over a 5-year period. Mortality estimates were derived from the literature.

Results: 1431 patients were screened at first-time follow-up visits and 1294 at routine follow-up visits. With a rate of effectiveness of 4.3% for the C-MIS, the NNT was 240 (min-max: 64-∞) to prevent one death. The corresponding NNS was 687 (min-max: 141-∞) in the cardiology clinic, 574 (min-max: 134-∞) in the vascular surgery clinic and 444 (min-max: 90-∞) in the vascular medicine clinic. Within 5 years, 10 (min-max: 0-58) deaths could be prevented in all three clinics together. With the effectiveness for the C-MIS for first-time and routine follow-up attendees, only six (min-max: 0-36) and zero (min-max: 0-25) deaths could be prevented, respectively.

Conclusion: In terms of the efficiency of the C-MIS in addition to NRT, there is some benefit for first-time, and no benefit for routine follow-up attendees in preventing death.
INTRODUCTION
To reflect the effectiveness of an intervention from a clinical perspective, the number needed to treat (NNT) is generally used in clinical trials.\(^1\) NNT is defined as the number of patients that need to be treated to prevent one adverse event, and is the reciprocal of the absolute risk reduction.\(^2\) However, in screening programmes, the NNT may not reflect the total effort that is necessary to identify all patients who could potentially be treated. To quantify this effort, a new statistic has been developed: the number needed to screen (NNS).\(^3\) NNS is defined as the number of patients that need to be screened to prevent one adverse event. NNS can be one of the components used to evaluate and compare the usefulness, relevance, efficiency, cost-effectiveness and feasibility of screening programmes.\(^4\)

NNT and NNS are sensitive to baseline risk, patient characteristics, clinical setting\(^5\) and the effectiveness of an intervention. For example, as quitting smoking is important in preventing the recurrence of cardiovascular events, many smoking cessation programmes have been initiated. However, their limited effects on quit rates have resulted in moderate-to-high NNTs.\(^6\)\(^-\)\(^11\) It is necessary to realize that the NNT in these programmes represents the number of patients that quit smoking as a result of the intervention, and not the number of adverse events or deaths that were prevented. To prevent one adverse event or death, the number of patients that need to participate in the programme is much higher (Appendix 1A). Moreover, a considerable number of patients that were approached were not interested in participating in a smoking cessation programme. These elements comprise the NNS (Appendix 1B). However, repeated screening for smoking behaviour and motivating smokers to participate in the programme impose more time constraints on healthcare providers.

There is an increasing need to quantify the effort that health care providers need to make to engage patients in attempting to quit smoking in relation to clinically relevant outcomes. Using data from a recently published smoking cessation intervention trial in cardiovascular outpatients,\(^11\) we studied the efficiency of the additional effect of the intervention, expressed as NNT, NNS and number of deaths prevented over a 5-year period. In other words, we investigated the number of patients that has to be screened and treated to prevent one death as a result of a minimal smoking cessation intervention.

METHODS
To calculate the NNT and NNS to engage patients in a smoking cessation programme, we used mortality rates from the literature\(^12\) and data from the trial mentioned earlier.\(^11\) This trial studied the effectiveness of a minimal contact behavioural counselling smoking cessation intervention (the Minimal Intervention Strategy for Cardiology patients (C-MIS: Appendix 2)), in addition to nicotine replacement therapy among patients who attended the cardiology, vascular surgery and vascular medicine (department where patients are...
assessed for integral vascular risk management) outpatient clinics for treatment of symptomatic atherosclerotic cardiovascular disease (CVD) at the Academic Medical Center in Amsterdam, The Netherlands. As there could be a difference in the prevalence of atherosclerotic disease and smokers, which might influence the NNS, data were studied separately for the three different outpatient clinics.

All patients aged over 17 years attending one of these outpatient clinics were eligible. From September 2001 to May 2002 all eligible first-time attendees of the clinics (e.g. the first 3 visits; Appendix 2) were recruited, until the target number of patients was reached (consecutive series of 500 cardiac, 500 vascular surgery and, for arbitrary logistical reasons, 250 vascular medicine patients). On the basis of the assumption that patients recently diagnosed with a smoking-related illness would be more motivated to participate in a smoking cessation programme than patients attending the outpatient clinic for a routine follow-up visit, we included an equivalent number of patients on a routine follow-up visit (e.g. more than 3 visits to the same clinic; Appendix 2). These patients were recruited from September 2002 to February 2003.

Patients were excluded if they did not attend their appointments. To compensate for these exclusions, more patients were recruited than the 2500 mentioned above ((500+500+250) * 2) patients. The patients were divided into six subgroups: three outpatient clinics stratified by two types of visit (first time and routine follow-up).

Outpatient clinic attendance for first-time or routine follow-up visits was identified by means of the Hospital Information System. The names of the patients identified were passed on to the clinicians, who were invited to verify the type of visit (first or routine), the presence of symptomatic atherosclerotic CVD (yes, no), the smoking status (yes, no), and the willingness to quit (yes, no). Missing data were completed after cross-checking with patients, clinicians and medical records.

The original trial was approved by the medical ethics committee of the hospital. The patients included in this trial were asked for written informed consent. To calculate the NNT and NNS, we used mortality rates derived from the literature and the cohort data from the above-mentioned trial. For the purpose of this study, the lowest bound of the effectiveness of the C-MIS was set at 0.01% (point estimate 4.3% (95% confidence interval (CI): -3.2 to 11.7%), under the hypothesis that the C-MIS would not increase the number of smokers.

To extend the scope of the NNS and deaths prevented to a 5-year period, the number of all attending patients was retrieved from clinic admission data contained in the hospital’s annual report for 2003. At the cardiology outpatient clinic approximately 6700 patients (9500 consultations) were seen annually, of whom 2075 (22%) were first-time attendees; at the vascular surgery clinic of 1975 patients (3000 consultations), 700 (24%) were first-time attendees; at the vascular medicine clinic of 1225 patients (2150 consultations), 575 (26%) were first-time attendees.
Analysis

Patient characteristics (i.e., symptomatic atherosclerotic CVD and smoking status) were described separately for each subgroup. Patients with symptomatic atherosclerotic CVD were furthermore described by age, sex, smoking status and willingness to quit smoking.

To calculate the NNT to prevent one death, using the C-MIS programme, we divided 1 by the multiplication of the calculated risk difference of dying while still smoking or dying after having stopped smoking, by the estimated effectiveness of the C-MIS (Appendix 1A). These numbers were derived from the literature.\textsuperscript{12} The effectiveness of the C-MIS with regard to quitting smoking was based on the results of the original trial\textsuperscript{11} and additional analysis from our dataset.

To subsequently calculate the NNS, we divided the NNT by the percentage of patients willing to quit smoking (Appendix 1B), derived from the cohort data of the original trial.

To quantify the total effort required from all three outpatient clinics to prevent deaths over a 5-year period, the number of patients per year visiting the outpatient clinics was multiplied by 5, by the percentage of patients with symptomatic atherosclerotic CVD, and by the percentage of patients who smoked. The result was subsequently divided by the NNS (Appendix 1C).

Sensitivity analysis

To examine the variability of NNS, we performed a sensitivity analysis for each of the estimates of the variables mentioned (risk of dying, patient characteristics, and the effectiveness of the intervention). The sensitivity analysis was performed using the analysis of extremes.\textsuperscript{13} Simultaneously each variable was given the most optimistic or pessimistic value as derived from the 95% CIs. This analysis was performed separately for each subgroup.

RESULTS

Patient characteristics, prevalence of symptomatic atherosclerotic CVD, and smoking

A total of 2725 patients visited the three outpatient clinics during recruitment (Figure 1). Of these, 1431 patients were first-time attendees and 1294 routine follow-up attendees. Patients attending the cardiology and vascular surgery outpatient clinics were older (mean age 59.8 and 61.6, respectively) than patients attending the vascular medicine outpatient clinic (mean age 48.0).

The prevalence of symptomatic atherosclerotic CVD was significantly higher in patients attending the vascular surgery outpatient clinic. This was also true for routine follow-up attendees on comparison with first-time attendees (Figure 1).

The prevalence of smoking was significantly higher in patients at the vascular surgery outpatient clinic than at the other clinics, and this is also applicable to first-time attendees compared to routine follow-up attendees (Figure 1).
**Figure 1:** Flowchart of patients in the total cohort

CVD, Cardiovascular disease. First visit: visit 1- visit 3; Routine follow-up visit: four or more visits. *n differs from N because of missing values. *p < 0.05, significant difference between the outpatient clinics. †p < 0.05, significant difference between the two types of visits.

**Patients with symptomatic atherosclerotic CVD: characteristics, smoking prevalence and willingness to quit**

Patients with symptomatic atherosclerotic CVD attending the cardiology and vascular surgery outpatient clinic were, on average, 13 and 15 years older than those attending the vascular medicine outpatient clinic (Table 1). The number of smokers was highest among those attending the vascular surgery outpatient clinic (37.9%).

The percentages of patients who smoked and patients willing to quit smoking were lower in all three outpatient clinics in the routine follow-up attendees, but only reached statistical significance in the vascular surgery outpatient clinic (Table 1).
Effectiveness of C-MIS, and impact on NNT and NNS

With an overall estimated effectiveness of 4.3% (95% CI: -3.2 to 11.7) for the C-MIS,\textsuperscript{11} and a risk difference of dying while still smoking or after quitting smoking of 9.8% (95% CI: 6.4%-13.4%),\textsuperscript{12} the NNT with the C-MIS is 240 patients who smoke with symptomatic atherosclerotic CVD to prevent one death. The most optimistic calculation shows the NNT to be 64. In the most pessimistic calculation, the NNT reaches infinity. The NNS for the cardiology, vascular surgery and vascular medicine outpatient clinics is 687, 574 and 444, respectively. In the most optimistic calculation, the NNS is 141, 134 and 90, respectively; in the most pessimistic calculation, these numbers reach infinity (Table 2).

For first-time attendees, the estimated effectiveness of the C-MIS is 6.5% (95% CI: -3.3% to 16.3%)\textsuperscript{11} and the NNT is 158. For the cardiology, vascular surgery and vascular medicine outpatient clinics, the NNS is 399, 312 and 237, respectively (Table 2).

For routine follow-up attendees, the estimated effectiveness for the C-MIS is -0.9% (95% CI: -10.1% to 8.4%),\textsuperscript{11} and the NNT and NNS reach infinity for all outpatient clinics (Table 2).

The yield of the C-MIS over a period of 5 years

Using the overall estimated effectiveness of the C-MIS over a 5-year period, a small number of deaths can be prevented: 4.5 in the cardiology, 4.5 in the vascular surgery and 0.9 in the vascular medicine outpatient clinic (10 in total). In the most optimistic calculation, these numbers are 28.3, 24.0 and 6.2, respectively (58 in total). In the most pessimistic calculation, no deaths can be prevented (Table 2).

---

### Table 1. Characteristics of patients with symptomatic atherosclerotic CVD in outpatient clinic

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Patients (N)</th>
<th>Age, mean (SD) (years)</th>
<th>Male, N (%)</th>
<th>Smoking, N (%)</th>
<th>Willing to quit, N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>496</td>
<td>64.2 (12.9)*</td>
<td>337 (67.9)</td>
<td>100 (20.2)*</td>
<td>35 (35.0)</td>
</tr>
<tr>
<td>first visit</td>
<td>230</td>
<td>62.1 (13.8)†</td>
<td>141 (61.3)†</td>
<td>53 (23.0)</td>
<td>21 (39.6)</td>
</tr>
<tr>
<td>routine follow-up visit</td>
<td>266</td>
<td>65.9 (11.9)†</td>
<td>196 (73.7)†</td>
<td>47 (17.7)</td>
<td>14 (29.8)</td>
</tr>
<tr>
<td>Vascular Surgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>749</td>
<td>66.1 (11.6)*</td>
<td>492 (65.7)</td>
<td>284 (37.9)*</td>
<td>119 (41.9)</td>
</tr>
<tr>
<td>first visit</td>
<td>357</td>
<td>65.8 (11.8)†</td>
<td>230 (64.4)</td>
<td>156 (43.7)†</td>
<td>79 (50.6)†</td>
</tr>
<tr>
<td>routine follow-up visit</td>
<td>392</td>
<td>66.4 (11.4)†</td>
<td>262 (66.8)</td>
<td>128 (32.7)†</td>
<td>40 (31.3)†</td>
</tr>
<tr>
<td>Vascular Medicine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>144</td>
<td>51.1 (9.3)*</td>
<td>105 (72.9)</td>
<td>37 (25.7)*</td>
<td>20 (54.1)</td>
</tr>
<tr>
<td>first visit</td>
<td>67</td>
<td>50.3 (10.4)</td>
<td>52 (77.6)</td>
<td>18 (26.9)</td>
<td>12 (66.7)</td>
</tr>
<tr>
<td>routine follow-up visit</td>
<td>77</td>
<td>51.8 (8.4)</td>
<td>53 (68.8)</td>
<td>19 (24.7)</td>
<td>8 (42.1)</td>
</tr>
</tbody>
</table>

CVD, cardiovascular disease. *p < 0.05 significant difference between the outpatient clinics. †p < 0.05 significant difference between the two types of visits.
When using the above-mentioned specific estimated effectiveness of the C-MIS for first-time and routine follow-up attendees, 6.3 (e.g. 2.6, 3.0, 0.7) and no deaths were prevented, respectively (Table 2).

### Table 2. NNS of smoking patients with symptomatic atherosclerotic CVD to prevent one death when quit smoking by a C-MIS programme, and the total yield of the programme in 5 years in terms of prevented deaths

<table>
<thead>
<tr>
<th></th>
<th>Patients visits per year (N)</th>
<th>Symptomatic atherosclerotic CVD % (95% CI)</th>
<th>Smoking % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cardiology</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total(^a)</td>
<td>6700</td>
<td>46.1 (43.1 - 49.1)</td>
<td>20.2 (16.6 - 23.7)</td>
</tr>
<tr>
<td>first visit(^b)</td>
<td>2075</td>
<td>42.7 (38.6 - 46.9)</td>
<td>23.0 (17.6 - 28.5)</td>
</tr>
<tr>
<td>routine follow-up visit(^c)</td>
<td>4625</td>
<td>49.4 (45.2 - 53.7)</td>
<td>17.7 (13.1 - 22.3)</td>
</tr>
<tr>
<td><strong>Vascular Surgery</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total(^a)</td>
<td>1975</td>
<td>68.7 (66.0 - 71.5)</td>
<td>37.9 (34.4 - 41.4)</td>
</tr>
<tr>
<td>first visit(^b)</td>
<td>700</td>
<td>60.4 (56.5 - 64.3)</td>
<td>43.7 (38.6 - 48.8)</td>
</tr>
<tr>
<td>routine follow-up visit(^c)</td>
<td>1275</td>
<td>78.6 (70.7 - 78.4)</td>
<td>32.7 (28.0 - 37.3)</td>
</tr>
<tr>
<td><strong>Vascular Medicine</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total(^a)</td>
<td>1225</td>
<td>25.8 (22.1 - 29.4)</td>
<td>25.7 (18.6 - 32.8)</td>
</tr>
<tr>
<td>first visit(^b)</td>
<td>575</td>
<td>22.2 (17.5 - 26.9)</td>
<td>26.9 (16.8 - 39.1)</td>
</tr>
<tr>
<td>routine follow-up visit(^c)</td>
<td>650</td>
<td>30.0 (24.4 - 35.6)</td>
<td>24.7 (15.6 - 35.8)</td>
</tr>
</tbody>
</table>

CI, Confidence Interval; C-MIS, Minimal Intervention Strategy for Cardiology patients; CVD, cardiovascular disease; NNS, number needed to screen; NNT, number needed to treat; RRR, relative risk reduction. Justification of used estimates and calculations. The calculations were based on the following assumptions:

1. number of patients visits per year: hospital annual report 2003
2. mean percentage of symptomatic atherosclerotic CVD for the cardiology, vascular surgery and vascular medicine outpatient clinics (Figure 1)
3. mean percentage of smokers and willingness to quit smoking for the cardiology, vascular surgery and vascular medicine outpatient clinics (Table 1)
   - \(^a\) overall effectiveness of the C-MIS: 4.3% (95% CI: -3.2% to 11.7%)\(^1\)
   - \(^b\) effectiveness of the C-MIS for first-time attendees: 6.5% (95% CI: -3.3% to 16.3%)\(^1\); additional analysis
   - \(^c\) effectiveness of the C-MIS for routine follow-up attendees: -0.9% (95% CI: -10.1% to 8.4%)\(^1\); additional analysis
   - \(^d\) using the effectiveness of the C-MIS of 0.01% as lowest bound
4. absolute risk of dying when smoking in coronary heart disease population: 27.1% (95% CI: 22% to 32%)\(^2\)
5. RRR for deaths by abstained smoking and smoking: 36% (95% CI: 29.0% to 42.0%)\(^2\)
6. calculated risk difference (absolute risk multiplied with RRR) for deaths when abstained smoking and smoking: 9.8% (95% CI: 6.4% to 13.4%)\(^2\); To calculate the NNT and NNS and deaths prevented, see appendix 1A, 1B and 1C. NNT of smoking patients with symptomatic atherosclerotic CVD: 240 (upper and lower bound: 64-156 740)\(^d\). NNT of smoking patients with symptomatic atherosclerotic CVD for the first visit: 158 (upper and lower bound: 46-156 740)\(^d\). NNT of smoking patients with symptomatic atherosclerotic CVD for the routine follow-up visit: 102 383 (upper and lower bound: 86-156 740)\(^d\)

When using the above-mentioned specific estimated effectiveness of the C-MIS for first-time and routine follow-up attendees, 6.3 (e.g. 2.6, 3.0, 0.7) and no deaths were prevented, respectively (Table 2).
DISCUSSION

When compared with other screening programmes, fewer patients in our study needed to be screened and treated to reduce smoking-related mortality. However, the NNS was influenced heavily by a range of effectiveness estimates of the C-MIS. Ideally, the NNT and NNS should be interpreted in the context of a randomized clinical trial, as in our study. The NNS is one of the criteria used to determine the efficiency of a screening programme. In our study, the impact of the C-MIS in terms of number of deaths prevented over a period of 5 years, varied widely (0-58). In the most pessimistic calculation, no deaths could be prevented, although it is reasonable to assume that the programme will prevent six or more deaths given the actual effectiveness of the C-MIS for the two types of visit (Table 2).

