Aspects of retrospective record review

– A matter of patient safety
ASPECTS OF RETROSPECTIVE RECORD REVIEW
– A matter of patient safety

KIRSTINA SCHILDMEIJER

LINNAEUS UNIVERSITY PRESS
Abstract


Background: Health Care is of great value but despite increased efforts to improve patient safety, many patients still suffer harm caused by healthcare, and even more patients have encountered incidents that could have caused harm. Adverse events can be detected by using retrospective record review. The Harvard Medical Practice Study and the Global Trigger Tool are such methods. Retrospective record review has shown better coverage than the commonly used clinical incident reporting system to identify patient safety information.

Aims: The general aim of the thesis was to evaluate, compare and expand retrospective record review methods for clinical use in health care. The specific aims were:

- To evaluate the agreement in judgments of adverse events between well-trained Global Trigger Tool teams from different hospitals.
- To describe strengths and weaknesses, from team members perspectives of working with the Global Trigger Tool method of retrospective record review to identify adverse events causing patient harm.
- To evaluate the feasibility and capability of two common retrospective record review methods, the "Harvard Medical Practice Study" method and the "Global Trigger Tool" method in detecting adverse events in adult orthopedic inpatients.
- To evaluate retrospective record review for the detection and characterization of no-harm incidents and compare findings with conventional incident reporting systems.

Methods: A random sample of fifty patient records was reviewed by a team from each of five hospitals according to the Global Trigger Tool method (I). The teams were interviewed in focus groups concerning their experiences of the Global Trigger Tool method (II). In papers III and IV, a random sample of 350 or 100 orthopaedic admissions was examined. The outcomes from the Harvard Medical Practice Study and the Global Trigger Tool methods were compared (III). In paper IV the Harvard Medical Practice Study method was also used for identifying no harm incidents.

Results: The number of identified adverse events differed between the teams, corresponding to a level of adverse events ranging from 27.2 to 99.7 per 1000 hospital days. Differences were also found in the assessment of level of harm and judgment of preventability. Four of the teams made similar assessments while the fifth identified three times as many adverse events compared to the other teams (I).

Eight categories with their strengths and weaknesses emerged from the focus group interviews. The team members were constant in their generally positive experiences of the Global Trigger Tool method, even if the teams over time altered the application of the method (II).

With the Harvard Medical Practice Study and the Global Trigger Tool methods combined, 160 adverse events were identified in 105 (30%) of the 350 records. The Harvard Medical Practice Study method identified 155 (97%) adverse events in 104 of 350 records compared with 137 (86%) adverse events in 98 records using the Global Trigger Tool method. The adverse events causing the greatest differences were the ones causing minimal or moderate impairment (III).

In paper IV, results showed that 118 no harm incidents were detected in 91 (26%) of the 350 patient records. Ninety-four of the 118 (80%) no-harm incidents were classified as preventable. Sixteen no-harm incidents were identified by the five conventional incident reporting systems. Of these, ten no-harm incidents were also found by the Harvard Medical Practice Study method.

Conclusions: Retrospective record review enabled detection of adverse events as well as identification of no harm incidents. It is considered a useful method. There were differences both in agreement between reviewer teams and between review methods. Joint preparations and discussions seemed to increase the level of agreement in judgment between reviewers. By adding retrospective record review for findings of no-harm incidents to conventional incident reporting, healthcare providers can gain new important information about commonly occurring, no-harm incidents in order to improve patient safety.
Abstract


**Background:** Health Care is of great value but despite increased efforts to improve patient safety, many patients still suffer harm caused by healthcare, and even more patients have encountered incidents that could have caused harm. Adverse events can be detected by using retrospective record review. The Harvard Medical Practice Study and the Global Trigger Tool are such methods. Retrospective record review has shown better coverage than the commonly used clinical incident reporting system to identify patient safety information.

**Aims:** The general aim of the thesis was to evaluate, compare and expand retrospective record review methods for clinical use in health care. The specific aims were;
- To evaluate the agreement in judgments of adverse events between well-trained Global Trigger Tool teams from different hospitals.
- To describe strengths and weaknesses, from team members perspectives of working with the Global Trigger Tool method of retrospective record review to identify adverse events causing patient harm.
- To evaluate the feasibility and capability of two common retrospective record review methods, the "Harvard Medical Practice Study" method and the "Global Trigger Tool" in detecting adverse events in adult orthopedic inpatients.
- To evaluate retrospective record review for the detection and characterization of no-harm incidents and compare findings with conventional incident reporting systems.

**Methods:** A random sample of fifty patient records was reviewed by a team from each of five hospitals according to the Global Trigger Tool method (I). The teams were interviewed in focus groups concerning their experiences of the Global Trigger Tool method (II). In papers III and IV, a random sample of 350 orthopaedic admissions was examined. The outcomes from the Harvard Medical Practice Study and the Global Trigger Tool methods were compared (III). In paper IV the Harvard Medical Practice Study method was also used for identifying no harm incidents.

**Results:** The number of identified adverse events differed between the teams, corresponding to a level of adverse events ranging from 27.2 to 99.7 per 1000 hospital days. Differences were also found in the assessment of level of harm and judgment of preventability. Four of the teams made similar assessments while the fifth identified three times as many adverse events compared to the other teams (I).

Eight categories with their strengths and weaknesses emerged from the focus group interviews. The team members were constant in their generally positive experiences of the Global Trigger Tool method, even if the teams over time altered the application of the method (II).

With the Harvard Medical Practice Study and the Global Trigger Tool methods combined, 160 adverse events were identified in 105 (30%) of the 350 records. The Harvard Medical Practice Study method identified 155 (97%) adverse events in 104 of 350 records compared with 137 (86%) adverse events in 98 records using the Global Trigger Tool method. The adverse events causing the greatest differences were the ones causing minimal or moderate impairment (III).

In paper IV, results showed that 118 no harm incidents were detected in 91 (26%) of the 350 patient records. Ninety-four of the 118 (80%) no-harm incidents were classified as preventable. Sixteen no-harm incidents were identified by the five conventional incident reporting systems. Of these, ten no-harm incidents were also found by the Harvard Medical Practice Study method.

**Conclusions:** Retrospective record review enabled detection of adverse events as well as identification of no harm incidents. It is considered a useful method. There were differences both in agreement between reviewer teams and between review methods. Joint preparations and discussions seemed to increase the level of agreement in judgment between reviewers. By adding retrospective record review for findings of no-harm incidents to conventional incident reporting, healthcare providers can gain new important information about commonly occurring, no-harm incidents in order to improve patient safety.
It may seem a strange principle to enunciate as the very first requirement in a hospital that it should do the sick no harm

(Florence Nightingale, 1855)

To my family
“It may seem a strange principle to enunciate as the very first requirement in a hospital that it should do the sick no harm”

(Florence Nightingale, 1855)

To my family
The thesis is based on the following original papers, which are referred to in the text by their Roman numerals.


List of papers

The thesis is based on the following original papers, which are referred to in the text by their Roman numerals.

I. Schildmeijer K, Nilsson L, Årestedt K, Perk J.

II. Schildmeijer K, Nilsson L, Perk J, Årestedt K, Nilsson G.


IV. Schildmeijer K, Muren O, Perk J, Pukk Härenstam K, Unbeck M, Nilsson L.
Abbreviations

AE  Adverse Event
GTT  Global Trigger Tool
HFMEA Health Care Failure Mode and Effect Analysis
HMPS  Harvard Medical Practice Study
IHI Institute for Healthcare Improvement
IVO The Health and Social Care Inspectorate (Swe: Inspektionen för Vård och Omsorg)
LÖF The Patient Insurance LÖF (Swe: Patientförsäkringen LÖF)
MBRR Marker Based Record Review (Swe: Markörbaserad journalgranskning)
NCC MERP National Coordinating Council for Medication Error Reporting and Prevention
PDSA Plan-Do-Study-Act model
RN  Registered Nurse
RRR  Retrospective Record Review
SKL Swedish Association of Local Authorities and Regions (Swe: Sveriges Kommuner och Landsting)
Content

Introduction ................................................................. 2
The scope of the problem ........................................ 2
Patient safety ......................................................... 3
Laws and regulations in the area of patient safety ......... 4
Stakeholders in assuring patient safety ...................... 4
  The patient’s role .................................................. 4
  The role of the healthcare staff ............................. 5
  The role of hospital management ......................... 6
  The role of society .............................................. 6
  Knowledge of improvement .................................. 7
Patient safety methods .............................................. 8
Clinical incident reporting ....................................... 9
Patient and relative incident reporting ................. 9
Retrospective record review .................................. 9
The Global Trigger Tool method ............................ 10
The Harvard Medical Practice Study method .......... 12
The Wimmera Clinical Risk Management Model ....... 13
Rational .................................................................. 13
Aims .................................................................... 15
  Specific aims ..................................................... 15
Methodology ............................................................ 16
  Design ............................................................... 16
  Setting ............................................................... 16
  Sample and inclusion criteria ............................. 16
  Procedure and data collection ............................ 17
Data analysis ............................................................ 19
  Statistical analysis .......................................... 19
  Qualitative analysis .......................................... 20
Ethical considerations .............................................. 21
Results ...................................................................... 23
  Paper I. Lack of consistency between reviewer teams using the Global Trigger Tool method .... 23
  Paper II. Experiences of strengths and weakness of the Global Trigger Tool method .... 24
  Paper III. Evaluating the Harvard Medical Practice Study method and the Global Trigger Tool method .... 24
  Paper IV. Finding no-harm incidents using the Harvard Medical Practice Study method .... 25
Discussion ............................................................... 27
  Discussion of findings ........................................ 27
  Methodological considerations .......................... 31
Conclusions .............................................................. 35
  Clinical Implications ....................................... 36
  Future research ................................................. 37
Populärvetenskaplig sammanfattning ......................... 38
Acknowledgments .................................................. 42
References ............................................................. 44
Appendix ................................................................... 55
  Table 1. The Global Trigger Tool method triggers used in papers I - III (Swedish version with 53 triggers) ................................................................................................................. 55
  Table 2. The screening criteria from Harvard Medical Practice Study method used in papers III - IV ............................................................................................................. 57
**Introduction**

This thesis was written in the context of caring science. Caring science is profession neutral, and its primary mission is to assist patients in becoming as healthy as possible. The responsibility for the care provided and care relationship is always the caregiver’s. Therefore, it is of the greatest importance to provide care enabling patients to feel safe and secure within the healthcare system. As a part of caring science, practical design implies that routines improve, whereby yesterday’s good performance may constitute today’s malpractice (1). The hope of the author of this thesis is that it will be read not only by healthcare staff members and hospital managers but by policy makers in the healthcare system as well.

**The scope of the problem**

Safety critical industries where catastrophic outcomes are a factor, such as in the aviation industry, are required to provide evidence of safety before undertaking hazardous operations. Approaches to safety are required to be organization-wide and referred to as “Safety management systems”, which include safety policies, safety plans, and a feedback loop to improve safety performance communicated to all relevant staff members (2). Healthcare is also a safety critical sector where outcomes can be devastating to those affected (3). Despite increased efforts to improve patient safety, many patients still suffer harm caused by healthcare (4, 5). The Institute of Medicine’s report “To Err is human” in 1999 indicated large problems in the health care system (4). Based on numbers of adverse events (AEs) detected in two studies, the Harvard Medical Practice Study (6) and the Utah and Colorado study (7), calculations were made resulting in between 44 000 and 98 000 patients in the United States dying every year from harm caused by healthcare. Results were extrapolated from the 33.6 million United States patients receiving hospital care in 1997 (4). From Swedish conditions during the years 2003-2004, Soop et al. (5) found that if the incidence of AEs in Swedish hospitals were extrapolated results corresponded to 105 000 AEs judged as preventable, and 630 000 extra days of hospitalization per year. Even if these results are somewhat dated, later research has shown that patient safety is a persistent problem. Baines et al. (8) found that the risk of experiencing a preventable AE was higher in 2008 than in 2004. Based on an extrapolation from studies published during 2008-2011, James (9) calculated the number of AEs leading to patient death caused by healthcare to at least 210 000 in the United States yearly.
Patient safety

The patient safety area uses different terms, sometimes for the same events. To clarify this, the World Health Organization (WHO) published a report in 2009 with definitions related to patient safety. Some are presented in Table 1 (10).