The strength of our study is that no assumptions were made about patient and cohort characteristics such as smoking status, willingness to quit smoking, prevalence of symptomatic atherosclerotic CVD, or the effectiveness of the C-MIS. The actual numbers were all obtained from the original trial. Our only assumption was that the risk of dying was independent of the underlying type of cardiovascular disease. This assumption follows the Joint British recommendations on the prevention of coronary heart disease, which state that patients with peripheral arterial disease and those suffering from coronary arterial disease should be managed similarly, as the aetiology and the major risk factors of these conditions are similar.

To interpret the relative magnitude of the calculated NNS, we compared our result with the NNS in other studies of different diseases. Of the screening strategies evaluated...
by Rembold, the largest clinical benefit was found for dyslipidaemia, if detection was followed by pravastatin treatment for 5 years. To prevent one death, the estimated NNS was 418 (95% CI: 235 to 79720). The NNS to prevent one death in Dutch general practice screening programmes is approximately 2560 for cervical cancer and approximately 2340 for the detection of hypertension in patients aged 55-75 years. This comparison suggests that the C-MIS for smoking cessation, assessed by the NNS and NNT, may be quite effective. Comparing only the NNS with other screening programmes leads to an incomplete assessment. Apart from the number of deaths prevented, other outcomes such as morbidity, balance of advantages and disadvantages, and cost-effectiveness need to be taken into account. In addition, smoking is a chronic ‘condition’ that requires repeated screening rather than a single approach.

Not only should the number of deaths prevented be counted as the benefit of an intervention, but also the prevention of morbidity. The literature shows a 32% relative risk reduction of nonfatal myocardial infarctions in cardiovascular patients who quit smoking. When repeating the same calculations for morbidity, as we did for the risk of death, a similar effect size is derived, doubling the overall effect of the intervention. Furthermore, the benefits of the intervention may increase over time. As follow-up of patients in the included trials ranged from 1 to 26 years, it is unlikely that our NNS will be influenced heavily when the timeframe is expanded. However, the effect size is likely influenced by the age of the participants.

When balancing advantages and disadvantages of the programme, screening for smoking behaviour and applying the C-MIS is not likely to have any disadvantages for the patient. Clinical practice guidelines on smoking cessation recommend that clinicians should identify, document and treat every smoker at each visit. However, in the Netherlands, smoking behaviour is only recorded in approximately 67% of the medical records of smokers. Furthermore, only 10% of Dutch physicians support the patient in their efforts to quit smoking. Lack of time, lack of incentive and lack of effective smoking cessation programmes in outpatient clinics are associated with these practices. Although the C-MIS is a structured and feasible programme and can be delivered by trained nurses, the clinician must still continue to verify and discuss the smoking habits of patients to motivate them to quit or to prevent relapses. In our original trial, we created optimal circumstances by identifying all patients who smoked, offering a personalized quit-smoking advice, delivering the C-MIS with the aid of trained nurses, and offering free NRT. These components may have increased patients’ willingness to quit smoking, resulting in an overly favourable NNT and NNS. Furthermore, we calculated the NNT and NNS as a single intervention for the additional effects of the C-MIS, which might also have resulted in optimistic estimates. Nevertheless, the calculations provide insight into the efforts clinicians have to make in screening and treating patients who smoke. However, comparisons with other screening programmes should be interpreted cautiously.
Finally, we did not calculate real costs, but helping smokers to quit is ranked among the top three most cost effective preventive services that clinicians can offer (asymptomatic) patients. Yet, we have insufficient knowledge about the cost-effectiveness in our population of older patients with smoking-related disease. The older the patient, the less likely it is that he or she is willing to quit, as older patients question the advantages of smoking cessation in view of their limited remaining life span. Although nurse-delivered interventions were found to be effective, especially in hospitalized cardiovascular patients, patients assigned to nurse-led clinics for secondary prevention of coronary heart disease did not show any improvement for smoking status at 1 year. Furthermore, evidence of effective cessation strategies in the cardiovascular (out)patient population is very limited. We showed that patients (even those who are older) with newly diagnosed cardiovascular disease are more receptive and willing to quit than routine follow-up attendees. However, the C-MIS was ineffective in the latter group.

On the basis of our findings, the implementation of the C-MIS in outpatient clinics may be questioned. In our centre, we decided not to proceed with the C-MIS after the trial.
REFERENCES

1. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. BMJ 1995; 310:452-4


APPENDIX 1A

Formula used for calculating the NNT of smoking patients with symptomatic atherosclerotic CVD by using the C-MIS.

\[
\text{NNT of patients who smoke with symptomatic atherosclerotic CVD} = \frac{1}{(\text{risk difference between dying while still smoking and dying after having quit smoking}) \times (\text{estimated effectiveness C-MIS})}
\]

Appendix 1B

Formula used for calculating the NNS of smoking patients with symptomatic atherosclerotic CVD by using the C-MIS.

\[
\text{NNS of patients who smoke with symptomatic atherosclerotic CVD} = \frac{\text{NNT of patients who smoke with symptomatic atherosclerotic CVD}}{\% \text{ willing to quit smoking}}
\]

Appendix 1C

Formula used for calculating the number of deaths prevented over 5 years in a cardiovascular outpatient clinic.

\[
\text{No. of deaths prevented per 5 year} = \frac{(\text{No. of patients visiting the outpatient clinic per year}) \times (5) \times (\% \text{ symptomatic atherosclerotic CVD}) \times (\% \text{ smoking})}{\text{NNS of patients who smoke with symptomatic atherosclerotic CVD}}
\]
Appendix 2: Definitions used in this study

C-MIS: A minimal contact behavioural smoking cessation intervention: Minimal Intervention Strategy for Cardiology patients. The C-MIS consists of 6 steps and is delivered by nurses, after a advice to quit smoking by the clinician.

Clinician

Nurse

1. advice to quit smoking

1 assess smoke profile
2 assess and enhance motivation
3 discuss barriers
4 set a quit date
5 provide self-help materials
6 plan after-care by telephone

2. pay attention to smoking behaviour during each visit at the outpatient clinic

Symptomatic atherosclerotic CVD: Clinically diagnosed occlusive or aneurysmal disease of coronary or peripheral arteries, or a strong suggestion that this was the cause of the patient’s symptoms.

Smoking: Patients who smoke one or more cigarettes or cigars a day at the time of assessment.

Willing to quit: Number of patients who actually participate in the trial, and patients who want to quit on their own.

First visit: Three or less visits (or patients with a newly diagnosed cardiovascular event) to the outpatient clinic cardiology, vascular surgery or vascular medicine of the Academic Medical Center in Amsterdam.

Routine follow-up visit: After the third visit (four or more visits) to the same outpatient clinic, the patient was addressed as routine follow-up patient.

C-MIS, Minimal Intervention Strategy for Cardiology patients; CVD, cardiovascular disease.
Long-term adherence to a local guideline on postoperative body temperature measurement: mixed methods analysis

MN Storm-Versloot
AM Knops
DT Ubbink
A Goossens
DA Legemate
H Vermeulen

ABSTRACT

Aim: To find out whether a successful multifaceted implementation approach of a local evidence-based guideline on postoperative body temperature measurements (BTM) was persistent over time, and which factors influenced long-term adherence.

Methods: Mixed methods analysis. Patient records were retrospectively examined to measure guideline adherence. Data on influencing factors were collected in focus group meetings for nurses and a plenary meeting with an interactive questionnaire for doctors.

Results: Records from 102 surgical patients were studied, totalling 1226 BTM. According to the guideline, an indication for BTM was present in 55% (679/1226). Actually, BTM were taken in 60% (736/1226), of which 55% (403/736) was in accordance with the guideline. The overall adherence rate to the guideline was 50% (617/1226). Belief in the advantages of the guideline and strong staff support appeared to facilitate long-term adherence. Barriers were; the controversial nature of the guideline, the lack of self-efficacy among nurses and doctors as to clinical judgement to identify an infection when refraining from BTM, and a lack of management and staff doctor support. Furthermore, newly appointed nurses and doctors were trained to measure BTM during their initial medical or nursing education, which was in contradiction with the guideline.

Conclusions: A multifaceted implementation strategy is not sufficient to maintain long-term adherence. To ensure long-term adherence, especially of controversial guidelines, adherence should be monitored and reported regularly over time. Strong staff support and leadership on all wards is crucial to maintain awareness. Medical and nursing curricula should include the pros and cons of taking BTM, combined with enhancing self-efficacy.
INTRODUCTION

Clinical guidelines are considered to improve the quality of care and close the research evidence-practice gap.\(^1,2\) Non-adherence to guidelines may lead to unnecessary diagnostics, suboptimal treatment, or even adverse events.\(^3,4\) Even though, it is still a challenge to implement evidence-based guidelines and their recommendations into clinical practice and achieve high adherence rates.\(^5-7\)

A large body of literature is available on the effective implementation of clinical guidelines, but the “magic bullet” does not exist.\(^8,9\) Passive implementation strategies, such as the presentation and publication of a guideline, will hardly change professional behaviour.\(^10\) To optimize adaptation, it has been proposed that guidelines should be simple and supported by active and multifaceted implementation strategies.\(^11-13\)

A multifaceted strategy includes presentations, personal visits, feedback, reminders, and letters.\(^14\) Such a strategy was used in order to break through unnecessary postoperative routine body temperature measurements (BTM), after the development of an evidence-based guideline in our hospital [Appendix 1].\(^15,16\) In short, the results of the studies showed that routine BTM could - and should - be abandoned. Serious postoperative infections were found to occur without an accompanying increase in temperature, while fever was frequently not accompanied by an infection. The strategy for the (short-term) implementation of this guideline was tailored to the facilitators and barriers, as indicated by nurses and doctors. The main barriers were self-efficacy and disagreement with the nature of the guideline as routine BTM is still considered as good patient care.

The multifaceted approach resulted in a 91% adherence to the guideline, six months after its release [Appendix 2].\(^17\) Thereafter, the responsibility for the use of the guideline was transferred to the staff surgeon and the head nurse on each ward involved. Although short-term guideline adherence may be achieved,\(^18,19\) regression to old habits is a common human flaw.\(^20\) The last was expressed by several nurses and staff surgeons on our wards. They observed signs of decline in adherence to the guideline and this was intercepted by the study group.

Subsequently, a study was performed on the facilitators and barriers to long-term adherence to this local guideline. The aim was to find out whether a successful multifaceted implementation approach is persistent over time, and which factors were influencing long-term adherence to evidence-based guidelines.

METHODS

Setting

This mixed-methods study with retrospective and prospective, as well as qualitative and quantitative, components was performed at the Department of Surgery of a 1000-bed university hospital in the Netherlands. The Department of Surgery consists of nursing
wards for short stay, traumatology, vascular surgery, and general surgery, which includes gastro-enterology, hepatic, pancreatic, and biliary surgery patients. Because this study merely monitored daily practice, approval by the local medical ethics review board was not needed.

**Guideline**

The guideline was approved by the medical and nursing management teams of the surgical departments in 2000. The guideline consists of the BTM recommendation and states that body temperature needs not to be measured postoperatively for surgical cases of non-infectious diseases [Appendix 1]. Indications for BTM were the presence of an infectious disease preoperatively, or patients contracted an infection postoperatively, or blood transfusion.

**Methods of evaluation**

*Assessment of long-term guideline adherence*

Quantitative data on long-term adherence to the guideline was retrospectively determined, as any prospective determination is likely to influence the behaviour of nurses and, thus, bias the results. From all elective and emergency surgical procedures performed during a three-month period in 2007, 25 records from each surgical ward for evaluation were selected at random. The three-month period was the same time frame as used in the initial guideline adherence measurement.

Two investigators (MNSV, AMK) independently examined patient records for BTM and their indications in accordance with the guideline. Four outcomes were possible: BTM could have been taken appropriately or inappropriately, or they could have been omitted appropriately or inappropriately. Disagreement about indications, which occurred in 16 patients (16%), were referred to a third investigator (DTU) and resolved by discussion.

Adherence rate was defined as the number of appropriately measured and omitted BTM relative to the total number of BTM. The total number of BTM was calculated following the definitions used in the diagnostic study, and the subsequent determination of the initial adherence rate. In these studies BTM were recorded twice a day until discharge or until a maximum of 14 days postoperatively.

**Focus group meetings with nurses**

For registered and student nurses from surgical wards, structured focus group meetings were organized to explore facilitators and barriers influencing long-term guideline adherence. Nurses could attend the meetings only once. The aim of the meeting was explained, that data would be processed anonymously, and that answers were not judged as being good or bad. Furthermore, it was emphasized that discussion on the topic was important.

Nurses were asked first, whether they were aware of the guideline and how they had learnt about it; second, whether they adhered to the guideline; third, if they had experienced staff support (staff surgeon or head nurse) regarding guideline adherence; and fourth, all
facilitators and barriers experienced by them regarding guideline adherence were listed. All nurses wrote down their appreciation of positive or negative reasons pertinent to guideline adherence. These reasons were plenary discussed and clustered into several positive or negative themes until the group was satisfied with the result. Any new themes that came up during the discussion were added. One investigator led the discussion (AMK) and another took field notes (MNSV). After each session, the impressions of both investigators were added to the field notes. Both investigators independently categorized the clustered themes under one of the six main factors that influences implementation according to Grol: 1) belief in or the change itself; 2) those who have to work with the changes; 3) social context; 4) organizational aspects; 5) economic, political, and legal aspects; and 6) aspects of the implementation strategy chosen. The last factor was irrelevant and not used, because the guideline was not implemented at this point in time. Disagreement was discussed with a third investigator (AG) and a final categorization was made by consensus. The focus group meetings were continued until saturation was reached: in the last two sessions no additional themes or insights were discovered.

**Plenary meeting with doctors**
A structured plenary meeting was organized with the staff surgeons, surgical residents and medical students from all surgical wards and they were confronted with an interactive questionnaire. The questions were derived from the nurses’ focus group results and addressed adherence to the guideline itself and whether the facilitators and barriers mentioned by the nurses agreed with the doctors’ views. In addition, details on position and specialization were asked.

**Data analysis**
Guideline adherence was expressed as a percentage. In the analysis of long-term adherence, a distinction was made between wards with (i.e. management enforces compliance with the guideline) and without ‘staff support’—considering that this is an obvious factor influencing guideline adherence. The plenary meeting with doctors was performed with the aid of an interactive audience response system (TurningPoint, Reivo Ltd, Theale, UK) and imported into SPSS v.14 (SPSS Inc., Chicago, Illinois, USA). Also, descriptive statistics of the data obtained during the plenary meeting and the guideline adherence rates were performed using SPSS.

**RESULTS**

**Characteristics of respondents and awareness of the guideline**
Forty-seven nurses from four surgical wards attended the focus group meetings. After seven meetings no additional information or insights had appeared. Nurses’ functions are shown in Table 1. Thirty-four out of 47 (72%) nurses reported to be aware of the guideline, mainly because they had participated in the original study 7 years earlier (59%) or because
they had been informed by colleagues (38%). Nurses from two out of the four wards stated they were not aware of whether or not the staff had decided to apply the guideline. Hardly any (15%; 3/20) of these nurses claimed to have followed the guideline in contrast with the two wards where nurses experienced staff support (95%; 21/22) (Table 1).