Table 1. Patient safety terms defined by WHO

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk</td>
<td>The probability than an incident will occur</td>
</tr>
<tr>
<td>Near miss</td>
<td>An incident which did not reach the patient</td>
</tr>
<tr>
<td>No-harm incident</td>
<td>An event reached a patient but did not result in discernible harm</td>
</tr>
<tr>
<td>Incident</td>
<td>An event or circumstance that could have resulted, or did result, in unnecessary harm to a patient</td>
</tr>
<tr>
<td>Error</td>
<td>A failure to carry out a planned action as intended or application of an incorrect plan</td>
</tr>
<tr>
<td>Adverse event</td>
<td>An incident that results in harm to a patient</td>
</tr>
<tr>
<td>Harm</td>
<td>Impairment of structure or function of the body and/or any deleterious effect arising there from, including disease, injury, suffering, disability and death, and may be physical, social or psychological</td>
</tr>
<tr>
<td>Healthcare associated harm</td>
<td>Harm arising from or associated with plans or actions taken during the provision of healthcare, rather than an underlying disease or injury</td>
</tr>
<tr>
<td>Patient safety</td>
<td>Patient safety is the reduction of risk of unnecessary harm associated with healthcare to an acceptable minimum</td>
</tr>
</tbody>
</table>

Patient safety covers both the absence of harm to patients and the actions taken to prevent harm. Healthcare staff workers struggle daily to make healthcare function well. At times they must also take extra measures to avoid causing harm when discovering that the standard of care is insufficient. Patient safety also includes structuring the care processes to eliminate known causes of errors (11).

The healthcare system has become more and more complex and extensive organizational changes are continuing processes (12-15). Human errors can be seen either as created by individuals or as a symptom of organizational problems (16, 17). The healthcare staff is responsible and connected to all levels of the healthcare system and is influenced by patients, teamwork, the environment, and the organizational context including hospital management (17-19). Therefore, errors must be seen in relation to the context where the healthcare staff is working (17, 20). During recent years great attention has been paid to patient safety and quality of care. This has led to discussions about the reasons for error, about an individual or system based view, and to the introduction of many patient safety tools and data collecting methods, all intended to reduce patient harm (11, 16, 21-23).
European studies report a prevalence of AEs in inpatient care between 9-12% (5, 24, 25). A review of conditions in Sweden of hospital admissions during 2003-2004, showed that 12.3% of patients receiving hospital care suffered an AE. On average, every patient suffering from harm had an increased length of hospital stay of six days (5). During 2012, all Swedish hospitals reviewed 20 randomly selected records per month (26) from January-March and reported their AEs to the Swedish Association of Local Authorities and Regions (SKL) (n= 3 900 records). In 14% of the records, an AE was discovered (27). In both Swedish investigations, a majority of these AEs were considered preventable. Healthcare associated infections were the most common AEs (5, 27).

Laws and regulations in the area of patient safety

According to “Act of healthcare” in the Swedish Statute Book (HSL 1982:763) from 1982, healthcare must be safe, and of high quality (28). To further increase patient safety the “Act of patient safety” (SFS 2010:659) has been added (29). It emphasizes all healthcare staff liability to undertake systematic proactive patient safety work and report events that have or could have caused patient harm, according to the “Lex Maria statute” (SOSFS 2005:28) (30). Healthcare managers should also provide information to the healthcare staff concerning investigation, notification and the decision from the “Inspectorate for healthcare” (IVO) and inform the patient of the notification and outcome of the decision of IVO (31).

Stakeholders in assuring patient safety

Healthcare concerns everyone, whether as patient, relative, healthcare staff member, administrator or citizen. It is not only important to provide safe care to patients, but for all involved, although all are affected differently. Thus, patient safety can be seen from different perspectives.

The patient's role

Patients and their relatives are the most attentive observers of care, and are those most likely to notice whether correct treatment has been administered. To obtain their feedback is therefore important. This can be difficult in some cases. Patients may be too ill, confused, or afraid to confront the healthcare staff, which can be experienced as uncomfortable (32-34). Despite this, studies have shown patients capable of reporting both AEs and no-harm incidents (34-36). Most events reported by patients are drug related, along with difficulty breathing and excessive pain (34, 36-38).
Healthcare should meet the requirements for providing good health, implying meeting patient needs for security in care and treatment (28). This is intended to establish a trusting relationship, which can emerge through the healthcare staff’s openness, commitment, trust and reliability (39, 40). Most people have great trust in healthcare (41, 42). Patients believe that when sick and vulnerable, healthcare staff members will look out for their interests and thus seek care with trust (43-45). Even in high risk situations such as in chemotherapy, patients maintain trust in patient safety effectiveness (46). A majority of patients feel protected from AEs, but those who have experienced an AE during their hospital stay, maintain a gloomier memory of their entire hospital experience (41), leading to less trust and confidence in the healthcare staff’s competence (47). Patients suffering an AE need opportunities to not only talk about the technical-practical dimensions, but also about personal ones (48). Inadequate responses from the healthcare staff adds to distrust. This affects the patient’s sense of safety, loss of psychological wellbeing, and undermines their trust and confidence in professional care (49). This may lead to poor communication between patient and staff, where the patient becomes reluctant to provide important information about their wellbeing, subsequently affecting the patient’s safety (50).

The role of the healthcare staff

The most cited principle of medical ethics is “Above all, do no harm” (51). This may have created a blame culture where mistakes are hard to deal with, as staffs perceive errors as intolerable (52). Everyone commits errors and misjudgments, and healthcare staff members are no exception to this rule. In addition, the complexity of today’s care must also be considered. Patients are treated by many such as physicians, registered nurses (RNs) and therapists whose perspective and priorities may differ both internally and from patients. RNs and physicians are key groups of healthcare providers. They handle hundreds or perhaps thousands of handoffs, memos, reports and notes (11). The RN’s role requires professionalism in order to provide good care in a number of situations. In nursing the RN must identify patient needs. RNs are responsible for implementing much hands-on care, administer vast amounts of medication, and monitor patients for complications and untoward reactions to treatment. They also provide continuous supervision to patients at risk for healthcare associated infections, falls and medication errors, etc. (11, 53, 54). Surgery, for example, causing injured nerves and blood vessels made by physicians, is also risk filled and AEs have commonly occurred (55-58).

Risk factors are always present and the healthcare staff may be the difference between success and failure (16, 59). Whereby RNs and physicians care directly for patients within the healthcare system’s complex structure, they are often at the “sharp end” of error as the last barrier between care and patient (60-63). Unfortunately,
these errors may lead to considerable consequences. RNs and physicians tend to see error more as personal fault rather than a fault of the system (22, 64). The event remains in the mind of the involved healthcare staff for many years affecting their personal and professional lives, thus increasing the risk for poor clinical performance (22). Studies have shown that they become second victims and suffer mental scars (14, 65, 66). Even colleagues and healthcare managers can be of the opinion that AEs are due to personal errors. In the complex healthcare of today, work situations are sometimes unspecified and unpredictable. The work of the healthcare staff often differs from what they have been trained for, requiring adjustments to existing resources and requirements. Errors should be regarded more as results of unexpected combinations of actions rather than failures, i.e. a system thinking (67).

**The role of hospital management**

For achieving care improvement, hospital management requires knowledge of quality issues such as patient safety tools and data collection. Time must also be allocated, as well as money and effort. Involvement by the hospital management in minimizing errors reduces risk while increasing risk awareness, thus preventing errors (68). Hospital managers and staff may recognize that safety issues have often common underlying causes and attempt to strike a balance between error-specific solutions and broader system improvements (69). Opportunities to improve care in the future involve continuous feedback and learning from errors in a no-blame environment (70-75).

**The role of society**

In addition to human suffering, patient harm also contributes to substantial economic costs to society. These costs include healthcare costs, costs to counties, lost income, and production loss. The direct costs to society in 2000 for additional bed days in the United Kingdom were calculated at 2 billion pounds (76). The same year the costs for preventable harm were calculated at 17-29 billion dollars in the United States (4). In a Danish study it was calculated that 10% of economic resources for healthcare was used for treatment and consequences to patients harmed by healthcare (24). Even if these calculations were several years old they indicate a current problem of major proportions. Jha et al. (77) investigated unsafe care and measured disability-adjusted life years lost. They estimated 421 million worldwide hospitalizations and approximately 42.7 million AEs. These AEs resulted in 23 million disability-adjusted life years lost in 2009, showing that unsafe care remains a major problem. An indication of the conditions in Sweden can be obtained by looking at The Patient Insurance LÖF (LÖF). LÖF certifies county and regional accountability to patients harmed in connection with healthcare. Compensation is provided if harm is considered preventable. During the year 2012, 12 900 notifications were reported. This was an increase of 9% compared to 2011. Of those who had sought compensation, patients or
relatives were compensated in 39% of the cases amounting to 477 million Swedish crowns. Examples of compensated cases during and after care were; infections, injured blood vessels and nerves related to surgery. Other common reasons for economic compensation were delayed or incorrect diagnosis (78).

Knowledge of improvement

All members of the healthcare system should strive to promote, protect or restore patient health, which requires will, imagination, and implementation. For this goal all within the healthcare system must realize two tasks: to provide healthcare and improve quality and safety (79). This is particularly expressed by the Swedish Statute Book in “Act of patient safety” (SFS 2010:659) (29), of the Swedish Society of Nursing (SSF) in its “Strategy for developing healthcare” (80) and in the “Description of RN competence” by the National Board of Health and Welfare (81).

All work can be seen as processes within a system. A process is a set of conditions and causes where steps are transferred during input and output by repeatedly coming together. Conditions and causes can be measured and collected. With the data collected, changes can be developed and tested (82). For improvement, knowledge is necessary. Within the healthcare system improvement is depended on knowledge of the subjects and discipline, i.e. professional knowledge, but also knowledge of improvement itself. Improvement knowledge consists of; systemic knowledge, knowledge of variation, knowledge of psychology, and knowledge theory. What is made, created or produced in the system leads to knowledge of the system. The understanding of variation over time is a key to learning, as well as is the utilization of the differences for improvement. Knowledge of psychology includes the knowledge of human reactions to change within a system and to a system. The knowledge of psychology is valuable whereas people respond differently to changes. The possibility for learning and building knowledge is created when theory and action are linked together (83). The link between professional and improvement knowledge is shown in Figure 1.
All improvement involves changes and should be a daily part of all work, in all parts of a system. Assessment of whether changes lead to improvement must include the measurement and analysis of patterns (84).

**Patient safety methods**

There are different methods for increasing patient safety. They can be broadly categorized into safety tools and data collection methods. Thus some examples have been selected to describe the patient safety area. Patient safety tools are used directly in healthcare practice. Examples of patient safety tools are S-BAR (85–88) and checklists (89–101) which may directly affect healthcare safety. Data collecting methods are important in indicating problem areas in the healthcare system. Examples of data collecting methods are WalkRounds (102, 103), Healthcare Failure Mode and Effect Analysis (HFMEA) (104–108), root cause analysis (29, 104, 105, 109–112), patient safety culture measurements (113–116), mortality and morbidity conferences (117–120) and quality registers (87, 121, 122).
Clinical incident reporting

The most common data collecting method within the healthcare system is clinical incident reporting. Incidents and risks are reported by some, while others focus only on severe incidents (22). Although described as an important component of systematic patient safety work, only a minority of AEs and no-harm incidents are reported in clinical incident reporting systems (123-130). Reasons for not reporting can vary, but the culture of blame may have an important impact (131). Other factors that may contribute to low reporting are problems involved in reporting, lack of time, lack of knowledge, lack of feedback, fear of disciplinary actions and litigation (22, 70-72, 75). There have been many suggestions for improving reporting systems (132, 133). The most important task is to make the reporting system based on system-oriented changes rather than individualistic target reforms (131).

Patient and relative incident reporting

Patients and relatives can also report incidents. This can be done by contacting clinics directly or via a Patient Mediator. The patient advisory committee could be contacted if patients or relatives found it difficult to contact or address complaints themselves (134). Complaints or notifications are reported according to the “Lex Maria statute” (31). Complaints can also be made to the Patient Insurance LÖF (135).

Retrospective record review

By measurement and intervention, it is possible to gain perspective on areas in need of improvement. This can be done through the identification of AEs by structured record review. Structured record review is often performed on randomly selected admissions, and can be carried out either prospectively or retrospectively. The retrospective record review method (RRR) has been shown to identify substantially more incidents compared to other methods (124, 127, 129, 130, 136–140).

The review process within the RRR method is performed by teams usually consisting of two primarily reviewers (RNs), a second reviewer (physician), and undertaken two stages. The first stage is carried out by the RNs following the RRR method. They screen records looking for triggers, e.g. “Pressure ulcer” (141) or for criteria, e.g. “Unplanned transfer from general to intensive care” (5). Some of the triggers/criteria are AEs per definition, e.g. “Healthcare associated infection”, while others are seen more as events that require attention, e.g. “Blood transfusion”. The differences between triggers and criteria could be described by triggers being more explicit, e.g. “Flumazinil”, while criteria are mostly implicit, e.g. “Adverse drug reaction”. For every trigger/criterion detected, a judgment is made by the nurses regarding whether the trigger/criterion reflect the presence of a potential AE or not. Records with potential AEs are then forwarded to the physicians for further
review, stage two. In stage two, the physician judges whether patient harm has occurred or not, based on findings in stage one.

There are three main types of RRR methods, the Global Trigger Tool method (GTT), the Harvard Medical Practice Study method (HMPS) and the Wimmera Clinical Risk Management Model. All methods focus on identifying patient harm, AEs. Since the focus of the current thesis is based on the GTT and the HMPS methods, the Wimmera Risk Management Model is only described briefly.

**The Global Trigger Tool method**

The GTT method has become an increasingly common RRR method in clinical work. The Institute of Healthcare Improvement (IHI) from Boston developed the GTT method in 2004 (142). GTT is a two stage review method for identifying patient harm. Through a retrospective review of randomly selected patient records the reviewers systematically look for key information, so called triggers (events) that may indicate possible patient harm. The GTT method is built on 54 primarily explicit triggers. The triggers are divided into six modules; care, medication, surgical, intensive, perinatal and emergency (141). These have been translated and adapted to Swedish conditions. The triggers “restraint use” in the care category and “other” in the medication category have been excluded. In the surgical category a new trigger, “occurrence of any postoperative complication”, has been added, resulting in a Swedish version consisting of 53 triggers (Appendix Table 1) (26). Specific activity versions have also been developed based on the original IHI method, e.g. versions for patients in intensive care (143), neonatal care (144), surgical care (137) and ambulatory surgery (145).

The GTT method is designed to utilize small samples over time. It recommends selecting a sample of 10 records every two weeks, i.e. 20 records monthly. Organizations with large resources could choose to select 40 records. Larger samples only provide small additional benefit. Records chosen for review should include complete records of patients 18 or older with a hospital stay of at least 24 hours. The inclusion period before and after the index admission is set at 30 days. Inpatients from rehabilitation and psychiatric wards shall be excluded whereby the GTT method triggers are not defined for this population. A time limit of 20 minutes per record is set for stage one because the intention of the GTT method is to provide useful information about trends and special causes suitable for improvement, but not for records to be read cover to cover. For the second stage, no time limit is set. The GTT method only includes commissions while omissions are negated as AEs (141).
The reviewer team should consist of at least three people, usually two RNs and a physician. Each RN reviews the patient record marking triggers in the GTT method worksheet, one per patient. If the RN finds that the marked trigger is associated with a potential AE, she/he describes and categorizes the harm. After completing the review, the two RNs compare results, come to a consensus, and fill in a new worksheet. Records with potential AEs are then given to the physician (141).

The physician does not examine the patient records, but assesses the two RNs' mutual decision in the second stage. Assessment as to whether an AE has occurred should be based on the patient’s perspective. The patients underlying disease/condition is not including in the AEs (26). The physician categorizes the harm adapted from the scale from the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) (146) and use the category E-I, i.e. “Temporary harm to the patient and required intervention” to “Patient death”. Category A-D is excluded as they do not cause harm (Table 2) (141).

Table 2. Severity scale used in the GTT method

<table>
<thead>
<tr>
<th>Categories for judgment of level of harm</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Circumstances or events that could lead to errors</td>
</tr>
<tr>
<td>B An error not affecting a patient</td>
</tr>
<tr>
<td>C An error occurred but not leading to harm</td>
</tr>
<tr>
<td>D An error occurred and led to a control or some kind of measure to ensure the patient was not injured</td>
</tr>
<tr>
<td>E Temporary harm to the patient and required intervention</td>
</tr>
<tr>
<td>F Temporary harm to the patient and required initial or prolonged hospitalization</td>
</tr>
<tr>
<td>G Permanent patient harm</td>
</tr>
<tr>
<td>H Intervention requiring life sustenance</td>
</tr>
<tr>
<td>I Patient death</td>
</tr>
</tbody>
</table>

The IHI suggests three ways to present the data collected. This can be done by run charts showing AEs per 1000 patient days, AEs per 100 admissions or as a percent of admissions with an AE. The review teams can, in addition to the run chart, present the volume of each category of harm. Events occurring prior to admission and present on arrival to the hospital could be included in a separate category. When prioritizing for improvement work, data can also be presented by type of AE (141).

In addition to Swedish conditions, a further aspect has been added to the GTT method, preventability. This is also judged by physicians. They determine if the AE is preventable on a scale from 1-6, from “Little to no evidence” to “Virtually certain evidence” that the AE was possible to prevent. The perception in the Swedish GTT handbook is that it is primarily preventable harm that one should focus on in the improvement of patient safety work. The preventability scale for the GTT method is the same used in developed versions of the method used in the HMPS (Table 3) (26). From the experiences of working with the GTT method in
Sweden, the need of revision and development of the Swedish GTT handbook from 2008 was emphasized. This was done during 2012. The intention was to provide a background for working with the GTT method, both on hospital and clinical levels, and provide clearer guidelines for judgment of harm and preventability. The new Swedish GTT handbook is called “Marker Based Record Review” (MBRR) (147).

**Table 3. Causation and preventability scale used in both the GTT and the HMPS methods**

<table>
<thead>
<tr>
<th>Categories for judgment of level of harm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>

The Harvard Medical Practice Study method

The HMPS method has been used in a number of large nationwide studies (5, 24, 25). The method is based on the California medical insurance feasibility study made in 1977 where a random sample of patients and hospitals offered one of the first large sample estimates of AEs (148). To investigate if the results from the HMPS method could be transferred to other states, the Utah and Colorado study was made with 15 000 discharged patient records during 1992 (149). The method is based on two (148, 150), or three (5, 24) stages of record reviews with no time limit. Both omissions and commissions are included in the review. In this thesis, the name HMPS is used for all studies based on the criteria.

To be included for record review the patient should be an adult inpatient. Psychiatric care and rehabilitation is often excluded from review. Stage one is often carried out by RNs, screening the patients records comprehensively for 18 screening criteria (Appendix Table 2). For every screening criterion detected, a judgment is made by the RNs as to whether it reflects the presence of a potential AE or not. Records with potential AEs are then handed over to a physician (150, 151). In the second stage the physicians confirm the AE and assess causation on a scale from 1–6. A score of four or higher is regarded as an AE. A similar scale is also used by the physicians for judging if the AE is caused by negligence in healthcare. (Table 3) (5, 25).
Severity of the AE is judged by the physician on a seven point scale (Table 4). (148).

### Table 4. Severity scale used in the HMPS method

<table>
<thead>
<tr>
<th>Severity scale</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Minimal impairment, recovery within 1 month</td>
</tr>
<tr>
<td>2</td>
<td>Moderate impairment, recovery within 1-6 months</td>
</tr>
<tr>
<td>3</td>
<td>Moderate impairment, recovery within 6-12 months</td>
</tr>
<tr>
<td>4</td>
<td>Permanent impairment, degree of disability less than 50%</td>
</tr>
<tr>
<td>5</td>
<td>Permanent impairment, degree of disability more than 50%</td>
</tr>
<tr>
<td>6</td>
<td>Contributed to patient death</td>
</tr>
<tr>
<td>7</td>
<td>Unable to determine</td>
</tr>
</tbody>
</table>

The Wimmera Clinical Risk Management Model

The Wimmera Clinical Risk Management Model is an RRR method for inpatient care and consists of eight general criteria (later a ninth criteria was added). Records of all patients discharged from the hospital are screened by a non-medical staff member. Records found with at least one criterion are then forwarded to a physician to determine if an AE is present. A multi professional surveillance committee then meets, discusses findings and assesses e.g. severity, from 1-7, i.e. from “Minor severity” to “Death” (152). AEs are also analyzed and ranked in order of risk severity. To prevent the event from recurrence, actions are then planned and implemented. Examples of effective actions are simplifying systems, standardizing procedures, using reminders and checklists, providing timely information and small-group interactive education. (153).

Rational

Despite efforts, AEs are still common in healthcare. It is of importance to reduce patient harm, not least because of the suffering of patients and their relatives. To become aware of incidents is of importance, as it is a first step in preventing them. Different methods have been developed in order to prevent patient harm. One of these methods is RRR. The RRR method has been found to identify substantially more incidents compared to other methods. Of the RRRs, the GTT and HMPS are the most common. The GTT method randomizes small samples over time and has 53 mostly explicit triggers while the HMPS method randomizes large samples with the intention of measuring incidence and generalizing results. The HMPS method has 18 chiefly implicit criteria. For both methods the AE must have occurred before or during the index admission and detected during and/or after index admission. The GTT method only includes commission while the HMPS method also includes omission. Despite research there remains a lack of knowledge of RRR. Whether comparisons are possible between independent experienced reviewer teams and how reviewer teams experience strengths and weaknesses of the RRR remain unknown. Nor have the feasibility of the GTT and the HMPS methods of the
same sample been investigated. From an improvement standpoint, it would be of interest to know whether RRR can be used in a more proactive manner, or merely for identifying patient harm. A deeper knowledge of the RRR could contribute to its development and understanding. This could be one way to improve patient safety work in clinical practice.
Aims

The general aim of this thesis was to evaluate, compare and expand retrospective record review methods for clinical use in healthcare.