During the plenary meeting, 42 doctors responded to the questionnaire. Doctors’ functions are presented in Table 1. Specializations and functions were representative for the total surgical staff. In general, 33% (14/42) of the doctors were aware of the guideline. Most of them (93%; 13/14) were staff surgeons.

**Assessment of actual long-term guideline adherence**

In total 102 patient records were reviewed. Baseline patient characteristics are given in Table 2. In these patients, a maximum of 1226 BTM could have been taken if they had been measured following the definitions used in the determination of the initial adherence rate (twice a day up until discharge or up to 14 days postoperatively). According to the guideline 679 out of 1226 (55%) BTM should have been taken. Actual BTM were taken in 736 out of 1226 (60%) of which 403 out of 736 (55%) was in accordance with the guideline. The overall adherence rate was 617 out of 1226 (50%) (Table 3) and did not differ between wards with and without staff support. However, the percentages of inappropriately BTM were much lower for the wards with staff support (110/277: 40%) than for those without (223/270: 83%).

<table>
<thead>
<tr>
<th>Table 1: Respondents’ functions and nurses’ self-reported guideline adherence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nurses’ functions</strong></td>
</tr>
<tr>
<td>Registered nurses</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Student nurses</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td><strong>Self-reported adherence</strong></td>
</tr>
<tr>
<td>Always</td>
</tr>
<tr>
<td>Sometimes</td>
</tr>
<tr>
<td>Never</td>
</tr>
<tr>
<td>Missing</td>
</tr>
</tbody>
</table>

**Doctors’ functions**

| Staff surgeons | 9 | 7 | 16 (38) |
| Surgical residents | 5 | 9 | 14 (33) |
| Medical students | 3 | 9 | 12 (29) |
| Total          | 17 (40%) | 25 (60%) | 42 |

Chapter 9
Factors influencing guideline adherence according to the nurses

The information retrieved from the focus group meetings are summarized below for each factor separately. Table 4 presents the most important facilitators and barriers.

1. **The belief in or the change itself**: as to abandoning routine postoperative BTM.

   Especially at the wards with staff support, almost all nurses stated that the main advantage of the guideline was the reliance on clinical judgement rather than on the routinely sampled, and possibly misleading body temperature. This advantage was mentioned particularly when a patient had clinical signs of an infection, without an increased body temperature.

   Other nurses believed many patients on their wards did not meet the guideline criteria, because these patients were at high risk of developing an infection.

---

**Table 2: Patient (record) characteristics and numbers of BTM**

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient records</td>
<td>102</td>
<td></td>
</tr>
<tr>
<td>Number of postoperative days (median, interquartile range)</td>
<td>6 (3-14)</td>
<td></td>
</tr>
<tr>
<td>Type of surgery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GE/HPB</td>
<td>34</td>
<td>33</td>
</tr>
<tr>
<td>Vascular surgery</td>
<td>29</td>
<td>28</td>
</tr>
<tr>
<td>Trauma surgery</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>General surgery/ other</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Preoperative infection</td>
<td>31</td>
<td>30</td>
</tr>
<tr>
<td>Postoperative infection</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Maximum possible number of BTM in all patients*</td>
<td>1226</td>
<td></td>
</tr>
<tr>
<td>Maximum possible number of BTM performed on indication**</td>
<td>679</td>
<td>55</td>
</tr>
</tbody>
</table>

* Based on definitions used in the diagnostic study\(^16\): a maximum of two recorded possible measurements per day until discharge or until a maximum of 14 days postoperatively

** Measurements to be taken according to the guideline: patients with preoperative infection, or patients contracted an infection postoperatively, or patients receiving blood transfusion

GE/HPB, gastroenterology/ hepatic, pancreatic, and biliary surgery; BTM, body temperature measurements.

**Table 3: BTM in all patients taken or not taken in accordance with the guideline**

<table>
<thead>
<tr>
<th>BTM according to guideline</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actually BTM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>403</td>
<td>333</td>
<td>736</td>
</tr>
<tr>
<td>No</td>
<td>276</td>
<td>214</td>
<td>490</td>
</tr>
<tr>
<td>Total</td>
<td>679</td>
<td>547</td>
<td>1226</td>
</tr>
</tbody>
</table>

\(^{A}\) BTM appropriately taken (percentage = A/[A+C] * 100%)

\(^{B}\) BTM inappropriately taken (percentage = B/[B+D] * 100%)

\(^{C}\) BTM inappropriately not taken (percentage = C/[A+C] * 100%)

\(^{D}\) BTM appropriately not taken (percentage = D/[B+D] * 100%)

\(^{A+D}\) Adherence rate: BTM appropriately taken plus the BTM appropriately not taken (percentage = [A+D]/[A+B+C+D] * 100%)

BTM, body temperature measurements.

Factors influencing guideline adherence according to the nurses

The information retrieved from the focus group meetings are summarized below for each factor separately. Table 4 presents the most important facilitators and barriers.

1. **The belief in or the change itself**: as to abandoning routine postoperative BTM.

   Especially at the wards with staff support, almost all nurses stated that the main advantage of the guideline was the reliance on clinical judgement rather than on the routinely sampled, and possibly misleading body temperature. This advantage was mentioned particularly when a patient had clinical signs of an infection, without an increased body temperature.

   Other nurses believed many patients on their wards did not meet the guideline criteria, because these patients were at high risk of developing an infection.
2. **Those who have to work with the change:**

At the wards without staff support, the majority of the nurses stated clinical judgement is subjective and varies as it depends on the knowledge, experience, and expertise of nurses and doctors. Therefore, routine BTM were obtained in order not to miss any infection. Moreover, some of these nurses stated, when an indication to BTM is present, it can wrongfully be forgotten, which might be harmful for the patient.

3. **Social context: Attitude and expectations**

Most nurses stated it would be easier to adhere to the guideline if doctors did so in the first place, but new doctors were usually unaware of the guideline and often requested for the BTM. Nurses needed to exonerate themselves from not taking the temperature. Furthermore, some nurses took BTM routinely and others only for the indications as described in the guideline. This lack of uniformity made it more difficult for nurses to discuss the value of BTM with patients.

---

**Table 4: Statements of nurses regarding facilitators and barriers for guideline utilization**

<table>
<thead>
<tr>
<th>Facilitators</th>
<th>Barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Innovation itself: the believe in or the change itself</strong></td>
<td><strong>To take measurements may be a useless ritual, but it is harmless, easy to do, and not time-consuming</strong></td>
</tr>
<tr>
<td>Trust in the scientific basis of the guideline</td>
<td><strong>To prevent superfluous diagnostic tests</strong></td>
</tr>
<tr>
<td>To prevent superfluous diagnostic tests</td>
<td><strong>To stimulate the development of clinical judgement</strong></td>
</tr>
<tr>
<td>To stimulate the development of clinical judgement</td>
<td><strong>To take measurements may be a useless ritual, but it is harmless, easy to do, and not time-consuming</strong></td>
</tr>
<tr>
<td><strong>Individual professional: those who have to work with the change</strong></td>
<td><strong>To take measurements may be a useless ritual, but it is harmless, easy to do, and not time-consuming</strong></td>
</tr>
<tr>
<td>Confidence in personal clinical judgement for identifying a postoperative infection.</td>
<td><strong>To take measurements may be a useless ritual, but it is harmless, easy to do, and not time-consuming</strong></td>
</tr>
<tr>
<td><strong>Social context: attitude and expectations</strong></td>
<td><strong>To prevent superfluous diagnostic tests</strong></td>
</tr>
<tr>
<td>Staff support in using the guideline</td>
<td><strong>To stimulate the development of clinical judgement</strong></td>
</tr>
<tr>
<td>Nurses correct each other when inappropriate measurements were taken.</td>
<td><strong>To take measurements may be a useless ritual, but it is harmless, easy to do, and not time-consuming</strong></td>
</tr>
<tr>
<td><strong>Organizational context</strong></td>
<td><strong>To take measurements may be a useless ritual, but it is harmless, easy to do, and not time-consuming</strong></td>
</tr>
<tr>
<td>Standard procedures on wards were given to newly appointed nurses, including the content of the guideline</td>
<td><strong>To take measurements may be a useless ritual, but it is harmless, easy to do, and not time-consuming</strong></td>
</tr>
<tr>
<td>It is always hard to find a thermometer on the ward</td>
<td><strong>To take measurements may be a useless ritual, but it is harmless, easy to do, and not time-consuming</strong></td>
</tr>
<tr>
<td><strong>Economic, political, and legal context</strong></td>
<td><strong>To take measurements may be a useless ritual, but it is harmless, easy to do, and not time-consuming</strong></td>
</tr>
<tr>
<td>None mentioned</td>
<td><strong>To take measurements may be a useless ritual, but it is harmless, easy to do, and not time-consuming</strong></td>
</tr>
</tbody>
</table>

BTM, body temperature measurements.
4. Organizational aspects

Patient population: Obviously, during the period of 7 years following the initial guideline implementation, organizational changes in departments and wards had occurred. Some wards accommodated other specialties than surgery as well, e.g. plastic surgery and urology. Therefore, specialties that had not participated in the primary study were not familiar with the guideline and doctors still requested BTM.

Staff support: Almost all nurses on two wards were poorly aware whether or not the staff had decided to apply the guideline. In contrast, new nurses on the two wards with staff support were told that routine postoperative BTM were not considered a routine practice to perform, but only in specific situations.

5. Economic, political, and legal aspects

Routine BTM were still regarded as an esteemed ritual in nursing, even in the education of student nurses and doctors.

Factors influencing guideline adherence according to the doctors

Among the doctors who were aware of the guideline, more claimed to adhere to the guideline on the wards with staff support (5/6) than on those wards without staff support (2/8). About half of the doctors (20/42: 48%) routinely asked for a patient’s temperature, especially if they were unaware of the guideline. The majority (21/30: 70%) of staff surgeons and surgical residents did not want to know the temperature on a daily basis, but did if patients had clinical signs of a postoperative infection. In contrast, medical students did want to know body temperatures routinely (7/12: 58%). Especially on the ‘no staff support’ wards, some doctors claimed that patients did not match the guideline because they were at high risk for postoperative infections.

<table>
<thead>
<tr>
<th>Table 5: Statements of doctors, divided by guideline awareness</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aware of guideline</strong></td>
</tr>
<tr>
<td>Adopting the guideline</td>
</tr>
<tr>
<td>Trusting the evidence</td>
</tr>
<tr>
<td>Asking routinely for the temperature</td>
</tr>
<tr>
<td>Almost never asking for the temperature</td>
</tr>
<tr>
<td>Want to know the temperature on a daily basis</td>
</tr>
<tr>
<td>Want to know the temperature when patients have clinical signs of an infection</td>
</tr>
<tr>
<td>Patient group fits into the guideline</td>
</tr>
<tr>
<td>Belief many patients are at risk of postoperative infection</td>
</tr>
<tr>
<td>Trust their own clinical judgement</td>
</tr>
<tr>
<td>Trust the nurses’ clinical judgement</td>
</tr>
<tr>
<td>NA, not applicable</td>
</tr>
</tbody>
</table>

Long-term guideline adherence
The amount of confidence in personal clinical judgement for identifying postoperative infections without BTM was highest for the staff surgeons, and lowest for medical students. Staff surgeons and medical students had more confidence in the nurses’ clinical judgement for identifying a postoperative infection than surgical residents did. Table 5 summarizes the doctors’ responses to the statements with a distinction for awareness of the guideline.

**DISCUSSION**

This study shows that an initially applied successful multifaceted implementation strategy does not guarantee long-term adherence to an evidence-based guideline. The overall adherence rate had decreased to 50%, after the initial adherence rate of 91%. Facilitators for long-term adherence were: a belief in the advantages of the guideline, and staff support. Barriers were, first, the controversial nature of the guideline with faded awareness, second, a lack of self-efficacy: e.g. distrust in personal clinical judgement to identify an infection when refraining from BTM, and third, for nurses a lack of management and staff doctor support.

Same facilitators and barriers have been reported in studies on short-term guideline adherence.\textsuperscript{22-24} One of the few studies focusing on long-term adherence showed that sustained adherence to guidelines is more likely if they contain clear recommendations, are evidence-based, are non-controversial, and do not demand a change in existing routines.\textsuperscript{22} The guideline studied here did not meet the latter two criteria, which may explain the declined adherence rate over time.

As to the controversial nature and changing existing routines, the guideline recommended only taking BTM when infection was diagnosed. This was a titanic change in daily practice, which was achieved initially, but not sustained over time on all wards. In general, student nurses as well as residents learned that routine BTM is considered as good patient care and these conceptions are still incorporated in medical and nursing curricula. Therefore, only on the wards with strong staff support and leadership, newly appointed nurses were notified with information on the content of the guideline and awareness and familiarity was sustained. Although these nurses were informed, newly appointed surgical residents were not. As a consequence, awareness in especially the surgical residents was fading.

As to self-efficacy, this study showed that even the majority of staff surgeons frequently asked for BTM when patients had clinical signs of an infection. This misinterpretation on the content of the guideline led to discussion among doctors and nurses about the validity of clinical judgement when to take BTM. Different misconceptions on BTM among graduates have also been described in literature.\textsuperscript{25} In general, nurses and doctors still consider BTM to be essential to support their clinical judgement and confirm clinical signs of infection\textsuperscript{26} and although the majority of staff surgeons were confident to identify an infection without BTM, the surgical residents were not. A lack of self-efficacy in doctors has been described
as a main barrier to guideline adherence and can be present in up to 65%. In addition, residents distrusted not only their own, but also the nurses’ clinical judgement. Distrust between doctors and nurses is depending on and related to doctors’ perceptions of the individual nurses’ competence. Nurses also expressed a lack of self-efficacy, but in contrast they stated that stimulating the development of clinical judgement was an advantage of the guideline and should therefore be supported and taught to student nurses.

As to staff support, nurses’ need for support from management and doctors was very important. A lack of staff support hampered long-term adherence. The same has also been described for successful implementation of guidelines in nursing practices.

This study is somewhat extraordinary as a mix of quantitative and qualitative, but also retrospective and prospective methods were used. Mixed-methods are increasingly used in implementation and quality improvement research. However, some limitations of this study should be mentioned. Adherence was measured only once and after a long time. Maybe an interrupted time series analysis would have given a better overview of the changes in adherence rate to the guideline and yielded earlier signs of non-adherence. Nowadays, the incorporation of long-term boosters through the development of quality indicators is recommended to monitor and report the change of behaviour. At that time we were not aware of that and, as a 91% adherence was reached, it was believed that the responsibility for endured adherence could be transferred to the staff surgeon and the head nurse on each ward. Furthermore, reasons for the lack in self-efficacy were not explored. This should be studied more thoroughly in future implementation studies.

Our findings indicated that an active multifaceted implementation strategy can be effective for deploying a guideline, but it is not sufficient to root practice change and to accomplish long-term adherence. A strong self-efficacy and staff support seemed necessary to maintain awareness and sustained adherence. Therefore, some recommendations may be useful for long-term adherence in guidelines. First, in especially controversial guidelines, adherence rate and barriers have to be monitored and reported over time. Enforcement through leadership is needed by health care providers at every professional level, and indicators to monitor the required changes. When adherence rate tends to decline, any barriers should be explored again and accordingly addressed with tailored interventions. Second, staff support on all wards is necessary to maintain awareness. Third, it is necessary to influence and build the self-efficacy of nurses and doctors. Fourth, guideline recommendations should be incorporated in medical and nursing curricula. This all together should enable to ensure long-term adherence of guidelines.
REFERENCES


Appendix 1

**Summary: development of the guideline ‘postoperative temperature measurements’ (2000)**

**Problem:** Widespread traditions in clinical practice associated with questionable effectiveness are still present. One such tradition is the routine use of body temperature measurements (BTM) in postoperative patients (at least twice a day until discharge) in order to detect or exclude an infection at an early stage. In mostly retrospective and unblinded studies it has been suggested that early postoperative fever was of little use in clinical practice.\(^{15}\) If fever is present, doctors are more likely to order additional diagnostic tests, but if the temperature is normal, they are less likely to consider the probability of an infection. This daily routine translates into countless BTM and medical actions per year. In the era of evidence based practice it has become necessary to rationalize these traditions.

**Aim of the study:** To prospectively assess the diagnostic accuracy of routine postoperative BTM by comparing them with the presence or absence of postoperative infection in a general surgical population.

**Results:** From 2282 BTM taken from 284 patients, a temperature $\geq 38^\circ$C appeared to have the following diagnostic parameters: Sensitivity: 37% (95% Confidence Interval [CI]: 16-62%); Specificity 80% (95% CI: 75-85%); Negative Predictive Value: 90% (95% CI: 89-91%); Positive Predictive Value: 8% (95% CI: 5-13%).