Specific aims

Paper I.
To evaluate the agreement in judgments of adverse events between well-trained Global Trigger Tool teams from different hospitals.

Paper II.
To describe strengths and weaknesses, from team member perspectives of working with the Global Trigger Tool method of retrospective record review to identify adverse events causing patient harm.

Paper III.
To evaluate the feasibility and capability of two common retrospective record review methods, the “Harvard Medical Practice Study” method and the “Global Trigger Tool” in detecting adverse events in adult orthopedic inpatients.

Paper IV.
To evaluate retrospective record review for the detection and characterization of no-harm incidents and compare findings with conventional incident reporting systems.
Methodology

Design

A multi-method approach including both quantitative and qualitative methods was used. All papers focused on the RRR (I - IV), two of which had a comparative design (I, III). One study was qualitative based on focus group interviews with team members working with RRR (II). In the fourth paper, the RRR was used for identifying no-harm incidents (IV). The papers according to the GTT method are referred to in the Swedish GTT handbook from 2008 (I - III).

Setting

Papers I and II included teams from five hospitals in the southeastern region of Sweden. Four were middle sized with about 200 beds, and the fifth was a university hospital with about 600 beds. Papers III and IV were performed at a university hospital in Stockholm, Sweden with about 450 beds.

Sample and inclusion criteria

To evaluate agreement in AE assessments between well-trained GTT teams, a random sample of 50 records was selected for review from one of the participating hospitals (220 beds) between October 2009 and May 2010. All records eligible for selection met two criteria for inclusion: (1) inpatients with at least 24 hours of hospital stay; (2) and patients over 18. Records from surgical, orthopedic, gynecological and obstetric, medical, psychiatric and geriatric clinics were included in the study. Teams consisting of two RNs and one physician each carried out an RRR (I). In order to describe reviewer teams’ experiences of the strengths and weaknesses of the GTT method (II), the same five teams as in paper I participated. In one of the teams one additional RN participated. A total of 16 team members were interviewed five physicians and eleven RNs (II).

To evaluate the feasibility and capability of the HMPS and GTT methods (III) and to evaluate RRR by using the HMPS method for the detection of no-harm incidents (IV), a random sample of 350 admissions
from a total of 3701 during 2009, was selected. Based on earlier studies, AEs would occur in approximately 16% of the 3701 admissions at the orthopedic department, which would be sufficient to estimate the prevalence of AEs with a 95% CI of 12.16–19.84% (25, 56, 151).

Two teams consisted of one RN and two physicians for the review of paper III, while paper IV consisted of only one review team. The index admission emanated from the orthopedic department, and the AE (III) or no-harm incident (IV) had to be related to orthopedic care. For inclusion at least one of the following three criteria had to be met in papers III and IV:

(i) The AE/no-harm incident had to be caused within 30 days before index admission, leading to index admission (III), or be detected during index admission.

(ii) The AE/no-harm incident had to occur and be detected during index admission.

(iii) The AE/no-harm incident had to be caused during index admission and detected within 30 days of index discharge from the Orthopedic Department. AEs in this criterion were not required to result in a new admission.

**Procedure and data collection**

Paper I. Paper copies of records selected from one of the included hospitals were sent to all RRR reviewers. No validation or consensus was achieved between teams before study start. Instead, they were told to review exactly as they would have had the records come from their own hospital. For the same reason, no definition of AE was presented to the teams before study start. The RRR process consisted of two stages. In stage one records from the random sample were reviewed independently by experienced RNs in each team. When ready, records were discussed and analyzed, and then a consensus was reached. Records with potential AEs were given to the team physician. In stage two, the team physician performed an independent review containing a potential AE in stage one. The physician investigated possible causes for AEs related to the healthcare provided. The physician also estimated the severity of each AE and degree of preventability. If a patient had more than one AE, each was included and counted separately.

Paper II. The focus group interviews were made at the team members’ workplaces. Interviews were carried out by a moderator and an observer. The moderator was responsible for facilitating the discussion and prompting team members to speak, while the assistant moderator recorded the sessions and took notes. An introductory question was used: What are your experiences of the strengths and weaknesses of the GTT method? The introductory question was followed by transition-questions and key questions, as described by Krueger and Casey (154). All interviews were recorded and transcribed verbatim. The moderator and
assistant moderator presented their first impressions immediately after the focus group interviews and compared these interpretations to those found in earlier focus groups.

Paper III. In this study, the review process was standardized. One team reviewed using the GTT method with the RN screening for triggers, and the other team reviewed in accordance to the method used in HMPS, while the RN screened for criteria. Before study start a written manual, including definitions and detailed examples for each method was developed, discussed and approved by all team members. Each team also independently reviewed eleven training records and a consensus process allowed discussions of AE assessments and related matters. Team members unfamiliar with the computer system received an hour’s theoretical training. The RN reviewing by the GTT method read seven computerized training records before study start to become familiar with the computerized record system. One team, reviewing in accordance to the HMPS method, also screened for no-harm incidents. A three stage reviewing process was performed. In stage one; the random samples were reviewed by the RNs, one for each method. They screened either for the presence of 53 triggers (the GTT method) or 18 criteria (the HMPS method). A decision was made by the RNs regarding potential AEs. Only records with potential AEs were forwarded to the physicians. Each physician reviewed half the records forwarded by the RN. In stage two, physicians performed an independent review, deciding whether patient harm had occurred. Healthcare causation was assessed using a 6-point scale (Table 3). A similar 6-point scale was used to judge the preventability of the AE (Table 3). Severity was judged using standardized scales for each method, since they were different (Tables 2&4). The HMPS team also assessed the nature of the AEs, according to diagnostic procedures, drug treatment, invasive or other procedures. The nurses’ review process was evaluated, as was the physicians’. The physicians included any additional AE found not included by the RNs in the first stage. Every tenth record deemed not containing a potential AE was screened by a physician in each team. To assess inter-rater reliability between physicians a random sample of the forwarded records was reviewed double blinded within each team. After independent review the physicians discussed the records which had been reviewed twice, and then reached a consensus. In a third stage all AEs found by means of both methods were compared and discrepancies analyzed. Events only identified by one of the methods were reviewed by the physicians within the other team.

Paper IV. The team consisted of one RN and two physicians. The review process for no-harm incidents took place simultaneously during review for AEs by the HMPS method team (III). A two stage process was performed. In stage one the random sample was screened by the RN for the presence of 18 screening criteria, and a decision was made by the RN if the criteria found reflected a potential no-harm incident. The RN forwarded records with potential no-harm incidents to the two physicians for further judgment. In stage two, physicians confirmed the screening criteria found by the RN. Physicians classified the no-harm incidents into
eight categories, describing their nature. Place of occurrence and contributing factors were also assessed. A comparison between the no-harm incidents found by HMPS method and one local and four nationwide systems for incident reporting was also carried out. The five systems were 1) the obligatory clinical incident reporting system at the hospital, 2) hospital reports, named Lex Maria, 3) malpractice claims reported to LOF, 4) the Medical Responsible Board and 5) patient/relative complaints reported to the Patient Advisory Committee.

Data analysis

Statistical analysis

Descriptive statistics were used to describe the study sample (I-IV). Descriptive statistics were based on level and distribution of data and were reported as frequencies, means, medians, standard deviations, quartiles and ranges (I, III-IV) (155).

To evaluate the pair-wise agreement between reviewer teams (I), absolute agreement in percent (%) and weighted kappa ($\kappa_w$) statistics was used. Bootstrapping with 10,000 replications was conducted to estimate a 95% confidence interval for the $\kappa_w$ and enable calculating an overall unweighted kappa ($\kappa$) for all teams (156). The $\kappa$ and $\kappa_w$ were interpreted according to the following guideline: poor ($< 0.00$), slight (0.00-0.20), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61-0.80) and almost perfect (0.81-1.00) (157) (I).

To compare the HMPS and GTT methods respective proportions of verified criteria and triggers, chi-square was used. In relation to AEs, the Mann-Whitney $U$ test was used to compare and analyze results between the two methods, i.e. the GTT and HMPS, regarding the numbers of criteria and triggers found by the team’s RNs. To assess the association between the methods regarding the RNs review time, the Spearman rank order correlation coefficient ($r_s$) was used. The Spearman rank order correlation coefficient was also used to investigate “learning curves” for the RNs, regarding review time (III).

Chi-square test was used to compare the proportions of no-harm incidents criteria. In order to analyze the differences between groups, the Mann-Whitney $U$ test was used. To identify the number of times a specific screening criterion identified a no-harm incident divided by the total number of times the screening criterion was found, the positive predictive value (PPV) was calculated and reported in percent (%) (IV).

Analysis was carried out by using SPSS 19 for Windows (IBM Corporation, Somers, NY, USA) and Stata for Windows 12.0 (StataCorp LP, College Station, TX, USA) (I), StatView® v5.0.1 (SAS Institute Inc.,
Ethical considerations

Ethical permission was granted by the Regional Ethical Review Board at Linköping University for papers I and II (2010/56-31 and 2010/399-32). For papers III and IV, the Swedish Ethics Committee of Stockholm gave its approval (2008/951-31/3). A confidentiality agreement was signed by the external reviewers of papers I, III and IV. All patient information remained confidential throughout the studies. Permission was obtained to enter the records from the Chief Medical Officer from the hospital in southeastern Sweden (I) and from the Head of the Orthopedic Department and the Chief Medical Officer at the university hospital, and permission was obtained for access to the databases according to the Personal Data Act (III-IV) (158).

In these research studies, all reviewers had worked in healthcare for several years and were well acquainted with confidentiality. For the focus group interviews, all included RNs and physicians received an email with a request to participate (II), and before the focus groups interviews started, they gave their written consent. The team members were informed that their participation was voluntary and could at any time terminate participation.

In healthcare practice, most patients are familiar with health care documentation requirements and usually consider healthcare staff members reading their medical histories from other clinics as positive. However, there is always a risk that patients may experience a feeling of discomfort if aware that persons not directly involved in their care are able to access their records for research purposes alone. Patients have the right to block their data to healthcare staffers not directly involved in patient care. Data used for systematic and continual development of healthcare quality improvement cannot, however, be blocked by patients. It is important to keep a risk-benefit perspective in mind. By becoming aware of AEs and no-harm incidents to patients, these areas can be targeted. The studies in this thesis may lead to improving awareness of AEs among healthcare staff members enabling preventive patient safety work and reducing patient harm. From this point of view patient benefit may outweigh potential inconvenience caused by healthcare professionals reading patient records.

SAS Campus Drive, Cary, NC, USA) (IV) and Excel 2007 (III-IV). The level of statistical significance was set at p<0.05 (I, III-IV).

Qualitative analysis

To describe strengths and weaknesses from team member perspectives of the Global Trigger Tool method, focus group interviews (II) were carried out and analyzed in accordance with the method described by Krueger & Casey (154). The method is used to identify patterns in data and to discover relationships between experiences. Research triangulation was used, i.e. the researchers independently compiled, analyzed and interpreted the same phenomena, the focus group interviews. The transcripts were read and reread by all researchers to become completely familiar with the data and to comprehend essential features. The text was coded by all researchers independently and one of the researchers (KS) established categories based on citations. Then all researchers discussed the categories, and changes were made until consensus was reached. Opinions with similar meanings were grouped into eight categories.
Ethical considerations

Ethical permission was granted by the Regional Ethical Review Board at Linköping University for papers I and II (2010/56-31 and 2010/399-32). For papers III and IV, the Swedish Ethics Committee of Stockholm gave its approval (2008/951-31/3). A confidentiality agreement was signed by the external reviewers of papers I, III and IV. All patient information remained confidential throughout the studies. Permission was obtained to enter the records from the Chief Medical Officer from the hospital in southeastern Sweden (I) and from the Head of the Orthopedic Department and the Chief Medical Officer at the university hospital, and permission was obtained for access to the databases according to the Personal Data Act (III-IV) (158).

In these research studies, all reviewers had worked in healthcare for several years and were well acquainted with confidentiality. For the focus group interviews, all included RNs and physicians received an email with a request to participate (II), and before the focus groups interviews started, they gave their written consent. The team members were informed that their participation was voluntary and could at any time terminate participation.

In healthcare practice, most patients are familiar with healthcare documentation requirements and usually consider healthcare staff members reading their medical histories from other clinics as positive. However, there is always a risk that patients may experience a feeling of discomfort if aware that persons not directly involved in their care are able to access their records for research purposes alone. Patients have the right to block their data to healthcare staff members not directly involved in patient care. Data used for systematic and continual development of healthcare quality improvement cannot, however, be blocked by patients.

It is important to keep a risk-benefit perspective in mind. By becoming aware of AEs and no-harm incidents to patients, these areas can be targeted. The studies in this thesis may lead to improving awareness of AEs among healthcare staff members enabling preventive patient safety work and reducing patient harm. From this point of view patient benefit may outweigh potential inconvenience caused by healthcare professionals reading patient records.
Important to bear in mind is the risk of the patient’s perspective being overshadowed whereby all record reviewing is based on the health care staff’s documentation. It is possible that patients would provide a different interpretation of how their hospital stay was conducted.
Results

This thesis is based on four papers which evaluate, compare and expand the RRR method through investigation and agreement between reviewer teams (I), evaluation of the strengths and weaknesses of RRR (II), evaluation of two RRR methods (III), and using RRR to identifying no-harm incidents (IV). The main findings from the respective papers are presented below.

Paper I. Lack of consistency between reviewer teams using the Global Trigger Tool method

Differences in the number of AEs detected by different, but equally experienced review teams were found in their agreement when reviewing the same sample of records with a median reviewing time for the RNs of ten minutes. According to the judgment of preventability in particular, there were large differences between the reviewer teams. Greatest agreement was found in the detection of healthcare associated infections. The five teams identified 42 different AEs for 16 patients, corresponding to a level of AEs ranging from 27.2 to 99.7 per 1000 hospital days. The weighted kappa values for agreement between reviewer teams ranged from 0.26 to 0.77 team-by-team, with a combined unweighted $\kappa$ of 0.45 (95% CI=0.26-0.63), corresponding to moderate reliability. Four of the teams made similar assessments while the fifth team identified three times as many AEs compared to the other teams. With that team excluded, the agreement between reviewer teams increased from $\kappa=0.45$ to $\kappa=0.65$. The most common harm found was healthcare associated infections, i.e. pneumonia, sepsis and urinary tract infection. Most AEs resulted in minor, transient harm. Most of the AEs were judged preventable (58%), but in some case the differences between the teams’ judgment differed from 1-5 (urinary tract infection) and 2-5 (pressure ulcer). The preventable scale is shown in Table 3. The proportion of physician’s judgment of the preventable AEs ranged from 33% to 82% between teams.
Paper II. Experiences of strengths and weakness of the Global Trigger Tool method

Eight categories emerged with their strengths and weaknesses from the focus group interviews; "Usefulness of the GTT", "Application of the GTT", "Triggers", "Preventability of harm", "Team composition", "Team tasks", "Team member’s knowledge development" and "Documentation".

The teams found the GTT method useful whereby it identified patient harm and could be used for RRR for different specialties. Even if the teams mentioned that they reviewed according to the Swedish GTT handbook, gradually changes in the methodology were made by the teams. As an example, the Swedish GTT handbook recommends the RNs to sit in pairs when reviewing, but the teams reported how RNs divided up the records into two sets, each being read respectively. Further, the time limit had been exceeded.

The teams considered the GTT method useful. They also mentioned that their own knowledge had been developed by years of reviewing, and could now better observe care and harm from a patient perspective. Documentation was considered generally poor. Teams felt that for Swedish conditions added judgment of preventability, was too subjective, but nonetheless provided an opportunity for reflection, “Could we have prevented this from happening?”

Paper III. Evaluating the Harvard Medical Practice Study method and the Global Trigger Tool method

A total of 160 AEs were identified in 105 (30%) of the 350 records with the HMPS and the GTT methods combined. The median reviewer time for the RN reviewing in accordance to the HMPS method was three minutes compared to eight minutes for the RN using the GTT method. The review time for the physicians was six minutes for both reviewers using the HMPS method and four and eight minutes for the physicians reviewing in accordance to the GTT method. After the second stage, the HMPS found 151 (94%) AEs in 100 (29%) of the 350 records. Of this, 131 (87%) were deemed preventable. For the GTT method, 99 AEs were found (62%) in 85 (24%) of the records and 77 (78%) were deemed as preventable. The third review stage showed that 155 of the 160 AEs (97%) were found by the HMPS method in 104 records compared with 137 (86%) AEs in 98 records using the GTT method. Of the duplicated reviewed records, the assessment of physicians regarding healthcare causation before team discussions, were coherent within the teams in 93% and 88% of the cases for the HMPS and GTT methods respectively and preventability in 100% and 95% of the cases, respectively. Most AEs resulted in minor, transient harm and the majority was
judged to be preventable. The main difference between the methods regarding severity and types of AEs was found among the ones causing minimal or moderate impairment. These were predominantly urinary retention, infiltrated intravenous infusions, pressure ulcers and healthcare-associated infections. The nature of AEs is shown in Table 5.

**Table 5. The nature of adverse events**

<table>
<thead>
<tr>
<th>Adverse events</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthcare associated infection</td>
<td>38 (23.8)</td>
</tr>
<tr>
<td>Pressure ulcer</td>
<td>25 (15.6)</td>
</tr>
<tr>
<td>Urinary retention</td>
<td>17 (10.6)</td>
</tr>
<tr>
<td>Skin damage</td>
<td>12 (7.5)</td>
</tr>
<tr>
<td>Effects of drug treatment</td>
<td>11 (6.9)</td>
</tr>
<tr>
<td>Lack of surgical results</td>
<td>10 (6.3)</td>
</tr>
<tr>
<td>Subcutaneous infusions</td>
<td>8 (5.0)</td>
</tr>
<tr>
<td>Pain</td>
<td>6 (3.8)</td>
</tr>
<tr>
<td>Deep vein thrombosis/pulmonary embolism</td>
<td>6 (3.8)</td>
</tr>
<tr>
<td>Nerve damage</td>
<td>5 (3.1)</td>
</tr>
<tr>
<td>Heart failure</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td>Bleeding</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td>Thrombophlebitis</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td>Others</td>
<td>13 (8.1)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>160 (100.0)</strong></td>
</tr>
</tbody>
</table>

**Paper IV. Finding no-harm incidents using the Harvard Medical Practice Study method**

The HMPS method found 118 no-harm incidents in 91 of the 350 (26%) records, corresponding to an average of 0.34 no-harm incidents per admission (range 0-3). Ninety-four (80%) of the 118 no-harm incidents were classified as preventable. The physicians’ assessment of the double blinded reviewed records (n=11), were coherent in all cases concerning healthcare causation and preventability.
Drug-related no-harm incidents were found in 66 of the cases, and of these 88% were considered preventable. Team factors (verbal and written communication, supervision and seeking help, and team leadership) were the most common contributing factors found and contributed to 100 (41%) of the no-harm incidents. The nature of the no-harm incidents are shown in Table 6.

<table>
<thead>
<tr>
<th>No-harm incidents</th>
<th>n (%)</th>
<th>Preventable n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug related, nursing care</td>
<td>40 (33.9)</td>
<td>40 (100)</td>
</tr>
<tr>
<td>Drug related, medical care</td>
<td>26 (22.0)</td>
<td>18 (69.2)</td>
</tr>
<tr>
<td>Nursing care, excl. drug related</td>
<td>18 (15.3)</td>
<td>13 (72.2)</td>
</tr>
<tr>
<td>System related</td>
<td>12 (10.2)</td>
<td>12 (100)</td>
</tr>
<tr>
<td>Anesthesia related</td>
<td>10 (8.5)</td>
<td>3 (30.0)</td>
</tr>
<tr>
<td>Surgery/invasive actions</td>
<td>7 (5.9)</td>
<td>4 (57.1)</td>
</tr>
<tr>
<td>Diagnostics related</td>
<td>3 (2.5)</td>
<td>3 (100)</td>
</tr>
<tr>
<td>Treatment, excl. drugs and surgical procedures</td>
<td>2 (1.7)</td>
<td>1 (50.0)</td>
</tr>
<tr>
<td>Total</td>
<td>118 (100.0)</td>
<td>94 (79.7)</td>
</tr>
</tbody>
</table>

When comparing the findings with five conventional local and nationwide incident reporting systems, the incident reporting systems identified 16 incidents. Of these, ten were found by RRR. Nine of the no-harm incidents were derived from the local incident reporting system. One no-harm incident was found in the nationwide incident reporting system, which was also identified by RRR. Most of the no-harm incidents were related to drug therapy.
**Discussion**

**Discussion of findings**

The general aim of the thesis was to evaluate, compare and expand RRR methods for clinical use in healthcare. Differences in agreement were found between non-prepared teams (I) but also between teams using different RRR methods (III). Reviewer teams considered the RRR as well functioning even if all had modified the original method over time. RRR was found useful not only in identifying AEs but also for the detection of no-harm incidents.

The RRR has limitations in agreement between teams but has potential for improvement (I-IV). Careful preparation before review start seemed to play an important role for the nature of similarities between the teams (I, III, IV). In the study where the teams had no extra validation or consensus discussions before study start concerning the definition of AEs, or how to consider AEs, moderate agreement between reviewer teams was found (I). This is similar to other studies, although participants had used different validation steps before study start (143, 144, 159-161). Four teams made similar assessments while one team differed considerably from the others (I). It appeared as if this team assumed that AEs and triggers were equal, but in the focus group interviews they looked upon the triggers as indicating substandard care, i.e. if care had been carried out correctly the patient would not have need to return to the hospital within 30 days (II).

In the focus group interviews (II) results indicated that all teams initially followed the Swedish GTT handbook (26), but had later made small changes in methodology over time. This could be one explanation for inter-rater disagreement between reviewer teams (I). Another possibility could be that the method was initially interpreted differently at different hospitals, as shown by von Plessen et al. (161), but this is unclear as the teams were told to review as they were accustomed to. Changing the method has also been found by others (159, 161, 162). The developer of the GTT method points out that the method should not be used for comparisons, but used for comparing one’s own results over time (141). Mattson et al. (159) challenges this. They found when calculating the measurement error of the GTT method, the mean values of harm rates were within the measurement error of the GTT method. This was especially clear when teams...