**Conclusions:** BTM are a poor indicator for detecting an infection; a patient may have an infection without fever and vice versa.

**Guideline content:** Routine postoperative BTM are abolished. BTM are indicated only if: patients underwent surgery because of an infection to monitor its progression; patients contracted an infection postoperatively; patients received a blood transfusion.
Appendix 2

**Summary: implementation of the guideline (1999-2001)**

**Introduction:** Before the start of the diagnostic study ‘postoperative BTM’ we realized that the acceptance of omitting routine BTM could be low in nurses and doctors. Nurses and doctors need to learn to rely more on their clinical judgement than relying on the BTM. Therefore, baseline facilitators and barriers were assessed by means of a questionnaire and by formal and informal meetings. The main barriers were: self-efficacy and agreement.

**Baseline description:** 50% of the doctors wanted routine BTM versus 8% of the nurses. For BTM on indication this was respectively 43% and 69%. At the start of the study, about 30% of doctors and nurses were afraid of missing an infection or detecting infections later if they had to rely solely on their clinical judgement. Approximately 75% of doctors and nurses expected that patients would feel unsure when BTM were not carried out.

**Multifaceted strategy:** Strategies used were based on these barriers: involvement of doctors and nurses in the prospective study, daily visits to the departments for providing information and feedback, regular information letters and presentations, study recognition, posters, reminders, patient information letters, a short guideline description with recommendation in the (electronic) protocols, and evaluation of the implementation strategy.

**Methods:** Evaluation of the implementation strategy was prospectively monitored. For a period of 3 months, guideline adherence was measured by chart review, where reasons for BTM were stated. A maximum of two BTM per day up until discharge or until 14 days postoperative were recorded. A post-questionnaire measured the self-reported adherence of doctors and nurses as well as their motives to be adherent.

**Results:** Adherence: The guideline was completely followed in 50% (37/75) patients, and partly in the other patients. BTM were taken in 14% (157/1086) of which 33% (55/157) was in accordance with the guideline. In total 91% (984/1086) BTM were in accordance with the guideline.

**Reasons for non-adherence:** Clinical complaints and symptoms of a postoperative infection. **Motives for adherence:** Motives of the 93% followers (doctors and nurses) of the guideline were based on scientific proof (59%), convincing arguments (37%), and always wanting to measure on indication (30%).
Do stakeholders in wound care prefer evidence-based wound care products? 
A survey in the Netherlands

AM Eskes
MN Storm-Versloot
H Vermeulen
DT Ubbink
ABSTRACT

**Introduction:** For several wound products compelling evidence is available on their effectiveness, for example, from systematic reviews. The process of buying, prescribing and applying wound materials involve many stakeholders, who may not be aware of this evidence, although this is essential for uniform and optimum treatment choice.

**Methods:** In this survey, we determined the general awareness and use of evidence, based on (Cochrane) systematic reviews, for wound products in open wounds and burns among wound care stakeholders, including doctors, nurses, buyers, pharmacologists and manufacturers. We included 262 stakeholders.

**Results:** Doctors preferred conventional antiseptics (e.g. iodine), while specialized nurses and manufacturers favoured popular products (e.g. silver). Most stakeholders considered silver-containing products as evidence-based effective antiseptics. These were mostly used by specialized nurses ($47/57; 82\%$), although only few of them ($9/55; 16\%$) thought using silver is evidence-based. For burns, silver sulfadiazine and hydrofibre were most popular. The majority of professionals considered using silver sulfadiazine to be evidence-based, which contradicts scientific results. Awareness and use of the Cochrane Library was lower among nurses than among doctors ($p<0.001$). Two thirds of the manufacturers were unaware of, or never used, the Cochrane Library.

**Conclusion:** Available compelling evidence in wound care is not equally internalized by stakeholders, which is required to ensure evidence-based decision making.
INTRODUCTION

To date, health care professionals are expected to keep abreast of current professional knowledge, and to apply research evidence in their daily practice in order to deliver the highest possible quality of care. Ideally, the evidence-based practice (EBP) paradigm promotes evidence-based decision making with patients in clinical practice, preferably derived from proper (Cochrane) systematic reviews, if any, or well-performed clinical trials. Reality, however, shows that 30-40% of patients receive care that is not in accordance with available high quality research evidence, while another 20-30% of patients receive care that is even contraindicated. Furthermore, the existence of guidelines does not guarantee its actual application. Guideline recommendations are followed in on average 67% of the treatment decisions made. From these figures, it is obvious that the EBP paradigm has not yet been adopted by all health care professionals in their daily practice.

In wound care, an additional phenomenon is apparent. The available therapeutic options to choose from may be influenced by many different health care stakeholders (i.e. doctors, nurses, manufacturers, buyers, pharmacists), as well as by the patients’ preferences. Although the experiential knowledge from all these stakeholders is a necessity, it is not a sufficient basis for clinical decision making. Hence, odds are high that the eventual treatment given is not evidence-based if one or more of these stakeholders do not make an evidence-based decision or are not aware of available high quality evidence. In other words, a joint venture is needed to make evidence-based wound treatment work.

Evidence-based wound care could also be seen as a challenge. Because of the lack of high quality research evidence or evidence-based guidelines to help choose the most appropriate form of local wound care and thus challenging evidence-based decision making. However, for some indications in wound care high quality evidence is available, which does make evidence-based treatment decisions possible. Recently, well-performed systematic reviews with recommendations as to the use of antiseptics for preventing and treating wound infections have been produced and disseminated at various (inter)national conferences and among different audiences. These reviews, for example, present high quality evidence about the effectiveness of honey dressings, iodine and silver sulfadiazine for specific wounds. For many years, iodine has been dissuaded because of its purported adverse effects, but was recently shown to be at least as effective as other antiseptics without serious harmful effects, such as a delay in wound healing, particularly in chronic and burn wounds. In contrast, silver sulfadiazine, although still the treatment of choice in burn wounds, was found not to counteract infections more than other antiseptic agents, while decelerating wound healing in patients with partial-thickness burns.

Because of the growing body of high quality evidence on the (in)effectiveness of certain (antiseptic) wound dressings or agents, it seems unethical to administer ineffective treatments or to withhold patients from the best available evidence-based treatments.
Therefore, the aim of this study was to explore the general awareness and use of compelling research evidence, based on available (Cochrane) systematic reviews, among several groups of wound care stakeholders in health care.

METHODS

Study setting
From April to September 2010, we contacted 31 Dutch medical centres, including all university (n = 8) and burn centres (n = 3), 13 home care institutions, 100 primary care facilities and 12 manufacturers of wound care products to take part in this study.

Participants
A representative, broad range of different health care professionals involved in wound care was recruited. These involved surgeons, plastic surgeons, dermatologists, general practitioners (GPs), surgical nurses, home care nurses, specialized wound care nurses, as well as manufacturers of wound care products, totalling eight professional groups. We aimed for about 25 professionals per group to obtain a full scale of possible answers. We included twice as many clinical and wound specialist nurses as they are key performers in daily wound care. Furthermore, interviews were planned with the heads of the hospital’s buyer and pharmacy departments.

Questionnaire
To assess awareness and use in daily practice of the available, high quality research evidence on antiseptics and wound care products by wound care stakeholders in the Netherlands, a short questionnaire was designed. It consisted of five questions, each relating to personal preference and awareness of evidence from (Cochrane) systematic reviews:

1. When considering an antiseptic dressing for an open wound, what would be your top 3 of wound care products?
2. For which antiseptic wound care products has the effectiveness been established, based on high-level evidence, for the treatment of open wounds?
3. What would be your top 3 wound care products for the local treatment of open partial-thickness burns?
4. For which wound care products has the effectiveness been established, based on high quality research evidence, for the treatment of open partial-thickness burns?
5. How often do you consult the Cochrane Library?

A list of various wound care products was given to choose from (Table 1 and Table 2). This list was based on common usage in daily practice and available evidence from (Cochrane) systematic reviews. Participants also had the opportunity to choose ‘other’. Furthermore, we noted the stakeholders’ age and function.
Collection of responses
We used different methods to collect the responses. First, during a national meeting of specialized wound care nurses and during presentations of Plastic Surgeons and Dermatologists we used electronic voting devices: ResponseCard™ keypads linked to TurningPoint for Microsoft® PowerPoint® (Turning technologies, Ohio, USA, version: 4.1.0.9020). No discussion was allowed. Second, we contacted medical centres and home care institutions by telephone. Third, we distributed 35 CDROMs with the questions to doctors and nurses of 12 Dutch medical centres. Fourth, some hospitals and home care centres requested we should attach the questionnaire to a cover letter addressed to the local doctors, nurses, and manufacturers, which was distributed via email. Health care professionals could only respond once to the questions.

Data analysis
Data were entered into PASW statistics 18.0 (SPSS Inc., Chicago, IL, USA). Statistical analyses included descriptive statistics of the top 3 choice of wound care products used, and awareness of evidence. The relation between age and the use of the Cochrane Library among the different groups were compared using the Kruskal-Wallis test, with P < 0.05 considered significant. The same test was used to detect differences between the professional groups. The Mann-Whitney U test was used to compare the awareness and use of the Cochrane Library among doctors and nurses.

We used the chi-square ($X^2$) statistic to whether the different methods of investigation (by email, phone, or plenary presentation) influenced the results.

RESULTS
A total of 262 professionals were included in this survey; 96 doctors, 143 nurses and 23 manufacturers of wound care products. The age distribution of the professionals is shown in Table 3. The age distributions are representative of the Dutch healthcare professional situation.

Completeness of data
In total, 126 out of 262 (48%) questionnaires were filled in completely, without any missing data. The main reason for missing data was a lack of experience with treatment of patients with burn wounds or not supplying three preferences. After comparing the complete questionnaires with the results of the incomplete ones, no substantial differences were found in the top 3. Therefore, we present the results of all answers given. No significant differences were found between a shorter (telephone or electronic voting devices) or longer (email or CDROM) answering time.
### Table 1: Choices of antiseptic products in the treatment of open wounds and awareness of evidence

<table>
<thead>
<tr>
<th>Professionals</th>
<th>Total (N)*</th>
<th>Chlorhexidine (%)</th>
<th>Eusol® (%)</th>
<th>Fucidin® (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Surgeons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usage</td>
<td>24</td>
<td>13 (54)</td>
<td>7 (29)</td>
<td>6 (25)</td>
</tr>
<tr>
<td>Evidence**</td>
<td>22</td>
<td>14 (64)</td>
<td>6 (27)</td>
<td>3 (14)</td>
</tr>
<tr>
<td><strong>Plastic surgeons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usage</td>
<td>25</td>
<td>8 (32)</td>
<td>13 (52)</td>
<td>14 (56)</td>
</tr>
<tr>
<td>Evidence**</td>
<td>25</td>
<td>12 (48)</td>
<td>8 (32)</td>
<td>11 (44)</td>
</tr>
<tr>
<td><strong>Dermatologists</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usage</td>
<td>37</td>
<td>15 (41)</td>
<td>11 (30)</td>
<td>19 (51)</td>
</tr>
<tr>
<td>Evidence**</td>
<td>37</td>
<td>17 (46)</td>
<td>7 (19)</td>
<td>15 (41)</td>
</tr>
<tr>
<td><strong>General practitioners</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usage</td>
<td>8</td>
<td>3 (38)</td>
<td>0 (0)</td>
<td>6 (75)</td>
</tr>
<tr>
<td>Evidence**</td>
<td>7</td>
<td>2 (29)</td>
<td>1 (14)</td>
<td>4 (57)</td>
</tr>
<tr>
<td><strong>Specialized nurses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usage</td>
<td>57</td>
<td>6 (11)</td>
<td>14 (25)</td>
<td>11 (19)</td>
</tr>
<tr>
<td>Evidence**</td>
<td>55</td>
<td>2 (4)</td>
<td>11 (20)</td>
<td>8 (15)</td>
</tr>
<tr>
<td><strong>Surgical nurses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usage</td>
<td>65</td>
<td>27 (42)</td>
<td>25 (38)</td>
<td>13 (20)</td>
</tr>
<tr>
<td>Evidence**</td>
<td>51</td>
<td>26 (51)</td>
<td>14 (27)</td>
<td>14 (27)</td>
</tr>
<tr>
<td><strong>Home care nurses</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usage</td>
<td>17</td>
<td>5 (31)</td>
<td>6 (38)</td>
<td>7 (44)</td>
</tr>
<tr>
<td>Evidence**</td>
<td>15</td>
<td>5 (33)</td>
<td>6 (40)</td>
<td>4 (27)</td>
</tr>
<tr>
<td><strong>Manufacturers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usage</td>
<td>23</td>
<td>2 (9)</td>
<td>6 (26)</td>
<td>4 (17)</td>
</tr>
<tr>
<td>Evidence**</td>
<td>21</td>
<td>1 (5)</td>
<td>1 (5)</td>
<td>1 (5)</td>
</tr>
</tbody>
</table>

NA, not applicable. *N is total number of participants who answered the question. **Evidence is number of respondents who have the opinion that the effectiveness of a particular product is evidence-based.

### Table 2: Choices of wound care products in the treatment of open partial-thickness burns and awareness of evidence.

<table>
<thead>
<tr>
<th>Professionals</th>
<th>Item</th>
<th>Total (N)*</th>
<th>Chlorhexidine (%)</th>
<th>Film (%)</th>
<th>Hydrofibre (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Surgeons</strong></td>
<td>Usage</td>
<td>23</td>
<td>0 (0)</td>
<td>1 (4)</td>
<td>11 (48)</td>
</tr>
<tr>
<td></td>
<td>Evidence**</td>
<td>23</td>
<td>1 (4)</td>
<td>1 (4)</td>
<td>10 (43)</td>
</tr>
<tr>
<td><strong>Plastic surgeons</strong></td>
<td>Usage</td>
<td>25</td>
<td>1 (4)</td>
<td>6 (24)</td>
<td>13 (52)</td>
</tr>
<tr>
<td></td>
<td>Evidence**</td>
<td>24</td>
<td>2 (8)</td>
<td>5 (21)</td>
<td>10 (42)</td>
</tr>
<tr>
<td><strong>Dermatologists</strong></td>
<td>Usage</td>
<td>34</td>
<td>2 (6)</td>
<td>6 (18)</td>
<td>11 (32)</td>
</tr>
<tr>
<td></td>
<td>Evidence**</td>
<td>33</td>
<td>3 (9)</td>
<td>4 (12)</td>
<td>5 (15)</td>
</tr>
<tr>
<td><strong>General practitioners</strong></td>
<td>Usage</td>
<td>6</td>
<td>1 (17)</td>
<td>0 (0)</td>
<td>2 (33)</td>
</tr>
<tr>
<td></td>
<td>Evidence**</td>
<td>7</td>
<td>2 (29)</td>
<td>0 (0)</td>
<td>2 (29)</td>
</tr>
<tr>
<td><strong>Specialized nurses</strong></td>
<td>Usage</td>
<td>56</td>
<td>2 (4)</td>
<td>4 (7)</td>
<td>49 (88)</td>
</tr>
<tr>
<td></td>
<td>Evidence**</td>
<td>57</td>
<td>3 (5)</td>
<td>1 (2)</td>
<td>36 (63)</td>
</tr>
<tr>
<td><strong>Surgical nurses</strong></td>
<td>Usage</td>
<td>50</td>
<td>6 (12)</td>
<td>12 (24)</td>
<td>23 (46)</td>
</tr>
<tr>
<td></td>
<td>Evidence**</td>
<td>41</td>
<td>9 (22)</td>
<td>12 (29)</td>
<td>19 (46)</td>
</tr>
<tr>
<td><strong>Home care nurses</strong></td>
<td>Usage</td>
<td>16</td>
<td>1 (7)</td>
<td>2 (13)</td>
<td>6 (40)</td>
</tr>
<tr>
<td></td>
<td>Evidence**</td>
<td>10</td>
<td>2 (20)</td>
<td>1 (10)</td>
<td>2 (20)</td>
</tr>
<tr>
<td><strong>Manufacturers</strong></td>
<td>Usage</td>
<td>21</td>
<td>0 (0)</td>
<td>4 (19)</td>
<td>16 (76)</td>
</tr>
<tr>
<td></td>
<td>Evidence**</td>
<td>19</td>
<td>1 (5)</td>
<td>3 (16)</td>
<td>11 (58)</td>
</tr>
</tbody>
</table>

NA, not applicable. *N is total number of participants who answered the question. **Evidence is number of respondents who have the opinion that the effectiveness of a particular product is evidence-based.
<table>
<thead>
<tr>
<th>Professionals</th>
<th>Item</th>
<th>Total (N)*</th>
<th>Chlorhexidine (%)</th>
<th>Honey (%)</th>
<th>Iodine (%)</th>
<th>Silver product (%)</th>
<th>Silver -containing dressing (%)</th>
<th>Other (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeons</td>
<td>Usage</td>
<td>23</td>
<td>0 (0)</td>
<td>2 (9)</td>
<td>11 (48)</td>
<td>20 (87)</td>
<td>2 (9)</td>
<td>4 (17)</td>
</tr>
<tr>
<td></td>
<td>Evidence**</td>
<td>23</td>
<td>3 (13)</td>
<td>2 (9)</td>
<td>15 (60)</td>
<td>20 (80)</td>
<td>4 (16)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Plastic surgeons</td>
<td>Usage</td>
<td>25</td>
<td>1 (4)</td>
<td>0 (0)</td>
<td>6 (25)</td>
<td>18 (75)</td>
<td>6 (25)</td>
<td>11 (32)</td>
</tr>
<tr>
<td>Dermatologists</td>
<td>Usage</td>
<td>34</td>
<td>2 (6)</td>
<td>4 (17)</td>
<td>21 (62)</td>
<td>15 (44)</td>
<td>14 (41)</td>
<td>11 (32)</td>
</tr>
<tr>
<td>General practitioners</td>
<td>Usage</td>
<td>6</td>
<td>3 (9)</td>
<td>4 (12)</td>
<td>8 (24)</td>
<td>8 (24)</td>
<td>7 (21)</td>
<td>NA</td>
</tr>
<tr>
<td>Specialized nurses</td>
<td>Usage</td>
<td>56</td>
<td>0 (0)</td>
<td>2 (33)</td>
<td>3 (43)</td>
<td>3 (43)</td>
<td>1 (14)</td>
<td>NA</td>
</tr>
<tr>
<td>Surgical nurses</td>
<td>Usage</td>
<td>50</td>
<td>2 (29)</td>
<td>2 (29)</td>
<td>2 (29)</td>
<td>2 (29)</td>
<td>1 (14)</td>
<td>NA</td>
</tr>
<tr>
<td>Home care nurses</td>
<td>Usage</td>
<td>16</td>
<td>3 (5)</td>
<td>1 (2)</td>
<td>31 (55)</td>
<td>40 (71)</td>
<td>16 (29)</td>
<td>0 (16)</td>
</tr>
<tr>
<td>Manufacturers</td>
<td>Usage</td>
<td>21</td>
<td>5 (10)</td>
<td>3 (6)</td>
<td>37 (74)</td>
<td>27 (54)</td>
<td>12 (24)</td>
<td>5 (10)</td>
</tr>
<tr>
<td></td>
<td>Evidence**</td>
<td>19</td>
<td>10 (24)</td>
<td>12 (29)</td>
<td>17 (41)</td>
<td>21 (51)</td>
<td>19 (46)</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3 (20)</td>
<td>1 (7)</td>
<td>12 (80)</td>
<td>9 (60)</td>
<td>4 (27)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 (20)</td>
<td>1 (10)</td>
<td>7 (70)</td>
<td>7 (70)</td>
<td>3 (30)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 (10)</td>
<td>0 (0)</td>
<td>5 (24)</td>
<td>8 (38)</td>
<td>10 (48)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2 (11)</td>
<td>1 (5)</td>
<td>1 (5)</td>
<td>3 (16)</td>
<td>8 (42)</td>
</tr>
</tbody>
</table>

** NA, not applicable. N is total number of participants who answered the question. **Evidence is number of respondents who have the opinion that the effectiveness of a particular product is evidence-based.
Table 3: Age distribution of professionals (in years)

<table>
<thead>
<tr>
<th></th>
<th>&lt;31</th>
<th>31-40</th>
<th>41-50</th>
<th>51-60</th>
<th>&gt;61</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeons (n=24)</td>
<td>5</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Plastic surgeons (n=25)</td>
<td>13</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Dermatologists (n=39)</td>
<td>21</td>
<td>9</td>
<td>4</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>General practitioners (n=8)</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Specialized nurses (n=57)</td>
<td>3</td>
<td>15</td>
<td>25</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Nurses (n=69)</td>
<td>38</td>
<td>8</td>
<td>9</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>Home care nurses (n=17)</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Manufacturers (n=23)</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>

Products for open wounds
The first choice of antiseptics used for the treatment of open wounds as given by the professional groups is shown in Table 4. The first choice within the groups of specialized nurses and manufactures for the local treatment of open wounds was silver-containing products. Doctors were more inclined to use conventional antiseptic products (e.g. Eusol®, chlorhexidine and iodine) than nurses, who more often chose popular antiseptic agents (e.g. silver products). Usage and awareness among the respondents of available evidence for each antiseptic wound care product is given in Table 1. In the majority of the professional groups (6 out of the 8 groups), over 50% of the respondents held the opinion that the effectiveness of silver-containing products is evidence-based. This opinion was also found in half of the groups for iodine and in 2/8 groups for chlorhexidine. In contrast, neither Eusol® nor Furacin® was considered supported by evidence. Notably, silver-containing products were mostly used by specialized nurses (47/57; 82%), but only few (9/55; 16%) stated to be aware of any evidence about the effectiveness of silver for open wounds. Remarkably, the number of doctors who used honey was lower than the number of doctors who stated to be aware of the evidence of its effectiveness. The opposite was observed in specialized wound care nurses and manufacturers.

Table 4: Stakeholders’ first choice of antiseptic products in the treatment of open wounds

<table>
<thead>
<tr>
<th>Professional Group</th>
<th>First Choice of Antiseptic Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeons</td>
<td>Chlorhexidine (42%)</td>
</tr>
<tr>
<td>Plastic surgeons</td>
<td>Eusol (28%)</td>
</tr>
<tr>
<td>Dermatologists</td>
<td>Iodine (65%)</td>
</tr>
<tr>
<td>General practitioners</td>
<td>Iodine (38%)</td>
</tr>
<tr>
<td>Specialized nurses</td>
<td>Silver-containing products (32%)</td>
</tr>
<tr>
<td>Surgical nurses</td>
<td>Chlorhexidine (20%)</td>
</tr>
<tr>
<td>Home care nurses</td>
<td>Iodine (59%)</td>
</tr>
<tr>
<td>Manufacturers</td>
<td>Silver-containing products (48%)</td>
</tr>
</tbody>
</table>
Products for burn wounds
The majority of professionals reported silver sulfadiazine, as their first choice for the local treatment of open partial-thickness burns (Table 5). In half of the groups (4/8), over 50% of the respondents answered that the effectiveness of silver sulfadiazine is evidence-based. This was also the case in 2/8 groups for hydrofibre and in 1 group for paraffin gauze, but none of the groups considered the effectiveness of chlorhexidine, film, honey, or iodine as evidence-based (Table 2). Strikingly, all groups often used paraffin gauze (with the exception of manufacturers), but only few in each group stated to be aware of any evidence about its effectiveness.

<table>
<thead>
<tr>
<th>Stakeholders</th>
<th>First Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeons</td>
<td>Silver sulfadiazine (67%)</td>
</tr>
<tr>
<td>Plastic surgeons</td>
<td>Silver sulfadiazine (76%)</td>
</tr>
<tr>
<td>Dermatologists</td>
<td>Silver sulfadiazine (32%)</td>
</tr>
<tr>
<td>General practitioners</td>
<td>Silver sulfadiazine (38%)</td>
</tr>
<tr>
<td>Specialized nurses</td>
<td>Hydrofibre (54%)</td>
</tr>
<tr>
<td>Surgical nurses</td>
<td>Silver sulfadiazine (29%)</td>
</tr>
<tr>
<td>Home care nurses</td>
<td>Paraffin gauze (41%)</td>
</tr>
<tr>
<td>Manufacturers</td>
<td>Hydrofibre (39%)</td>
</tr>
</tbody>
</table>

Awareness and use of the Cochrane Library
Figure 1 shows the awareness and use of the Cochrane Library in daily practice.
Doctors
All doctors were aware of the Cochrane Library; almost half of them used it once or twice a year. Surgeons (10/24; 42%) and plastic surgeons (10/25; 40%) tended to use the Cochrane Library more frequently (monthly or more often) than dermatologists (6/39; 15%) and GPs (1/8; 13%). However, this difference was not statistically significant (p = 0.724). Surprisingly, more than a quarter of the dermatologists and GPs was aware of the Cochrane Library but never used it.

Nurses and manufacturers
Awareness and usage of the Cochrane Library was lower among nurses than among doctors (p < 0.001). Nevertheless, this awareness was much higher in specialized nurses than in surgical and home care nurses (p = 0.001), of whom more than 75% stated to be unaware. The usage of the Cochrane Library among these nursing groups was poor, although more than a quarter of the specialized wound care nurses used it on a monthly basis. Among the manufacturers, two thirds were not aware of, or never used, the Cochrane Library.

Age and use of the Cochrane Library
No significant relations were found between age and awareness or use of the Cochrane Library among any of the groups.

Buyer and pharmacy departments
From the interviews with representatives of the buyer and pharmacy departments we learned that they had no preference and were not aware of any evidence regarding the effectiveness of the wound care products and could therefore not answer the questions posed. They merely ordered and delivered the products as requested by the doctor or nurse taking care of the wound patients.

DISCUSSION
Choices made in wound care by the various stakeholders are not always evidence-based, despite the availability of compelling research evidence from systematic reviews with recommendations for practice. For example, most healthcare professionals hold the opinion that the effectiveness of silver-containing wound care products, in particular silver sulfadiazine, is established, but this is in contrast with the results of three Cochrane systematic reviews, which suggest silver sulfadiazine has more disadvantages compared with other antiseptics. These reviews report on evidence that is sometimes over 10 years old, but this has not reached the minds of present-day wound care professionals yet. The latency between the publication of evidence and its integration in daily practice may take a quarter of a century, and the scientific and clinical realms seem to have reconciled themselves with this notion.
There is an ongoing discussion about the usefulness of Cochrane systematic reviews.\textsuperscript{13,14} Most of these reviews end by concluding that the volume and quality of the existing research is low, the consistency of study designs is lacking (e.g. regarding study endpoints), few replication studies exist, meta-analysis is usually impossible because of heterogeneity of the studies, and most included studies are at high risk of bias.\textsuperscript{13} Therefore, clinicians often receive no recommendations what to do in daily practice.\textsuperscript{14} On the other hand, reality shows that many published trials have methodological inadequacies. Therefore, it is important that Cochrane reviews highlight these methodological inadequacies, so that researchers pay more attention to the methodological quality of future research. In the case of the absence of compelling evidence, clinicians should rely on expert-opinion and consensus-based guidelines to assist clinical decision making.\textsuperscript{13}

Findings in our study are supported by previous cross-sectional studies. Knops et al. showed that surgeons use only about half of the convincing evidence.\textsuperscript{3} Four other studies investigated the awareness and use of the Cochrane Library; they concluded that there is little awareness,\textsuperscript{15-17} and subsequently little use of the Cochrane Library among health care professionals.\textsuperscript{16-18} From a study performed by Sigouin et al. it is known that differences between professional groups exist, as they found a significant difference in favour of oncologists compared to oncology nurses related to awareness of the Cochrane Library.\textsuperscript{15} Therefore, our results seem also true for other countries than only the Netherlands.

To turn the tide, Adamsen et al. proposed to develop an education strategy to provide stakeholders with evidence-based knowledge that empowers them to make evidence-based decisions.\textsuperscript{19} Nevertheless, a recently published systematic review showed that there is insufficient evidence about which types of interventions are effective to encourage the use of systematic reviews by professionals in clinical decision making.\textsuperscript{20} Although it is accepted that not all professionals should be involved in research, stakeholders should be able to critique and apply research pertinent to their area.\textsuperscript{21,22} This is in accordance with the conceptual framework Strauss et al. described, in which they propose that professionals can practice evidence-based medicine in one of three modes – as a doer, a user, or a replicator.\textsuperscript{23} Journal clubs may also be used as a feature to keep abreast of the latest research evidence and enable continuing professional education.\textsuperscript{24}

Not solely educational features will bridge the gap between evidence and practice. Other strategies are also needed. For example, a multidisciplinary local wound care committee, including those able to search and present relevant evidence, should coordinate the wound care policy within an institution or region. In addition, this may reduce the variability between professionals and institutions in the use of wound care materials (e.g. wound care materials used for the treatment of donor site wounds).\textsuperscript{25} Furthermore, opinion leaders and managers should be involved as they are important in improving and promoting evidence-based care.\textsuperscript{26} Moreover, to improve the change of professional practice and implementation of evidence-based recommendations, barriers should be identified and dealt with.\textsuperscript{27}
It is not only the task of the professionals to bridge the gap. Scientists should reach out and carry out reliable and relevant research and produce readable information.\textsuperscript{28} This is seldom the case; scientists pay relatively little attention to the implementation of the findings of their research in routine clinical care and usually use passive approaches to disseminate information (e.g. publication in professional articles). These approaches are generally ineffective, and at best, result only in small changes in practice.\textsuperscript{29} Hence, scientists and professionals should work together to investigate relevant clinical questions derived from daily practice. Finally, the awareness-to-adherence model which describes seven stages (awareness, acceptance, applicable, available and able, acted on, agreed to, and adherence) may help to get insight if the transfer between the different stages is insufficient.\textsuperscript{30,31} If this is the case, specific interventions could be used to improve this. For example, electronic scanning and alert services may be useful to help stakeholders to become aware of important changes, such as the journal of Evidence-Based Medicine and Evidence-Based Nursing.\textsuperscript{31} Furthermore, stakeholders should act aptly in terms of internalizing convincing evidence in their daily practice. In some cases, simple reminders help to act correctly.\textsuperscript{31} Finally, stakeholders should not forget the role of the patient. Patients are to be informed adequately on potential risks and benefits to improve their adherence to the wound protocol after discharge from hospital.

A limitation of this study is the small scope of the (Cochrane) systematic reviews we used for the assessment of awareness and usage of evidence. However, the wounds involved (possibly with the exception of burn wounds) reflect daily practice in wound care and may well be indicative of the situation for other indications. Second, it could be possible that stakeholders rarely turn to the Cochrane Library for answers to clinical dilemmas. Currently, an increasing number of medical schools and residency programmes are instituting curricula for teaching evidence-based principles.\textsuperscript{32} Therefore, modern stakeholders in wound care should be aware of the available evidence in the Cochrane Library, but its use is not (yet) sufficiently implemented.

Third, we used a self-reported questionnaire, which may have led to socially desirable answering and, subsequently, to an overestimation of Cochrane Library usage.\textsuperscript{33} Yet, this does not change the inferences from our study. Fourth, the questionnaire used is not validated. However, at present, there is no validated questionnaire available. We tried to obtain a first insight in this problem and to make the issue clear for future, more focused studies. Fifth, it is unclear whether the respondents answered reliably, that is, were they really aware of the evidence if they stated there is evidence of effect. Hence, we may even have overestimated their awareness. Lastly, the number of GP respondents in this study was limited, despite of our efforts to contact them. Therefore these results should be interpreted with caution. However, our results from GPs seem to be in accordance with two previous studies.\textsuperscript{18,34} They found that despite the preferences of GPs for evidence-based information (e.g. systematic reviews, and randomized clinical trials),\textsuperscript{34} the majority of respondents were unaware of, or did not use, the Cochrane Library.\textsuperscript{18,34}
Present-day reality is that producing systematic reviews with recommendations and disseminating the results does not naturally result in more awareness and use of the evidence in the Netherlands. Using wound care products while contradicting evidence is available endorses this statement. Although our results may not be surprising, it is important that the basic premise has been confirmed by a quantitative analysis to invoke improvement actions. The present availability of compelling research evidence and the positive attitude toward evidence-based practice, $^3,17,35-37$ should make evidence-based decision making in wound care possible.
REFERENCES


<table>
<thead>
<tr>
<th></th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>31.</td>
<td>Glasziou P, Haynes B. The paths from research to improved health outcomes. Evid Based Nurs 2005;8:36-8</td>
</tr>
</tbody>
</table>
Chapter 11

Epilogue
EPILOGUE

Three main key tools in quality of care

Evidence based practice (EBP), continuous quality improvement (CQI) and clinical research may be considered essential tools to bridge the chasm between actual and desired quality of health care, as illustrated in the figure.\textsuperscript{1-5} The first, EBP, integrates the best available scientific evidence into clinical practice. This should improve patient safety and patient-centered outcomes, and foster CQI.\textsuperscript{4,5} EBP focuses on “doing the right things at the right time for the right patient,” based on evidence and on improving evidence-based decision making, whereas CQI aims at “doing things right” based on local processes.\textsuperscript{6} EBP, CQI and research can be seen as a triangle of three main tools to solve clinical problems and deficiencies in quality of health care and to promote knowledge circulation.\textsuperscript{7} Only if proper research has been done the right thing to do for the patient becomes clear. And only if there is evidence for effectiveness, it can become clear how these right things can be done correctly (CQI). Understanding and applying these three tools empowers health care professionals to detect clinical issues related to quality of care, to develop new knowledge needed to address these issues, and to implement the results of research in daily practice and education.\textsuperscript{7} Thus, the combination of these three main tools should eventually help to bridge the quality chasm.