---

**Table 6. Nature of no-harm incidents**

<table>
<thead>
<tr>
<th>Nature of no-harm incidents</th>
<th>n (% )</th>
<th>Preventable n (% )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug related, nursing care</td>
<td>40 (33.9)</td>
<td>40 (100)</td>
</tr>
<tr>
<td>Drug related, medical care</td>
<td>26 (22.0)</td>
<td>18 (69.2)</td>
</tr>
<tr>
<td>Nursing care, excl. drug related</td>
<td>18 (15.3)</td>
<td>13 (72.2)</td>
</tr>
<tr>
<td>System related</td>
<td>12 (10.2)</td>
<td>12 (100)</td>
</tr>
<tr>
<td>Anesthesia related</td>
<td>10 (8.5)</td>
<td>3 (30.0)</td>
</tr>
<tr>
<td>Surgery/invasive actions</td>
<td>7 (5.9)</td>
<td>4 (57.1)</td>
</tr>
<tr>
<td>Diagnostics related</td>
<td>3 (2.5)</td>
<td>3 (100)</td>
</tr>
<tr>
<td>Treatment, excl. drugs and surgical procedures</td>
<td>2 (1.7)</td>
<td>1 (50.0)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>118 (100.0)</td>
<td>94 (79.7)</td>
</tr>
</tbody>
</table>
estimated the harm level E (resulting in temporary harm and requiring intervention) and F (resulting in temporary harm and requiring initial or prolonged hospitalization). This is in line with paper III, were two RRR methods were compared, the HMPS and the GTT methods. These results showed differences between the two reviewing methods. The main differences found seemed to be diverging views of how to look at AEs causing minor or moderate impairment. These were predominantly urinary retention, infiltrated intravenous infusions, pressure ulcers and healthcare-associated infections. The HMPS method includes both commission and omission, while the GTT method only includes commission. Adopting a more proactive view of patient safety, omissions have been added to the Swedish MBRR handbook from 2012. It has also been mentioned that patient harm can be clearly linked to negligence e.g. when assessments, measurements or treatment are delayed or absent, they should be seen as preventable (147). When analyzing the nature of no-harm incidents, omissions were found common, e.g. when important medication such as antibiotic and thrombosis prophylaxis was prescribed but not administered, or hygiene precautions were missed (IV). Omissions due to lack of prophylactic treatment of pulmonary embolism or wound infection can be a no-harm incident in one case but result in major harm in another depending on various barriers and rescue strategies by the healthcare staff (163).

To increase agreement between teams is possible and desirable. von Plessen et al. (161) suggests training records, a written manual including definitions and examples of how to view AEs. All of these interventions were carried out in the study according to papers III and IV. The result showed a higher agreement between physicians within the teams than in paper I, where no validation or consensus was made before study start. Although results should be interpreted with great caution since the double-blinded sample was small and comparison was only between two physicians. Soop et al. (5) also showed high agreement between the physicians’ initial independent reviews with kappa values 0.80 for the assessment of AEs and 0.76 for the judgment of preventability.

The Swedish addition of assessment of preventability of AEs was difficult to carry out according to the teams (I - II). In contrast to the recommendations from the IHI the Swedish version of the GTT method also includes judgment of preventability. While the IHI considered there was a risk that the reviewers would focus too much on judging preventability (141), the Swedish GTT handbook from 2008 considered that the primary goal should focus on the preventable AEs as part of an ongoing quality improvement (26). In paper I, the judgment of preventability differed significantly between teams, and was also considered as too subjective (II). The judgment of preventability could differ from 1-5 on the preventability scale for the same AE, judged by the team’s physicians (I). The teams mentioned that it was easier to identify AEs than to
decide if they were preventable (II). Instead of judging preventability, it should be asked; “Could we have done this differently?” (II). This may lead to discussions about not only doing the job but also improving it (83). Another way is to look at every AE as being preventable as Classen et al. (136), who mention that AEs now considered as not being preventable in the future might be judged as indicating substandard care. This is also in line with Dahlberg & Segesten (1), who mention that what was seen as good practice yesterday may be considered malpractice today.

Lack of documentation is an obstacle when reviewing records (II), previously observed by others (34, 164, 165). In the focus group interviews the teams considered record documentation insufficient (II), which was also included as team factors, the most common contributing factor of the no-harm incidents (IV). The RRR is completely dependent on documentation which may pose a limitation. It is only possible to review what has been documented (166). If RNs and/or physicians have not documented their observations or actions, they cannot be read or researched. Also, omissions included in the HMPS method but excluded in the GTT method can be difficult to find in the records. Omissions can provide valuable information about quality of care (167). Lack of documentation can be an omission in itself. Ball et al. (167) showed that 47% of the RNs stated that they had missed developing or updating nursing care plans on their latest work shift. Also, medication lists have often been shown to be incomplete (168). On the other hand, given that many AEs are still found by the RRR it must be considered a useful method for identifying AEs.

Different types of RRRs identify different amounts of AEs (III). To reach the goal of not harming patients, knowledge of measurements, strengths, and weaknesses is important (II) (169). More AEs were found with the HMPS method than with the GTT method (III). The criteria of the HMPS method are more implicit, and reviewing should be comprehensive. This may not have a great impact in terms of time consumption if the length of hospital stay is only a few days, but with longer hospital stays it may be more time consuming. In paper I with patients from different clinics, the median time for RNs’ reviewing was ten minutes using the GTT method, and in paper III, from an orthopedic context, the median time was three minutes for HMPS method reviewing, and eight for reviewing according to the GTT method. The review time for physicians (III) was six minutes for both reviewers using the HMPS method, and four and eight minutes for the physicians reviewing according to the GTT method. Context knowledge and knowledge of the computerized system may have had an impact on the differences between reviewer time for the HMPS method and the GTT method in paper III, but results still show that time factor not necessary must be an obstacle to perform RRR (22, 170-172).
Most of the differences in findings between the HMPS and the GTT methods were, according to the GTT method, not identified and/or rejected AEs classified as minor events, e.g. urinary retention and infiltrated intravenous infusions (III). When working with improvement, patient safety interventions are needed to reduce major as well as minor AEs, as minor AEs may get worse and cause serious harm (173). Maximizing learning is possible when working with minor AEs depending on the often large number of events, which cause patient suffering and massive economic consequences (25).

Except for identifying AEs, the RRR method can be used to identify no-harm incidents (IV). An important reason for analyzing and acting on common and often preventable no-harm incidents is that an incident not leading to an AE in one case, might be a sentinel of serious system defects resulting in major harm in the next. No-harm incidents provide valuable information about why something did not happen, making it possible to recognize actions taken to prevent AEs from occurring (174). When in the process of reviewing for AEs, it may not be so time consuming to include a search for no-harm incidents as well. It should also be possible to review for no-harm incidents according to the GTT method (141), which uses an adapted version of the NCC MERP scale of severity (146). When reviewing for AEs the levels A-D are excluded as they comprise risks and no-harm incidents (171). The problems with identifying AEs resulting in minor or moderate impairment, as found in paper III evaluating the HMPS and the GTT methods, could be possible to overcome if including no-harm incidents level C (an error occurred in a patient but did not lead to any harm) and D (an error occurred in a patient and which led to the control or any kind of measure to ensure the patient is not harmed) when reviewing. This is new and recommended in Sweden when reviewing for AEs by the handbook of MBRR from 2012 as a contribution of this study (147).

RRR identifies substantially more no-harm incidents compared to other incident reporting systems. In paper IV, the no-harm incidents found by RRR were compared to five conventional local and nationwide incident reporting systems and were shown to identify considerably more no-harm incidents. When comparing AEs with two other methods Classen et al. (136) found that RRR identified at least ten times more AEs compared to other methods. Regarding paper IV, only a small overlap was found between different methods for identifying incidents and the result confirms other results (123, 127, 140). No system of itself is complete, showing that more than one method is needed (169). Together they can provide a clearer picture with a composite perspective and thus used to encourage hospital managers to better utilize acquired information (123, 127, 128, 175).

Collecting data can lead to change. One principle in quality improvement concepts is that it is not possible to improve things if they are not measured (176). Even if data collection by RRR of itself fails to increase
patient safety focusing on the data may aid in recognizing when useful information is generated and eventually reveal a pattern. Continually reviewing over time maximizes learning from the data. When a pattern is revealed it can be valuable to use the data, develop and test a change. One way to do this is by using the Plan-Do-Study-Act models (PDSA) in parallel for each small improvement step to evaluate the impact of the changes and learn about alternatives (82). The PDSA model is shown in Figure 2.

![PDSA model](image)

**Figure 2. The PDSA model (82)**

To present data and measurements over time, e.g. the number of AEs per 100 admissions, is a core measurement of the GTT method protocol. Neither this aspect was investigated in paper I-III nor statistical process control showing improvement over time. If data is available both before and after a change, statistical process control makes it possible to determine when a change has occurred. The teams in the focus group interviews (II) all talked of the importance of bringing results back to the clinics, but how this was done was not discussed. As part of improvement knowledge it can be useful that data is analyzed and presented on both a hospital and clinical level. This reveals patterns and indicates variation depending on changes or random variation (82).

**Methodological considerations**

In research, the most useful, productive and appropriate method for the research question should be used (177). A better understanding of the research problem could be reached by combining quantitative and qualitative methods than by either approach alone. To use mixed methods could either be done by applying
the concept in one single study or in a series of studies (178). In this research, a quantitative approach was used in three papers (I, III, IV) and a qualitative approach was used in the paper comprising focus group interviews (II).

All teams had at least three years of experience of RRR or working with patient safety issues (I-IV). In the first paper, 50 records were used (I). The records from the hospital included in this paper may not be representative for other teams in other hospitals but records were randomly selected and reviewed by five teams from five hospitals. Classen et al. (142) also used 50 records in the original evaluation of the GTT method.

The records from paper I came from different specialties (surgery, orthopedics, gynecology and obstetrics, internal medicine, psychiatry and geriatrics) while the other randomly selected records were selected from an orthopedic department (III, IV). AEs are more common in surgical specialties (25, 58, 125, 137, 149, 150, 179-181) which can make the results from papers III and IV difficult to generalize, but the aim was to evaluate RRR methods on a local level. In these studies, 350 random admissions were reviewed. From the number of the department’s admissions, this is about 10% of all admissions in 2009.

The aim of paper I was to let teams with several years of RRR experience review as they were used to and compare their assessments. Because of this, no validation or consensus was made between the teams before the start of the study. Instead, all five teams were instructed to review the records, selected from one of the five hospitals in the same way they would have reviewed records from their own hospitals. Nothing was said about the Swedish GTT handbook or to what extent they normally used it. No specific definition of an AE was presented before the study start, but it was implicit that the AEs found should be caused by healthcare and not the patients underlying disease. Four of the teams made similar assessments, which may indicate that other teams from other hospitals would perform assessments such as these four included in the study.

Trustworthiness comprises dependability, confirmability, and transferability, and is needed in all qualitative research strategies. The investigator must reflect carefully, deal with and report potential bias or error sources (177). The moderator’s previous experiences may have an impact on dependability and should be explained. The moderator of the focus group interviews had worked with the GTT method for four years on a hospital level, while the assistant moderator had experience of qualitative research and focus group interviews (II). The focus group interviews were carried out with the same teams (except that an additional RN attended one of the teams) as in paper I and had all worked together for a long time. From the first study, the team
members were familiar with the moderator of the focus group interviews based on the first paper (I). This may have had an impact on their answers by telling the moderator what they thought she wanted to hear, but it could also have encouraged members to be critical. From the moderator and assistant moderator’s views the teams took the opportunity to speak freely about both the strengths and weaknesses of the GTT method. Perhaps the answers would have been different if team members had been changed and new teams had been formed. On the other hand, the teams’ experiences seemed very similar, which may confirm that the results would have been rather like those achieved.