The studies described in this thesis contribute to narrowing the chasm by enlarging knowledge development and knowledge use through the promotion of evidence-based decision making (Chapters 2 to 6). The effect of the generated evidence should be to improve quality of care by means of timely and patient-centered care (Chapter 7), valid efficiency parameters (Chapter 8), and effective implementation (Chapters 9 and 10). The latter studies show we may have contributed to narrowing the chasm, but knowledge implementation needs ongoing attention in the future.

Current nursing issues hindering knowledge implementation and evidence based decision making

For nurses, two main issues may hinder knowledge implementation and evidence-based decision making: first, interdisciplinary collaboration and communication; second, shortcomings in the delivery of evidence-based care.

Patient care environments have become more complex; therefore, interdisciplinary collaboration and coordination of care is a prerequisite to the delivery of high-quality patient care.\textsuperscript{8-10} The cornerstone of interdisciplinary collaboration is effective, skilled communication and understanding and appreciating each other’s professional role\textsuperscript{11} when setting joint priorities in patient care goals. Unfortunately, priorities in patient care tend to differ between doctors and nurses.\textsuperscript{12} In setting joint priorities, nurses should be proficient in reasoning skills in order to discuss the professional and moral arguments of their priorities, otherwise nurses’ contributions will be neglected. Thus, nurses should be proficient not
only in communication and collaboration, but also in clinical reasoning skills. At present, nurses still tend to rely on personal experience and communication with colleagues rather than on scientific sources of knowledge and rigorous quality evaluation. This is mainly due to a lack of a thorough and quantitative scientific basis for many nursing decisions and a deficit in academic training. Structured evidence-based tools, such as the triage systems (Chapters 2 and 3), as well as knowledge about the clinical relevance of monitoring vital signs (Chapter 5), may supplement personal experience and preferences with externally validated criteria, leading to an objective judgment. In addition, improving nurses’ knowledge of scientific jargon and reasoning based on research findings, something to which doctors are already more accustomed, will facilitate communication between multiple disciplines. Furthermore, a joint effort by nurses and doctors appears essential in order to arrive at a uniform care policy based on evidence and guidelines (Chapters 9 and 10). Several chapters in this thesis showcase the fact that conducting research in which nurses and doctors collaborate and in which the priorities of both are addressed stimulates inter-disciplinary collaboration and communication.

The second nursing issue concerns the actual delivery of evidence-based care in practice. At present, EBP and CQI are gradually integrated in initial nursing curricula and post-graduate education for nurses. The upscaling of the current nursing education to a more academic level, to facilitate deliverance of evidence-based care by nurses, is promising. Today’s nurses have a positive attitude towards EBP and agree that practicing EBP improves quality of care. Therefore, the ground seems fertile to promote the delivery of evidence-based care. Still, in practice there are a number of barriers present to employment of EBP and CQI, for example time constraints, knowledge gaps, awareness or availability of evidence, and management support. We encountered similar difficulties in our attempt to achieve lasting adherence to a local guideline on postoperative temperature measurements (Chapter 9) and when studying knowledge and use of evidence in national wound care (Chapter 10). Although adherence could be achieved in the short-term (Chapter 9), awareness could not be maintained without strong management support. Also, disseminating results of research does not naturally result in more awareness and use of the evidence (Chapter 10).

**Promoting EBP and CQI in clinical practice**

To overcome the barriers mentioned above, healthcare and educational organizations should encourage intertwinement of EBP, CQI, and research, and stimulate interdisciplinary collaboration using all three key tools. This requires a facilitating organizational culture, in which changes and novel findings from research can flourish to employ CQI, and in which CQI triggers an EBP cycle to check for evidence. In turn, both CQI and EBP can result in ideas or even need for further research, as we have shown in this thesis. For example, from a CQI perspective an accurate triage system was warranted to improve quality of care. By following the EBP steps, the literature showed a lack of evidence. This generated
a research question on the validity and accuracy of triage systems (Chapters 2 and 3). Using the results of our research, we decided which system to implement and contributed to CQI (Chapter 7). This cyclic and iterative process can and should also start with clinical reasoning about routine clinical procedures. This is exemplified by our studies in Chapters 5 and 6, in which all available evidence on routinely performed vital sign measurements and the use of silver dressings in preventing wound infection was summarized. The results are now being used for CQI. These examples show that disciplines involved should reflect on and question their daily professional actions and decisions, to detect discrepancies between present-day and desired quality of care.

Suggestions of ways to achieve intertwinement of these tools are, in the first place, the encouragement of an interdisciplinary rather than multidisciplinary collaboration by the current management. While “multidisciplinary” collaboration entails the mere involvement of different professionals, e.g. doctors and nurses, in patient care, “interdisciplinary” collaboration means all health-care professionals involved realize that decision-making depends on sharing information and formulating common patient-centered goals. This is exemplified in this thesis by the tasks nurses perform in an interdisciplinary setting in the areas of triage (Chapter 2 and 3), behavioral interventions (Chapter 4), patient monitoring (Chapter 5), and wound care (Chapter 6). Additional aspects investigated here are the evaluation of patient flow and satisfaction after implementing a triage system (Chapter 7) and the efficiency and feasibility of screening programs by quantifying the (nursing) effort needed (Chapter 8).

Second, collaboration between the medical and nursing discipline is not only necessary in patient care and education, but also in research programs. At present, research organizations and programs still focus on medical models, while nurses tend to focus more on distinct patient-centered outcomes. In general, the medical model usually lacks nursing care outcomes. Given the many areas of nursing research and the limited academic nursing capacity, research by nurses should be embedded in existing research groups using the existing research organization and infrastructure. Management and leadership interventions and approaches should facilitate this. Chapters 4 and 8 are examples of interdisciplinary collaboration in existing research groups in medical psychology, clinical vascular, cardiology, and nursing care.

Third, to create or strengthen an academic culture of evidence-based CQI, academic thinking and ongoing organizational support is needed. Furthermore, research capacity and capability building is necessary to conduct research that facilitates evidence-based decision-making in nursing. To make this possible in daily clinical practice, a balance needs to be achieved between the workload of daily activities for the primary process and the academic or research activities. A joint venture of role-modeling by different stakeholders is desirable, but is not yet available, to really make EBP work and to enhance the quality of care of patients as became clear from our research on guideline adherence described in Chapter 9.
Fourth, we learnt from Chapter 10, but also from previous studies,\textsuperscript{31,32} that clinical nurses are generally less aware of sources of existing evidence than doctors. This calls for knowledge translation and circulation, which is a dynamic and iterative process and moves along a continuum from awareness of, and adherence to, available evidence to improvement of the patients’ health. Knowledge translation can be used in CQI initiatives and may be helpful to close the evidence-practice gap.\textsuperscript{33} To facilitate the use of evidence in practice, practical tools such as clinical pathways and evidence-based recommendations can be used.\textsuperscript{34}

Fifth, in this thesis research utilization efforts were studied among a range of disciplines and stakeholders involved in health care (Chapters 9 and 10). Implementing new knowledge and changing clinical practice should be done in an interactive, dynamic way, in which personal experiences are valued and attitudes and beliefs are addressed.\textsuperscript{35-37} The effect of these team interventions should be assessed in different models of care delivery to determine how these interventions can induce the desired change. Identification of new evidence, and assessment of whether it offers new information that might change the latest standards in clinical practice, should be accelerated. Several higher impact research journals address this by explicitly stating what a new publication adds to the present body of knowledge. Updating of current practice implies a critical evaluation of this new knowledge and what that means for all stakeholders in clinical practice (Chapter 7 and 8).\textsuperscript{38} With this in mind, we studied patient satisfaction and effects on waiting time after the implementation of a triage system (Chapter 7). Furthermore, the NNS statistic can help to quantify the efforts needed to evaluate the usefulness and efficiency of interventions (Chapter 8). This evaluation cannot be done without collaboration and professional discussion on the consequences of the proposed change.\textsuperscript{39}

Finally, the studies described in this thesis illustrate the value of collaboration between nurses and doctors, without which none of these research projects would have succeeded. The results are relevant to both disciplines, as well as to managers and educators. Furthermore, the interdisciplinary character of the studies fosters the implementation process. From Chapters 9 and 10 we can conclude that the success of implementing such evidence depends on the ability of a team which functions in a coordinated, effective manner and communicates scientific arguments and findings. This seems fitting in the current era, in which the creation of a culture of continuous change and the determination of effective strategies to inform and engage all stakeholders in ways that help ensure the delivering of safe, cost-effective, and high-quality care are becoming prerequisites.

**Future research**

In the coming years, research in the nursing realm will require prioritization of patient-oriented research topics on the cutting edge of various disciplines involved. This can be realized by using explicit criteria to determine its societal priority, for example based on the mnemonic of the 8 D’s: Death, disease, disability, drug effects, discomfort, dysfunction,
dissatisfaction and dollars.\textsuperscript{40} A high score on one or more of these items implies a higher priority of the research topic. Another tool with which to assess priority are the criteria formulated by the scientific advisory board of the Dutch Association for Nurses and Caregivers.\textsuperscript{41} Research topics should be distinguishable from other research programs; should concern general patient care problems; and should have an impact on nursing care or care problems for specific patient groups. We agree on these additional criteria, but the criterion of uniqueness to some extent conflicts with our idea of interdisciplinary collaboration. Nursing research themes can be derived from current (inter)national perspectives, organizational research spearheads, and patient associations, but also from questions formulated by clinical nurses on the wards. The latter will enable demand-driven research and ensure the delivery of useful scientific knowledge for the nursing professions. From this interaction between bottom-up and top-down processes, themes can be included in interdisciplinary research programs, in which nurses need to have their own role, focus, and research budget.

The absence of a long-term national research agenda with set priorities leads to fragmented research themes. To avoid fragmentation, these national research programs should set out the (long-term) framework and programs based on explicit societal and scientific criteria, as mentioned above. Currently, patients, who are the actual consumers of research findings, are not included in the prioritization process. They should be encouraged and involved to do so in the future. The general scope of the currently prioritized research themes for nursing comprises self-management and limitations in daily activities and self-care.\textsuperscript{42} However, this is at odds with the medical way of prioritization, which is mostly illness-specific. With the arguments given in this thesis, we discourage mono-disciplinary prioritization and offer a passionate plea for interdisciplinary, evidence-based care and research.

**Figure:** Relationship between CQI, EBP and research
CQI: **continuous quality improvement.**
Based on the six sigma philosophy, in which the PDCA (Plan, Do, Check, and Act) cycle has expanded to the DMAIC (Kumar 2010).\(^1\)

- D  define opportunity
- M  measure performance
- A  analyze opportunity
- I  improve performance
- C  control performance

**EBP: Evidence Based Practice (Sackett 2000).**\(^2\)
- ask an answerable question
- acquire: find the best evidence
- critically appraise and evaluate the evidence
- apply the evidence in combination with clinical experience and patient values
- assess and evaluate outcomes

**Research**
- Based on the empirical cycle (De Groot 1981).\(^3\)
- observation: collecting and organization of empirical facts: forming hypothesis
- induction: formulating hypothesis
- deduction: deducting consequences of hypothesis as testable predictions
- testing: testing the hypothesis with new empirical material
- evaluation: evaluating the outcome of testing or else.
REFERENCES


Summary
SUMMARY

Improving quality of care is important for patients and healthcare professionals. Unfortunately there is a chasm between present-day and desired quality of healthcare. Closing this quality chasm is advocated.

Evidence Based Medicine is a tool to improve quality of healthcare by maintaining professional knowledge and fostering lifelong learning. Not only doctors but also nurses have increasingly adopted this paradigm in their decision making regarding patient care. This has led to the more general term Evidence Based Practice (EBP). EBP promotes the integration of new research knowledge into practice and fosters continuous quality improvement (CQI).

This thesis contributes to the body of knowledge of nursing practices, the promotion of evidence-based decision making, and efforts to overcome the challenges in the implementation process. For each decision investigated, its effect on the quality of care is measured.

Part 1: Evidence-based decisions in nursing

**Novel practices**

Overcrowding of Emergency Departments (EDs) means accurate triage systems are necessary in order to avoid delays in critical patient care. In the Netherlands, a guideline on triage was released in 2004, in which triage was defined as a dynamic decision process to prioritize the urgency need of patients entering the ED. Despite evidence of validity and reliability for the Emergency Severity Index (ESI) and a lack of research evidence for the Manchester Triage System (MTS), the latter was promoted as the preferred triage system in the Netherlands. To make an evidence-based decision about whether the MTS or the ESI should be implemented at our ED, the validity and inter- and intra-observer agreements for both triage systems were investigated.

**Chapter 2** deals with a prospective observational comparative study, in which the amount of undertriage is compared: it looks at the validity of two structured triage systems (MTS and ESI) and the local informally structured triage system (ISS) in use, and their relation to resource use, hospital admission, and length of stay (LOS) in a large patient group. Triage ratings in 421 patients treated by ED physicians were compared with a reference standard. Undertriage occurred in 20%, 11% and 8% in the ESI, MTS and ISS, respectively. In all 900 patients triaged, the urgency levels across all three systems were associated with a significant increase in resource use, hospitalization rate, and LOS. Although the ESI showed the highest percentage of undertriage and the ISS the lowest, it does seem preferable to use a verifiable, formally structured triage system. In **Chapter 3** inter and intra-observer agreement of the MTS and the ESI is compared, using paper-based clinical scenarios. A group of nurses who were familiar with the triage systems and another group who had
just trained in them judged the same scenarios. Interobserver unweighted kappas were 0.76 (95% CI 0.68 to 0.83) for MTS and 0.46 (95% CI 0.37 to 0.55) for ESI. Intra-observer unweighted kappas were 0.84 (95% CI 0.73 to 0.94) for MTS and 0.65 (95% CI 0.59 to 0.72) for ESI. These results showed that the MTS had a higher inter and intra-observer agreement than the ESI. Both studies are helpful, as they underpin with evidence clinical decisions about which triage tool to use (as well as the decision made in our ED) and build up the body of knowledge regarding the validity and agreement of triage tools.

The second example of a novel procedure was the introduction of a smoking cessation strategy in cardiovascular outpatients. Systematic reviews show that different smoking cessation programs, for example nicotine replacement therapy (NRT), are effective in different settings and patient populations. Specifically, the Minimal Intervention Strategy (MIS) is propagated by the Dutch expertise center on tobacco control (STIVORO). The MIS is a structured behavioral smoking cessation program and was adapted for cardiology inpatients (C-MIS); nurses are largely responsible for the intervention. This C-MIS was marginally effective in inpatients, but its effectiveness in the cardiovascular outpatient setting was unknown. Therefore, Chapter 4 addresses the effectiveness of this nurse-led intervention, in addition to NRT, to support smoking cessation in cardiovascular outpatients. Results from this randomized clinical trial showed that self-reported abstinence was 17% in the NRT group and 21% in the NRT + C-MIS group (AR 4%; 95% CI -6 to 14%). The MIS intervention was found ineffective and should not be implemented. This study was helpful in the decision process about whether or not to implement the C-MIS in daily nursing practice in cardiovascular outpatient clinics.

Routine clinical procedures

Studies fueled by uncertainty about the effectiveness of certain procedures in routine clinical care are presented in the next two chapters. Worldwide, several governmental institutions have developed guidelines for the timely identification of acutely ill medical patients, and recommend the use of early warning scores (EWS) or related systems. The consequences of these guidelines’ implementation are a substantial increase in measurements of vital signs, with doubtful clinical relevance. Therefore a systematic review is conducted, as described in Chapter 5, and the clinical relevance of routinely measured vital signs in medical and surgical hospitalized patients to detect mortality, septic or circulatory shock, ICU admission, bleeding, re-surgery or infection is determined. Vital signs included body temperature, heart rate, blood pressure, oxygen saturation and respiratory rate. A total of 14 studies were included in this systematic review: one prospective accuracy study and 13 observational studies. Although some discriminative likelihood ratios were found, the accompanying predictive values were low to moderate. There is scarce evidence available on the clinical relevance of routine vital-sign measurements. Hence, no evidence-based recommendation could be distilled from the current literature. There is a need for further prospective diagnostic accuracy and prognostic research on single or combination routine
vital-sign measurement to support daily routine and clinical decision making, and to avoid
unnecessary practices.

Another widely used routine is the use of silver sulfadiazine (SSD) in patients with
burn wounds. Many health care providers believe that SSD prevents wound infection and
promotes wound healing. Other silver-containing dressings have also become popular,
despite the absence of robust evidence for their effectiveness. Hence, a systematic review
(Chapter 6) is conducted to establish the effect of silver-containing dressings and topical
agents for the prevention of wound infection and the promotion of wound healing in
uninfected wounds. A total of 26 studies were included, 20 of which concerned patients
with burns. Heterogeneity of treatments and outcomes precluded meta-analyses. This
systematic review shows that there is insufficient evidence to establish whether silver-
containing dressings or topical agents promote wound healing or prevent wound infection.
However, some evidence pleads against the use of SSD for this purpose.

The results of both studies illustrate that, for some routine clinical practices, surprisingly
little evidence is available to support clinical decision making.

Part 2: Effect of evidence-based decisions on the quality of care

Influence of evidence-based decisions on patient care

Based on the results described in chapters 2 and 3, the MTS was implemented at our
hospital’s ED. Implementation of an accurate triage system could positively influence the
waiting time of critically ill patients and thereby the quality of care for these patients.
Waiting times at EDs are also correlated with patient satisfaction. Yet waiting times and
patient satisfaction have not been studied for the MTS. Chapter 7 deals with a prospective
observational before–after study, in which the influence of the implementation of MTS on
waiting times, LOS and patient satisfaction was determined in two consecutive patient series
(n=1808). Waiting times did not decrease after implementation of the MTS, but were better
distributed over urgency levels. In the After Implementation group, patient satisfaction was
significantly lower on provision of information and opportunity to explain their problems,
but waiting time and the feeling that their problem had been sorted out scored better. No
significant differences were found between triaged and non-triaged patients. The better
distribution over urgency levels justifies the decision to use the MTS. Still, implementing MTS
on its own is not sufficient to improve the efficiency and quality of EDs.

Smoking cessation is an important factor in reducing cardiovascular mortality, but
considerable effort is needed to successfully persuade patients to quit smoking. Treatment
effects are mostly presented as the number needed to treat (NNT), but this does not reflect
the total effort necessary to identify all patients who could or should potentially be treated.
To determine the usefulness, relevance and efficiency of screening programs, the number
needed to screen (NNS) may be a useful measure. Based on the prospectively sampled
cohort data for the study described in Chapter 4, we studied in Chapter 8 the efficiency of the C-MIS and determined the NNT and NNS in relation to the number of deaths prevented over a five-year period. Subgroup analysis was performed for patients attending the clinics for first or routine follow-up attendees. The NNS was 687 (minimum (min)–maximum (max): 141–∞) in the cardiology clinic, and 574 (min–max: 134–∞) in the vascular surgery clinic. With the C-MIS for first-time and routine follow-up attendees, only six (min–max: 0–36) and zero (min–max: 0–25) deaths, respectively, could be prevented. In terms of the effectiveness of the C-MIS in addition to NRT, there is some benefit for first-time attendees but no benefit for routine follow-up attendees in preventing death.

Chapters 7 and 8 exemplify the consequences of the new knowledge gained in chapters 2-4 for different stakeholders in clinical practice.

**Implementation challenges to change routine care**

The last two chapters describe two challenges for the implementation of changes in routine care. Despite a large body of knowledge available on effective implementation strategies of evidence-based guidelines into clinical and nursing practice, the ‘magic bullet’ does not exist. Multifaceted implementation strategies have been proposed to implement guidelines. In the review described in Chapter 5, fourteen studies were included: 13 observational studies and one prospective accuracy study. The last was performed by our own study group and assessed the diagnostic accuracy of routine postoperative body temperature measurements (BTM). By using a multifaceted implementation strategy, unnecessary BTM on surgical wards were abolished. Although short-term guideline adherence rates were high (91%), regression to old habits remains a common human flaw. Therefore the first challenge, i.e. long-term adherence, is described in Chapter 9, in which a mixed-methods analysis was used to determine if adherence was persistent over time and which facilitators and barriers contributed to adherence over time. Patient records were examined to see whether the recorded BTM complied with the guideline recommendation. Overall adherence rate was found to be 50%. Facilitators were belief in the advantages of the guideline and staff support. Barriers were the controversial nature of the guideline, the lack of self-efficacy among nurses and doctors regarding clinical judgment to identify an infection without the use of BTM, and a lack of management and staff doctor support. Therefore the conclusion was that to ensure long-term adherence, guidance is needed on every ward and indicators should be developed to monitor the required changes. Moreover, guidelines and their recommendations should be an integral part of the initial training period, standard education, and knowledge transfer of both nurses and physicians.

The second challenge is dealing with one of the barriers to adherence, i.e. awareness of the available evidence. It is impossible for doctors and nurses to keep abreast of all newly published research evidence. Therefore, available evidence is increasingly aggregated into systematic reviews (SR) and clinical guidelines. Despite active dissemination of the results of SRs and guidelines, 30–50% of patients receive care that is not in accordance with available
evidence. Based on the results of Chapter 6, Chapter 10 describes a cross-sectional study investigating the awareness and use of evidence of 262 wound-care stakeholders (doctors, nurses, buyers, pharmacologists and manufacturers) in their choice of wound dressings. Doctors preferred conventional antiseptics (e.g., iodine), while specialized nurses and manufacturers favoured popular products (e.g., silver). Most stakeholders considered silver-containing products to be evidence-based antiseptics, which contradicts scientific results. Nurses and manufacturers in particular were unaware of, or never used, the Cochrane Library. The results show that available high-quality evidence in wound care is not internalized equally by the various stakeholders, as is required to ensure evidence-based decision-making. As specialized nurses performed best in this regard, it appears that they can be most helpful in closing this quality chasm.

To solve clinical problems and to promote knowledge circulation, interdisciplinary collaboration in continuous quality improvement, EBP, and research should be seen as three complementary key tools in bridging the chasm between actual and desired quality of care.
Chapter 12

Samenvatting
SAMENVATTING

Het verbeteren van de kwaliteit van zorg is belangrijk voor patiënten en beroepsbeoefenaars van de gezondheidszorg. Helaas bestaat er een kloof tussen de huidige en gewenste kwaliteit en het is van belang deze kloof te overbruggen.

Een instrument om de kwaliteit van de gezondheidszorg te verbeteren is Evidence Based Medicine (EBM). EBM bevordert niet alleen de professionele kennis, maar ook het levenslang leren. Behalve artsen zijn ook verpleegkundigen dit paradigma in toenemende mate gaan gebruiken in hun besluitvorming met betrekking tot de patiëntenzorg, waarbij het verpleegkundig handelen in de praktijk ondersteund wordt door resultaten uit wetenschappelijk onderzoek. Dit heeft geleid tot de meer algemene term Evidence Based Practice (EBP). EBP bevordert de integratie en implementatie van nieuwe kennis uit wetenschappelijk onderzoek in de praktijk en is daarmee een pijler voor continue kwaliteitsverbetering (CQI).

Dit proefschrift draagt bij aan de kennis van verpleegkundigen, de bevordering van evidence based besluitvorming, en het aangaan van een aantal uitdagingen in het implementatieproces. Het resultaat van vrijwel elk onderzoek in deel 1 van dit proefschrift leidde tot een evidence based besluit in de verpleegkundige praktijk. Hierbij werd ook de implementatie en het effect op de kwaliteit van de zorg gemeten. Dit staat beschreven in deel 2.

Deel 1: Evidence based besluitvorming in verpleegkunde

Nieuwe procedures

Het eerste voorbeeld van een nieuwe procedure op basis van verkregen bewijsmateriaal betreft het invoeren van een nieuw triagesysteem op de afdeling Spoedeisende Hulp (SEH). Door de grote toename van het aantal patiënten op de SEH zijn gestructureerde triagesystemen noodzakelijk om vertraging in de zorgverlening aan ernstig zieke patiënten te voorkomen. In de Nederlandse richtlijn ‘Triage op de SEH’ uit 2004 wordt triage gedefinieerd als een dynamisch beslissingsproces waarbij de urgentie van patiënten wordt bepaald. Het Manchester Triage Systeem (MTS) en de Emergency Severity Index (ESI) zijn twee gestructureerde triagesystemen die hiervoor gebruikt kunnen worden. In de richtlijn werd een voorkeur uitgesproken voor het MTS, ondanks een gebrek aan wetenschappelijk bewijs over de validiteit en betrouwbaarheid van dit systeem, terwijl dit bewijs er wel was voor het ESI.

Om tot een evidence-based besluit te komen of het MTS of het ESI geïmplementeerd moest worden op de SEH van het Academisch Medisch Centrum bij de Universiteit van Amsterdam (AMC), hebben we de validiteit en de inter- en intra-beoordelaar overeenkomsten voor beide triagesystemen onderzocht.
Hoofdstuk 2 beschrijft een prospectieve observationele vergelijkende studie, waarin de validiteit werd vergeleken binnen de twee gestructureerde triage systemen (MTS en ESI) en het op dat moment nog in gebruik zijnde ‘informele’ triase systeem van de SEH. Verder wordt een relatie gelegd tussen urgentiecategorieën en gebruik van diagnostiek, ziekenhuis opname en verblijfsduur op de SEH. De hogere urgentiecategorieën van alle drie triagesystemen bleken bij 900 getrierte patiënten geassocieerd met een aanzienlijke toename van diagnostiek en verlenging van ziekenhuisopname en verblijfsduur.

De urgentiecategorieën van 421 patiënten, welke alleen door een SEH arts werden behandeld, werden vergeleken met een referentiestandaard. Als een maat voor validiteit werd het percentage ondertriage bepaald. Hoewel het ESI het hoogste percentage ondertriage had en het ‘informele’ triagesysteem het laagste, is het aan te bevelen om een formeel gestructureerd triagesysteem te gebruiken om uniformiteit en transparantie in triage te bevorderen.

In hoofdstuk 3 wordt de inter en intra-beoordelaar overeenkomst van het MTS en het ESI vergeleken met behulp van papieren klinische scenario’s. Achtteien verpleegkundigen verdeeld in drie groepen; een groep ervaren in het gebruik van het MTS, een groep ervaren in het ESI, en een groep opgeleid in beide triagesystemen, beoordeelden dezelfde scenario’s. De resultaten toonden aan dat het MTS een hogere inter en intra-beoordelaar overeenkomst had dan het ESI.

Beide studies helpen bij de klinische beslissing welk triagesysteem gebruikt kan worden op een SEH, in het bijzonder de beslissing die wijzelf gemaakt hebben voor de SEH van het AMC.

Het tweede voorbeeld van een nieuwe procedure was de invoering van een stoppen-met-roken interventie bij cardiovasculaire patiënten op de polikliniek. Gebaseerd op systematische literatuur overzichten blijken verschillende stoppen-met-roken programma’s, zoals nicotine vervangende therapie (NVT), ter ondersteuning van stoppen met roken effectief in verschillende patiëntengroepen. Daarnaast wordt vooral de minimale interventie strategie (MIS) gepropageerd door de Stichting Volksgezondheid en Roken. De MIS is een gestructureerde gedragsinterventie voor het stoppen met roken en is aangepast voor klinische cardiologische patiënten (C-MIS) bij wie verpleegkundigen grotendeels verantwoordelijk zijn voor de interventie.

In hoofdstuk 4 wordt het gerandomiseerde klinische onderzoek beschreven waarin de effectiviteit van de combinatie van de C-MIS met NVT vergeleken is met het gebruik van alleen NVT bij 385 cardiovasculaire patiënten op de polikliniek. Resultaten van dit onderzoek toonde aan dat de C-MIS interventie niet effectief was.

Deze studie was doorslaggevend bij het besluitvormingsproces voor het niet standaard uitvoeren van de C-MIS in de dagelijkse praktijk bij cardiovasculaire patiënten op de polikliniek.
Onzekerheid over routinematige klinische handelingen

Er zijn veel routinematige handelingen in de klinische praktijk waarvan de effectiviteit betwijfeld wordt. In de volgende hoofdstukken worden hiervan twee beschreven.

Wereldwijd zijn door verschillende overheidsinstellingen richtlijnen ontwikkeld die leiden tot een tijdige identificatie van acuut zieke patiënten, en deze raden hierbij het gebruik aan van de ‘early warning scores’ of gerelateerde systemen. Een gevolg van deze richtlijnen is een aanzienlijke toename van het meten van vitale functies, zoals bijv. lichaamstemperatuur, polsfrequentie, bloeddruk, ademhalingsfrequentie, en zuurstofsaturatie. Echter de klinische effectiviteit van het meten van deze vitale functies is twijfelachtig.

Hoofdstuk 5 geeft een systematisch literatuur overzicht weer waarin de effectiviteit van routinematig gemeten vitale functies bij interne en chirurgische klinische patiënten werd onderzocht in relatie tot mortaliteit, septische of circulatoire shock, intensive care opname, bloeding, heroperatie of infectie. De 14 studies geïncludeerd in dit systematische literatuur overzicht leverden enkele discriminatieve likelihood ratios op, maar de bijbehorende predictieve waarden waren laag tot matig. Er blijkt dus weinig wetenschappelijk bewijs te zijn om het doen van deze routinematige metingen te onderbouwen. Om onnodige metingen te voorkomen en besluitvorming mogelijk te maken ten aanzien van deze dagelijkse routine is verder onderzoek nodig naar de diagnostische accuratesse en prognostische waarde van de individuele bijdrage of combinatie van vitale functies.

Een ander veel gebruikte routinehandeling is het gebruik van zilver sulfadiazine (SSD) bij patiënten met brandwonden. Veel verpleegkundigen en artsen zijn er van overtuigd dat SSD helpt om wondinfectie te voorkomen en wondgenezing te bevorderen. Ook andere zilver bevattende wondverbanden nemen in populariteit toe, ondanks de twijfelachtige effectiviteit.

Zoals in hoofdstuk 6 beschreven, is daarom een systematisch literatuur onderzoek uitgevoerd waarin de effectiviteit is bepaald van zilver bevattende verbanden ter preventie van wondinfectie en het bevorderen van wondgenezing bij niet geïnfecteerde wonden. Er werden 26 studies geïncludeerd, waarvan 20 studies patiënten met brandwonden betroffen. Er bleek onvoldoende wetenschappelijk bewijs te bestaan om vast te stellen of zilver bevattende verbanden wondgenezing bevorderen of wondinfectie voorkomen. In het bijzonder pleit het bewijs zelfs eerder tegen het gebruik van SSD.

De resultaten van deze twee studies illustreren dat er voor een aantal routinematige klinische handelingen verbazingwekkend weinig wetenschappelijk bewijs beschikbaar is om deze handelingen te onderbouwen.
Deel 2: Effecten van evidence-based keuzes op de kwaliteit van zorg

Gebaseerd op de resultaten van hoofdstuk 2 en 3, werd het MTS geïmplementeerd op de SEH van het AMC. Gestructureerde triagesystemen kunnen de wachttijd van urgente patiënten positief beïnvloeden. Wachttijden hangen samen met patiënt(on)tevredenheid en door het verkorten van de wachttijd kan de patiënttevredenheid verbeterd worden. Er zijn nog geen studies uitgevoerd die deze effecten voor het MTS beschrijven.

In hoofdstuk 7 worden een observationele voor- en nameting beschreven, waarin de effecten van het gebruik van het MTS zijn onderzocht bij twee consecutieve patiëntengroepen \((n=1808)\) op wachttijden, SEH verblijfsduur en patiënttevredenheid. De wachttijden namen niet af, maar waren wel beter verdeeld over de verschillende urgentiecategorieën. De patiënttevredenheid bleek significant te verminderen met betrekking tot de informatieverstrekking en de gelegenheid om hun probleem uit te leggen. Wel waren patiënten tevredener met betrekking tot de wachttijd en ze hadden meer het gevoel dat hun probleem was opgelost. Er waren geen significante verschillen tussen getrieerde en niet getrieerde patiënten. De betere distributie van wachttijden over de urgentiecategorieën rechtvaardigt de beslissing om het MTS te gebruiken. Het gebruik van het MTS alleen is echter niet voldoende om efficiëntie en kwaliteit van zorg te verbeteren.

Stoppen met roken is een belangrijke factor bij het reduceren van cardiovasculaire mortaliteit, maar het kost beroepsoefenaars van de gezondheidszorg veel inspanning om patiënten inderdaad te laten stoppen. De ‘number needed to treat’ \((NNT)\) is een veelgebruikte maat die het aantal patiënten aangeeft dat behandeld moet worden om 1 extra patiënt te laten stoppen met roken. Deze maat reflecteert echter niet de totale inspanning die hiervoor geleverd moet worden, namelijk de identificatie van alle patiënten die potentieel behandeld kunnen worden. Om te bepalen of een screeningsprogramma bruikbaar, relevant en efficiënt is, kan de ‘number needed to screen’ \((NNS)\) gebruikt worden. Gebruikmakend van de data uit de cohortstudie en de effectiviteit van de C-MIS uit de gerandomiseerde studie zoals beschreven in hoofdstuk 4, is zowel een NNT als een NNS berekend.