To ensure confirmability the transcribed interviews were returned to the team members with the question of whether they felt that the text reflected the interviews. Investigator triangulation was used to validate findings from the focus group interviews. The transcripts were read and reread by all researchers to gain a sense of content, and the researchers made notes and headings in the margins to include all aspects. Categories based on citations were established by one of the researchers. Then the categories were discussed by all researchers and a consensus was reached. Krueger & Casey (154) recommend at least three focus group interviews, and in study II five focus groups were conducted. To ensure dependability, a careful description of the sampling procedure and data analyses was presented. From the five teams (Team I–Team V) each citation was given a number for data reporting purposes to show evidence of reporting across responses. From the view of transferability, the analysis was based on propositions of focus group interviews from five teams, but it is impossible to be sure of how well they reflect other teams’ opinions from other hospitals or circumstances.

Context knowledge and experience of the computerized system could be reasons for the differences found between the teams’ judgments in paper III. This is in line with results presented by Sharek et al. (182). In paper III, one team per method was used. Another result could perhaps have been reached by using a crossover methodology, or by letting the teams change RRR method after reviewing half the admissions. The physicians of the team reviewing according to the GTT method rejected minor events as AEs in the third review step to a greater degree than the physicians using the HMPS method. To be classified as a minor AE (category E) in the severity scale used in the GTT method an intervention should have taken place. The perception of what is classified as an intervention could have affected the number of AEs found, and could also have affected the inter-rater reliability within the team using the GTT method. The HMPS method uses a severity scale more inclusive of minor AEs. The perception of minor AEs could have been affected by both the physician’s own view within the teams or to fundamental differences between the RRR methodologies described. Other results may have been shown if different teams, more teams, teams with
other experiences or records from other medical specialties had carried out the RRR. In order to improve this knowledge, more studies are needed.

The results from papers III and IV came from one orthopedic department from one Swedish hospital which may have influenced the findings. The review in stage one was made by one RN and may have been different if it had been made by another RN or by two RNs. Even if it was a small number of double reviewed records, the inter-rater agreement between the two physicians was high in their judgments made in the second step.
Conclusions

The general aim of the thesis was to evaluate, compare and expand retrospective record review methods for clinical use in healthcare.

- When evaluating the agreement in judgments of adverse events between well-trained Global Trigger Tool teams from different hospitals a moderate agreement was seen between reviewer teams. Participating in regular network meetings appeared to increase agreement between teams. If the Global Trigger Tool method is to be used in comparisons between hospitals it is important that training is offered and regular collaboration encouraged.

- When describing strengths and weaknesses from the team members' perspective of working with the Global Trigger Tool method the method was considered to be a useful tool and adaptable to different clinical specialties. Despite various adaptations of the tool the method still fulfilled its purpose of identifying patient harm in spite of modifications made by the teams.

- The evaluation of the feasibility and capability of the Harvard Medical Practice Study and Global Trigger Tool methods to identify adverse events in adult orthopedic inpatients showed that the Harvard Medical Practice Study method identified more adverse events than the Global Trigger Tool method. The main differences were found among the adverse events causing minimal or moderate impairment where the latter identified fewer adverse events.

- Using the Harvard Medical Practice Study method as a retrospective record review provided a means to detect and characterize no-harm incidents. Furthermore, it identified more no-harm incidents compared to conventional incident reporting sources. By adding this method to conventional incident reporting, healthcare providers can gain a clearer picture of commonly occurring, no-harm incidents in order to improve patient safety.
Clinical Implications

- When using retrospective record review it is of importance to have a standardized manual with clear definitions. It is equally relevant that teams come together to train retrospective record review, to meet and discuss findings and how to judge adverse events. If the ambition is that the agreement between teams should increase, this needs to be managed on a national level. Regional and local meetings can be performed between the national meetings, to keep the debate alive.

- To gain improvement of patient safety, it is of importance to look upon many of the retrospective record review findings as preventable. When the results from the reviewing are reported by the teams the staff on the wards should ask themselves the question; “Could we have done this differently?” Working with retrospective record review develop the team members and give them the opportunity to use their findings for educating other healthcare providers, making changes in clinical practice and thus improving patient safety work.

- An advantage with retrospective record review is that other things can be found while reviewing, that can be of importance to improve, e.g. the quality of documentation. If the goal with the retrospective record review is to achieve improvement on a local clinical level, it can be an advantage to review records from the own specialty as it may be easier to understand why certain things are done. It can is also be easier to give feedback to the own specialty.

- The choice of retrospective record review method may influence the findings of adverse events. Trained reviewers reviewing records from patients with short hospital admissions can benefit from using the Harvard Medical Practice Study method as this provides a more in-depth analysis of the records.

- Retrospective record review can also be of use for the identification of no-harm incidents in parallel with reviewing for adverse events on a clinical level. This can be done with no great effort and need not be too time consuming. In a proactive patient safety environment identification of no-harm incidents may well play a significant role, as it may prevent some of these incidents from becoming patient harm events. Here lies a demanding challenge for safety conscious health workers!
Future research

- Recently the Swedish handbook, of patient safety has been published (2012). Studies are recommended to evaluate the effect of applying the models recommended in the handbook on improving the comparability between hospital teams.

- The study where the HMPS method and the GTT method were compared was made in an orthopedic clinic with short hospital stay. Further studies are needed to investigate the feasibility of these methods in other specialties with longer duration of admissions.

- In contrast to other sectors in society such as the aviation industry healthcare has as yet not found a structured way to really deal with no-harm incidents. Further studies are needed from which healthcare staff will be able to learn how to use no-harm incidents in order to increase patient safety.
Populärvetenskaplig sammanfattning


Det övergripande syftet med avhandlingen är att utvärdera, jämföra och utvidga strukturerad journalgranskningens kliniska användbarhet inom hälso- och sjukvård.

Avhandlingen består av fyra artiklar, baserade på tre olika material. Delstudiernas syften är följande:

- Att utvärdera överensstämmelsen i bedömningar av skador mellan erfarna Global Trigger Tool team från olika sjukhus (I).

- Att utifrån erfarna Global Trigger Tool teams perspektiv, beskriva metodens styrkor och svagheter för att identifiera händelser som orsakar patientskador (II).
Fem fokusgruppsintervjuer genomfördes med 16 deltagare från fem olika journalgranskningsteam, för att undersöka deras erfarenheter av att använda GTT metoden. Fokusgruppsintervjuer användes för att belysa teammedlemmarnas erfarenheter av metodens styrkor och svagheter. Intervjuerna analyserades genom en speciell metod utvecklad för fokusgruppsintervjuer (II).

I studie I fick teamen inte någon genomgång av definitioner, inga konsensusdiskussioner genomfördes innan studiestart och de fick inte heller någon detaljerad manual utan ombads granska enligt handboken för "Strukturerad journalgranskning" som de brukade. Resultatet visade att det fanns skillnader mellan de bedömningar som de granskade teamen gjorde. Skillnaderna återfanns såväl vad gällde antalet patientskador, nivån på skador och inte minst vad gällde bedömningen av undvikbarhet. Totalt identifierade teamen 42 olika skador, motsvarande en nivå på 27.2 till 99.7 skador per 1000 vårddagar. Fyra av teamen gjorde liknande bedömningar medan det femte teamet identifierade tre gånger så många skador...
som de övriga. De fyra team som gjorde liknande bedömningar ingick i ett regionalt nätverk där man träffades årligen för diskussioner angående vårdskador. Den vanligaste skadan som teamen fann var vårdrelaterade infektioner (I).

De flesta av skadorna bedömdes som undvikbara (58%) och här återfanns stora skillnader mellan de bedömningar som teamen gjorde. Detta gällde till exempel en patient som drabbats av en urinvägsinfektion och undvikbarheten av skadan bedömdes av teamens läkare ligga mellan 1-5, där 1 stod för "inget verkligt belägg för undvikbarhet" och 5 för "starka belägg för undvikbarhet". Ett annat exempel gällde en patient som drabbats av trycksår. Där skiljde sig läkarnas bedömningar mellan 2-5, där 2 stod för "svagt till ringa belägg för undvikbarhet" och 5 för "starka belägg för undvikbarhet". Procentsatsen för de undvikbara skadorna varierade mellan 33% till 82% (I).

Vid fokusgruppsintervjuerna framkom åtta kategorier med både styrkor och svagheter av att använda GTT metod. Teamen ansåg att GTT var en användbar och viktig metod, även om bedömningen av patientskadans undvikbarhet framstod som alltför subjektiv. Trots att teamen menade att de utförde sina granskningar enligt handboken för GTT metod, var det tydligt att de trots allt hade gjort vissa förändringar av metoden över tid. Teamen berättade till exempel att de istället för att sitta och granska tillsammans, vilket är rekommenderat enligt GTT metoden, delade sjuksköterskorna numera upp journalerna i två högar och granskade hälften var. Man hade även överskridit den rekommenderade tidsgränsen på 20 minuter. Teamen ansåg att deras viktigaste men också svåraste uppgift, var att informera de berörda klinikerna om resultaten av granskningarna. Teamen menade också att det var lättare att granska journaler från sin egen verksamhet (II).

Vid utvärderingen av HMPS och GTT metoderna identifierades sammanlagt 160 olika skador i 105 (30%) av de granskade 350 journalerna. HMPS metoden fann 155 av de 160 (97%) skadorna i 104 (30%) journaler, i jämförelse med GTT metoden som fann 137 (86%) skador i 98 (28%) journaler. Den största skillnaden återfanns i gruppen "skada utan bestående men". Samstämmigheten mellan läkarna inom respektive team, var hög i den här studien. Alla granskare hade fått en utförlig manual att följa, träningshantering användes och diskussioner genomfördes kring hur patientskador skulle bedömas innan studien påbörjades (III).

När HMPS metoden samtidigt användes för att identifiera tillbud, återfanns 118 tillbud i 91 (26%) av de 350 journalerna som granskades, av vilka 94 (80%) bedömdes som undvikbara. Därefter gjordes en jämförelse med fem system för incidentrapportering, ett lokalt och fyra nationella. Totalt identifierade de fem systemen
16 tillbud, av vilka tio också återfanns genom retrospektiv journalgranskning. Den vanligaste typen av tillbud var relaterad till läkemedelshantering (n=66) och 88% av dessa bedömdes som undvikbara (IV).

Retrospektiv journalgranskning har visat sig vara en användbar metod, både vad gäller identifiering av skada och av tillbud. Retrospektiv journalgranskning identifierar fler skador i jämförelse med andra metoder. Gemensam undervisning och träning i hur bedömningar ska göras kan ge en ökad samstämmighet mellan olika granskningsteam. Om metoderna ska användas för jämförelser mellan kliniker och sjukhus, är en samsyn vad gäller bedömningar viktig.

Genom att även använda retrospektiv journalgranskning för identifiering av tillbud kan ny och viktig information erhållas, som i sin tur kan användas för att förbättra patientsäkerheten och därmed förhindra att ytterligare patienter drabbas av vårdskador.
Acknowledgments

This thesis has been carried out at the School of Health and Caring Sciences, Faculty of Health and Life Sciences, Linnaeus University, with financial support from FORSS. I would like to express my sincere gratitude to everyone who has supported me during my years as a PhD student. This thesis would not have been possible without you. I would especially like to thank:

**Joep Perk**, my chief supervisor, who made my research possible. Your energy, inspiration and ideas are endless. You have truly inspired and encouraged me.

**Lena Nilsson**, my co-supervisor for always being there. I could never count all the hours you have spent on reading my texts. You will finally get some time now for other things! Without your experience of patient safety and RRR my studies would not have been possible.

**Kristofer Årestedt**, my co-supervisor for all support and positive thinking. Your knowledge of quantitative research has been of great value to me.

**Gunilla Nilsson**, co-author of paper II with her extensive experience of qualitative studies and focus group interviews, and for her good spirit and pleasant meetings.