Hoofdstuk 8 beschrijft deze berekening bij 2725 patienten, waarbij de NNT en NNS ook gerelateerd werden aan het aantal patiënten bij wie overlijden is voorkomen over een periode van 5 jaar. De NNS was 687 in de cardiologie polikliniek, en 574 in de vaatchirurgische polikliniek. Met een C-MIS effectiviteit voor patiënten die voor het eerst de polikliniek bezochten of patiënten die voor een routinecontrole kwamen, werden respectievelijk slechts zes en nul doden voorkomen over een periode van 5 jaar.

Hoofdstuk 7 en 8 geven beide aan welke consequenties de nieuwe kennis uit hoofdstukken 2 t/m 4 kunnen hebben voor diverse betrokkenen in de klinische praktijk.
Implementatie: uitdagingen om routinematige zorg te veranderen
Twee uitdagingen om routinematige zorg te veranderen worden beschreven in de laatste twee hoofdstukken. Ondanks de toegenomen kennis over effectieve implementatie-strategieën van evidence based richtlijnen in de klinische praktijk, bestaat de ‘magic bullet’ niet. Een gecombineerde implementatiestrategie lijkt op dit moment het beste om richtlijnen in te voeren.

Hoofdstuk 5 betrof een systematisch literatuur onderzoek waarin 14 studies waren geïncludeerd. Eén daarvan was een diagnostische studie die door onze eigen onderzoeksgroep is opgezet en uitgevoerd. Hierin werd de diagnostische accuratesse bepaald van routinematige postoperatieve metingen van de lichaamstemperatuur. Door gebruik te maken van een gecombineerde implementatiestrategie werden overbodige metingen op de chirurgische afdelingen afgeschaft. Drie maanden na invoering van dit nieuwe beleid was de compliantie hoog (91%). Het is echter bekend dat mensen na verloop van jaren weer terugvallen in oude routines.

De eerste uitdaging, het bereiken van een lange termijn compliantie, is daarom beschreven in hoofdstuk 9. Gebruikmakend van een mixed methods onderzoek werd bepaald wat de lange termijn compliantie aan de richtlijn was, en tevens welke bevorderende en belemmerende factoren aanwezig waren om compliantie op langere termijn te behouden. Op basis van een beoordeling van 102 statussen bleek de compliantie aan de richtlijn na zeven jaar 50%. Bevorderend voor een lange termijn compliantie waren ondersteuning van het afdelingsmanagement en geloof in de voordelen van de richtlijn. Belemmeringen waren de controversiële aard van de richtlijn, het niet kennen van de richtlijn, gebrek aan vermogen van verpleegkundigen en artsen om op basis van de eigen klinische blik een postoperatieve infectie te identificeren, en een gebrek aan ondersteuning vanuit het afdelingsmanagement. Om lange termijn compliantie te borgen, dient dit gemonitord te worden door gebruik te maken van indicatoren die de gewenste veranderingen weergeven. Verder zouden richtlijnen met aanbevelingen een integraal onderdeel moeten uitmaken van de opleiding tot verpleegkundige of arts.

De tweede uitdaging betreft een betere kennis van de beschikbare resultaten uit wetenschappelijk onderzoek. Het is voor verpleegkundige en artsen onmogelijk om op de hoogte te zijn van alle nieuw gepubliceerde onderzoeken. Daarom worden de beschikbare resultaten in toenemende mate samengevat in systematische literatuur overzichten en richtlijnen. Toch krijgt 30-50% van alle patiënten nog niet de zorg die in overeenstemming is met deze wetenschappelijke onderzoeksresultaten.

Gebaseerd op de resultaten uit hoofdstuk 6, beschrijft hoofdstuk 10 een cross-sectioneel onderzoek bij 262 betrokkenen (artsen, verpleegkundigen, inkopers, apothekers en fabrikanten van verbandmiddelen) waarin werd nagegaan welke soort antiseptische verbanden zij in de praktijk als eerste keus gebruiken bij de behandeling van wonden. Verder werd onderzocht in hoeverre men op de hoogte is van de evidence voor het gebruik van deze
verbanden. Artsen hadden een voorkeur voor conventionele antiseptische verbanden (bijv. jodium), terwijl gespecialiseerde wondverpleegkundigen en wondfabrikanten een voorkeur hadden voor de modernere verbanden (bijv. zilver). De meeste betrokkenen beschouwden de zilver bevattende verbanden als evidence-based, wat niet in overeenstemming is met de wetenschappelijke literatuur. Vooral verpleegkundigen en fabrikanten van verbandmiddelen waren niet op de hoogte van, of gebruikten nooit de Cochrane Library. Deze resultaten geven aan dat de resultaten uit wetenschappelijk onderzoek naar wondzorg niet bij alle betrokkenen voldoende bekend is, terwijl dit een vereiste is voor evidence-based besluitvorming. De gespecialiseerde wondverpleegkundigen scoorden hierin het beste en lijken daarom bij uitstek degenen te zijn die kunnen helpen om de kloof tussen huidige zorg en gewenste zorg te overbruggen.

Om dergelijke klinische problemen op te lossen en de kenniscirculatie te bevorderen is verdergaande interdisciplinaire samenwerking noodzakelijk, waarbij het continu bevorderen van de kwaliteit van zorg, EBP en onderzoek als drie complementaire instrumenten gezien kunnen worden om de kloof tussen de huidige en gewenste kwaliteit van zorg te overbruggen.
List of publications
INTERNATIONAL PUBLICATIONS


NATIONAL PUBLICATIONS


Storm-Versloot MN. Hoe betrouwbaar zijn triagesystemen op de Spoedeisende Hulp: het Manchester Triage Systeem vergeleken met de Emergency Severity Index. Verpleegkunde 2007;22(4).

Storm-Versloot MN. De effectiviteit van zilververbanden voor geïnfecteerde wonden: een systematisch literatuuronderzoek Verpleegkunde 2007;22(4).

Storm-Versloot MN. (Hoe betrouwbaar zijn triagesystemen op de Spoedeisende Hulp: het Manchester Triage Systeem vergeleken met de Emergency Severity Index)?. Bewaarbijlage EBP Bijzijn 2007.


Tump E, Storm-Versloot MN. Wat is de beste verzorging van een externe fixateur? Nursing 2010;16(7/8):39.


A p p e n d i c e s

Dankwoord
DANKWOORD

‘Trots’ heb ik als paranimf naast Dr. Hester Vermeulen en Dr. Loes Wiggers mogen staan. Beide keren vond ik heel bijzonder, maar nu is het mijn beurt om mijn proefschrift te verdedigen. Een ding is zeker, het schrijven van een proefschrift kan niet alleen gedaan worden en door de jaren heen hebben veel mensen direct of indirect hier een steen(tje) aan bijgedragen. Allen ben ik hiervoor dankbaar, en een aantal wil ik in het bijzonder noemen.

Prof.dr. D.A. Legemate, beste Dink. Ik wil je danken voor de kans die je me bood om als verpleegkundige bij jou onderzoek te mogen doen, eerst de ‘TEMP’ studie en daarna de ‘SMOKE’ trial. Dat dit ooit tot een promotie zou leiden lag toen nog niet in de lijn der verwachting. Wel was je bereid mij op mijn eigen manier en tempo te begeleiden naar dit moment. Wat ik zeer bewonderde was het feit dat je bij het kritisch nakijken van ‘mijn’ artikelen, je vaak toch altijd de vinger op de zere plek wist te leggen. En ondanks al je kennis dook je zelf ook weer de boeken in om zeker te weten of de gebruikte methodologie echt goed is.

Prof.dr. J.C. Goslings, beste Carel. Vanwege de onderzoeken die op triage gebied werden uitgevoerd werd je gevraagd als 2e promotor. Dank voor je input en vooral ook voor de snelle reacties op vragen en het nakijken van artikelen.

Dr. D.T. Ubbink, beste Dirk. Je hebt samen met Hester aan de wieg gestaan om mij aan het promoveren te krijgen. Je bijdrage, samenwerking en (methodologische) ondersteuning zou ik niet hebben willen missen. Ik heb veel van je geleerd en de wondcongressen die we samen bezocht hebben waren erg gezellig. Ik heb vooral nog mooie herinneringen aan Praag met al zijn glaswinkeltjes. Ons excuus om deze te bezoeken was dat we voor de regen wilden schuilen.

Dr. H. Vermeulen, lieve Hester. Je was vanaf het begin mijn voorbeeld en samen hebben we veel lief en leed gedeeld. Samen hebben we de opleiding Klinische Epidemiologie voor Verpleegkundigen gevolgd en daarna ook nog ‘even’ de Master of Science. Mijn eerste onderzoeken waren met jou. Hij was een drijvende kracht en je had een droom om Evidence Based Practice voor verpleegkundigen op de kaart te krijgen. Het was fijn dat ik daar een onderdeel in mocht zijn en hierin mee kon groeien. Toen bijna ging promoveren reden Dirk, jij en ik in de auto terug vanuit een congres in Rostock. Een van jullie beiden stelde toen de vraag wanneer ik ging promoveren. Na een positief besluit, was ik dan ook erg blij dat jij (en Dirk) mijn copromotoren wilden zijn. Dit proefschrift zou nooit tot stand zijn gekomen zonder jullie aflatende hulp, en discussies. Mijn dank aan jou (maar ook aan Dirk) is daarom niet in woorden uit te drukken.

Hooggeleerden heren van de promotiecommissie, Prof. Dr. P.J.M. Bakker, Prof. Dr. R.J. de Haan, Prof. Dr. R.J.P.M. Scholten, Prof. Dr. A.B. van Vugt, en Prof. Dr. T. Achterberg, dank voor uw bereidheid het proefschrift op zijn wetenschappelijke waarde te beoordelen. Beste Piet, het was zeer constructief dat ik het afgelopen jaar elke week bij jullie op de KPI mocht komen om aan mijn proefschrift te werken.
Jan Luitse: wat fijn dat ik onder jouw inspirerende leiding onderzoek mocht doen op de SEH naar de verschillende triagesystemen. We waren het er over eens dat het onwenselijk was om het MTS te implementeren zonder dat er enige evidence over de effectiviteit was. Ondanks dat je het altijd erg druk hebt, was het altijd mogelijk om met je te discussiëren over de klinische waarde van de gevonden resultaten. Ik ben dan ook heel blij dat je hier als mijn paranimf naast mij wil staan.

Anne Eskes: samen met jou een onderzoek doen was ‘wervelwinderig’. Het was fijn dat je mijn paranimf wil zijn en dat we de promotieperielen samen konden delen. Ik hoop dan ook over een paar maanden bij jouw promotie aanwezig te zijn. Verder vond ik het erg leuk dat we samen naar Australië gingen en de voorpret voor deze reis gaf ook de broodnodige afleiding.

Voor zover nog niet benoemd, wil ik voor de onderzoeken die beschreven staan in dit boekje heel veel mensen danken voor de prettige samenwerking. Ook aan jullie is het te danken dat dit boekje er is gekomen. Voor de ‘Smoke’ trial waren dat Prof.dr. Hanneke de Haes, Dr. Ellen Smets, Dr. Loes Wiggers, Prof.dr. Ron Peters, en Lucas van Loenen. Voor de ‘Triage’ studie: Erik Schinkel, Rob Köhlinger, Daniëlle Schutte, Johan Kappelhof, Vanessa Chin a Choi, Nienke van Lammeren, Maaike van der Kamp, en Dr. Remco Rosmulder. Voor de ‘Temp’ studie en ‘Vital signs’: Dr. Astrid Goossens, Anouk Knops, Lotte Verweij, Jeroen Ludikhuize, Faridi van Etten-Jamaludin, en Prof.dr. Piet Bakker. Voor de ‘Zilver’ review: Jarne van Hattem, Niels Bos, Prof.dr. Nicky Cullum, Sally Bell-Syer, en Ruth Foxlee.

Niet alle studies waarbij ik betrokken ben geweest als auteur zijn in dit boekje opgenomen. Ook deze mensen wil ik graag danken voor het vertrouwen die ze in mij stelden en de bijdrage die ze hebben geleverd aan mijn wetenschappelijke ontwikkeling. Een persoon wil ik hierbij in het bijzonder noemen: Dr. Liselot Hoornweg, het was heel fijn om met je samen te werken aan de AJAX trial en ‘uiteraard’ aan een review.

Als afsluiting van de Master opleiding EBP moest ik een wetenschappelijk artikel schrijven. Hierbij moet de naam van Dr. Rien de Vos genoemd worden die mij begeleidde. Je leerde me heel veel, maar pas veel later vielen al je opmerkingen op zijn plek. Daarnaast heeft Daphne Lees verschillende artikelen van mij geredigeerd en ik wil haar danken dat ze van mijn Dunglish altijd iets moois wist te maken.

Ook dank aan alle hoofdverpleegkundigen, verpleegkundigen en artsen van de chirurgische afdelingen, spoedeisende hulp en poliklinieken, waar ik de ruimte kreeg om een aantal onderzoeken uit te kunnen voeren. Echter deze onderzoeken waren nooit mogelijk geweest zonder de medewerking van patiënten, dus ook hen wil ik danken.

De eerste jaren als onderzoeker heb ik de kamer gedeeld met de transplantatie-coördinatoren. Fred, José en Janneke, dank voor de gezelligheid, maar ook de vele discussies over de zin van het leven. Onze kamer was soms overvol, zeker toen Hester er ook nog zat met alle studenten en jullie met jullie collega’s, maar het was een fijne en waardevolle tijd en ik denk er met plezier aan terug.

Ondanks dat mijn functie veranderde, kreeg ik van Betty Schinkel de ruimte en de mogelijkheid om door te gaan met mijn promotie. Dank voor je interesse en je vertrouwen in mij. Een nieuwe functie hield ook nieuwe kamergenoten in. Daisy, ik vind het jammer dat je niet bij mijn promotie aanwezig kunt zijn, want je bent naar India vertrokken. Dank voor de gezelligheid en het af en toe een hart onder de riem steken. Ik hoop dat je het komende jaar een mooie tijd hebt in India waar je later met plezier terug kunt kijken.

Ingrid, onderzoek doen was niet jouw ding, maar wel de mijne. Onze ‘krachten’ vulden elkaar mooi aan en ik hoop nog lang met je samen te mogen werken. Verder wil ik vooral de stafmedewerkers Kwaliteit en Zorg van het ‘eerste uur’: Jolanda, Lydia en Jonne danken voor het feit dat ze ieder op hun eigen manier er waren voor mij. Jonne, ik vind het zo lief van je dat je me hebt geholpen bij het ontwerpen van de cover. Dank je wel.

Natuurlijk wil ik ook mijn familie en vrienden bedanken. Jullie interesse en vertrouwen in mij zijn belangrijk geweest om dit proefschrift af te ronden. Lieve Ineke, als ‘grote’ zus, zat je me vroeger altijd achter mijn vodden om iets af te maken, voor ik aan iets nieuws begon. NU, HET PROEFSCHRIFT IS AF! Roedi, jij inspireerde me op het juiste moment door het sturen van een foto. Lieve schoonzussen en zwagers. Ik kan me geen betere schoonfamilie wensen, en vind het heel bijzonder dat jullie blijven, want je bent altijd achter mijn vodden om iets af te maken, voor ik aan iets nieuws begon.

Lieve papa en mama, door jullie ben ik wie ik ben. Ik dank jullie voor de mogelijkheden die jullie hebben gegeven om mij verder te kunnen ontwikkelen. Wat zouden jullie vandaag trots geweest zijn. Ik zal altijd aan jullie denken.


Lieve Nico, dank voor de vanzelfsprekendheid en liefde waarmee je mij altijd onvoorwaardelijk steunde om mezelf te ontwikkelen. Soms ging dat ten koste van je eigen hobby ‘vliegen’, omdat je de laatste tijd heel veel thuiszaken overnam. Zonder jouw steun en liefde was het dan ook nooit gelukt. Ik hoop de toekomst nog heel lang met je te mogen delen.
CURRICULUM VITAE


Als verpleegkundig onderzoeker werkte ze mee aan richtlijnenstudies: de TEMP studie, de Laxeerstudie en het postoperatieve gebruik van een maagsonde, met als doel om lokale Evidence Based aanbevelingen te formuleren. Ze deed ook onderzoek naar triagesystemen en zilververbanden. Daarnaast werkte ze mee aan de Smoke Trial, AJAX trial en PSS hip trial.

Op dit moment is ze werkzaam als stafmedewerker Kwaliteit en Zorg met als aandachtsgebied Evidence Based Practice en indicatoren.

Marja is getrouwd met Nico en ze hebben samen een tweeling: Philip-Jan en Jasperina.