**Maria Unbeck**, colleague and co-author of papers III and IV. Your hawk eyes miss nothing!

**Kiku Pukk Härenstam, Olav Muren, Urban Jurgensen** and **Peter Henriksson**, for creative collaboration and co-authorship.

**Ulla Peterson** and **Hasse Rutberg** for valuable comments at my half-time seminar. This made me see things in a different light and made it possible for me to go further.
Marie Oscarsson and Motzi Eklöf, examiners at my final seminar. Your enthusiasm and wise comments encouraged me to complete my thesis.

To all team members, for your cooperation with doing retrospective record reviews and for letting me conduct focus group interviews with you.

Solbritt Andersson at the University Library for her patience with me and my struggles with EndNote.

Matts Thorsson for design of figures and tables.

Colleagues, with special thanks to Birgitta and my former colleague Anki, for your support, pep talks and laughter.

My mother Elsie, who always told me to keep on studying.

My brother Birger, for all inspiring discussions.

My brother Anders, for all sporting events we have discussed and watched together. They were welcome breaks during my studies.

My friends and neighbors, for not giving up the hope of me finishing my thesis, and for time spent with music, travelling, book discussions and other nice things.

Mikael’s children; Jim with girlfriend Johanna, Roger with his wife Cecilia, and Malin, for their interest and support.

Jenny, Simon and Samuel, my children, who always had great faith in my ability to finish my thesis. You were, as always, right! And to Diego, Kae and Lill, for making me and my children happy.

Mikael, even if you still don’t understand why, I’m happy every day that you are in my life.
References


115. Sveriges kommuner och landsting. Patientsäkerhetssatsning 2013 överenskommelse mellan staten och Sveriges kommuner och landsting om förbättrad patientsäkerhet. [cited 2013 October 6]; Available from: http://www.skl.se/mediaservice/mediaservice.axd?MediaArchive_FileID=70ca1d0e-47d4-4a03-9ccd-22a9f481b697&FileName=underskrivna%C3%96K+Patients%C3%A4kerhetssatsning+2013.pdf.


104. Händelseanalys & implementation. Thibault G, Gandhi TK

105. Frankel A, Grillo SP, Baker EG, Huber CN, cause analysis system in action.


22a9f481b697&FileName=underskrivna%C3%96K+Patients%C3%A4kerhetssatsning+2013.pdf


121. Sveriges kommuner och landsting. Nationella kvalitetsregister. [cited 2013 October 6]; Available from: http://www.kvalitetsregister.se/


127. Olsen S, Neale G, Schwab K, Psaila B, Patel T, Chapman E, Vincent C. Hospital staff should use more than one method to detect adverse events and local real-time record review may all have a place. Qual Saf Health Care. 2007;16(1):40-4.


170. Christianss-Dingelhoff I, Smits M, Zwaan L, Lubberding S, van der Wai G, Wagner C. To what extent are adverse events found in patient records reported by patients and healthcare professionals via complaints, claims and incident reports? *BMJ Health Serv Res*. 2011;11:49.
Table 1. The Global Trigger Tool method triggers used in papers I-III (Swedish version with 53 triggers).

**Care Module**
- C1 Transfusion of blood product
- C2 Decrease of hemoglobin
- C3 In hospital stroke
- C4 Code/arrest/rapid response team
- C5 Dialysis
- C6 Positive blood culture
- C7 Emboli or Deep Vein Thrombosis
- C8 Patient fall
- C9 Pressure ulcers
- C10 Readmissions within 30 days
- C11 Healthcare associated infections
- C12 Transfer to higher level of care
- C13 Treatment
- C14 Other

**Surgical Module**
- S1 Return to surgery
- S2 Change in procedures
- S3 Admission to intensive care post-op
- S4 Intubation/reintubation/Continuous Positive Airway Pressure/Biphasic Positive Airway Pressure
- S5 X-ray intra-op or post-op or in Post-anesthetic Care Unit
- S6 Intra-op or post-op death
- S7 Mechanical ventilation greater than 24 hours
- S8 Intra-op adrenaline
- S9 Increased post-op troponin level
- S10 Change in anesthetic type
- S11 Consultant in the recovery room
- S12 Post-op complication
- S13 Abnormal pathological anatomical diagnosis
- S14 Insertion of catheters during surgery
- S15 Op time > 6 hours
- S16 Organ removal during op


182. Sharek PJ, Parry G, Goldmann D, Bones K, Hackbarth A, Resar R, Griffin FA, Rhoda D, Murphy C, Landrigan CP. Performance characteristics of a methodology to quantify adverse events over time in hospitalized patients. Health Serv Res. 2011;46(2)654-78.
Appendix

Triggers and screening criteria are listed below.

Table 1. The Global Trigger Tool method triggers used in papers I - III (Swedish version with 53 triggers).

<table>
<thead>
<tr>
<th>Care Module</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 Transfusion of blood product</td>
</tr>
<tr>
<td>C2 Decrease of hemoglobin</td>
</tr>
<tr>
<td>C3 In hospital stroke</td>
</tr>
<tr>
<td>V4 Code/arrest/rapid response team</td>
</tr>
<tr>
<td>C5 Dialysis</td>
</tr>
<tr>
<td>C6 Positive blood culture</td>
</tr>
<tr>
<td>C7 Emboli or Deep Vein Thrombosis</td>
</tr>
<tr>
<td>C8 Patient fall</td>
</tr>
<tr>
<td>C9 Pressure ulcers</td>
</tr>
<tr>
<td>C10 Readmissions within 30 days</td>
</tr>
<tr>
<td>C11 Healthcare associated infections</td>
</tr>
<tr>
<td>C12 Transfer to higher level of care</td>
</tr>
<tr>
<td>C13 Treatment</td>
</tr>
<tr>
<td>C14 Other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Surgical Module</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 Return to surgery</td>
</tr>
<tr>
<td>S2 Change in procedures</td>
</tr>
<tr>
<td>S3 Admission to intensive care post-op</td>
</tr>
<tr>
<td>S4 Intubation/reintubation/Continuous Positive Airway Pressure/Biphasic Positive Airway Pressure</td>
</tr>
<tr>
<td>S5 X-ray intra-op or post-op or in Post-anesthetic Care Unit</td>
</tr>
<tr>
<td>S6 Intra-op or post-op death</td>
</tr>
<tr>
<td>S7 Mechanical ventilation greater than 24 hours</td>
</tr>
<tr>
<td>S8 Intra-op adrenaline</td>
</tr>
<tr>
<td>S9 Increased post-op troponin level</td>
</tr>
<tr>
<td>S10 Change in anesthetic type</td>
</tr>
<tr>
<td>S11 Consultant in the recovery room</td>
</tr>
<tr>
<td>S12 Post-op complication</td>
</tr>
<tr>
<td>S13 Abnormal Pathological Anatomical Diagnosis</td>
</tr>
<tr>
<td>S14 Insertion of catheters during surgery</td>
</tr>
<tr>
<td>S15 Op time &gt; 6 hours</td>
</tr>
<tr>
<td>S16 Organ removal during op</td>
</tr>
</tbody>
</table>
Table 2. The screening criteria from Harvard Medical Practice Study method used in papers III - IV.

<table>
<thead>
<tr>
<th>Screening criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The index admission was an unplanned admission related to previous healthcare management within 30 days</td>
</tr>
<tr>
<td>2. Unplanned readmission after discharge from index admission within 30 days including outpatient visits</td>
</tr>
<tr>
<td>3. Hospital-incurred patient injury or no-harm incident</td>
</tr>
<tr>
<td>4. Adverse drug reaction</td>
</tr>
<tr>
<td>5. Unplanned transfer from general to intensive care</td>
</tr>
<tr>
<td>6. Unplanned transfer to another acute hospital</td>
</tr>
<tr>
<td>7. Unplanned return to the operating room</td>
</tr>
<tr>
<td>8. Unplanned removal, injury or repair of an organ during surgery</td>
</tr>
<tr>
<td>9. Other patient complication</td>
</tr>
<tr>
<td>10. Development of neurological deficit not present on admission</td>
</tr>
<tr>
<td>11. Unexpected death</td>
</tr>
<tr>
<td>12. Inappropriate discharge to home.</td>
</tr>
<tr>
<td>13. Cardiac or respiratory injury related to abortion or delivery</td>
</tr>
<tr>
<td>14. Healthcare associated infection or sepsis</td>
</tr>
<tr>
<td>15. Dissatisfaction with care documented in the patient’s medical record</td>
</tr>
<tr>
<td>16. Documentation or correspondence indicating litigation</td>
</tr>
<tr>
<td>17. Any other undesirable outcome not covered above</td>
</tr>
</tbody>
</table>

**Medication Module**

<table>
<thead>
<tr>
<th>Medication Module</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1  Clostridium difficile positive stool</td>
</tr>
<tr>
<td>M2  Activated Partial Thromboplastin- time &gt; 100 s</td>
</tr>
<tr>
<td>M3  International Normalized Ratio &gt; 6</td>
</tr>
<tr>
<td>M4  Glucose &lt; 3 mmol/l</td>
</tr>
<tr>
<td>M5  Serum creatinine greater than 2 times baseline</td>
</tr>
<tr>
<td>M6  Vitamin K administration</td>
</tr>
<tr>
<td>M7  Antihistamine</td>
</tr>
<tr>
<td>M8  Flumazinil</td>
</tr>
<tr>
<td>M9  Naxolone</td>
</tr>
<tr>
<td>M10 Anti-emetic</td>
</tr>
<tr>
<td>M11 Over-sedation/hypotension</td>
</tr>
<tr>
<td>M12 Abrupt medication stop</td>
</tr>
</tbody>
</table>

**Intensive Care Module**

<table>
<thead>
<tr>
<th>Intensive Care Module</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1  Pneumonia</td>
</tr>
<tr>
<td>I2  Readmission of intensive care</td>
</tr>
<tr>
<td>I3  In-unit procedure</td>
</tr>
<tr>
<td>I4  Intubation/reintubation</td>
</tr>
</tbody>
</table>

**Perinatal Module**

<table>
<thead>
<tr>
<th>Perinatal Module</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1  Apgar-points &lt; 7</td>
</tr>
<tr>
<td>P2  Transportation/transmission of mother/child</td>
</tr>
<tr>
<td>P3  Magnesium/terbutalin</td>
</tr>
<tr>
<td>P4  Serious lacerations injuries</td>
</tr>
<tr>
<td>P5  Instrumented delivery</td>
</tr>
</tbody>
</table>

**Emergency Department Module**

<table>
<thead>
<tr>
<th>Emergency Department Module</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1  Readmission to Emergency Department within 48 hours</td>
</tr>
<tr>
<td>E2  Time in Emergency Department &gt; 6 hours</td>
</tr>
</tbody>
</table>
Table 2. The screening criteria from Harvard Medical Practice Study method used in papers III - IV.

<table>
<thead>
<tr>
<th>Screening criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The index admission was an unplanned admission related to previous healthcare</td>
</tr>
<tr>
<td>management within 30 days</td>
</tr>
<tr>
<td>2. Unplanned readmission after discharge from index admission within 30 days</td>
</tr>
<tr>
<td>including outpatient visits</td>
</tr>
<tr>
<td>3. Hospital-incurred patient injury or no-harm incident</td>
</tr>
<tr>
<td>4. Adverse drug reaction</td>
</tr>
<tr>
<td>5. Unplanned transfer from general to intensive care</td>
</tr>
<tr>
<td>6. Unplanned transfer to another acute hospital</td>
</tr>
<tr>
<td>7. Unplanned return to the operating room</td>
</tr>
<tr>
<td>8. Unplanned removal, injury or repair of an organ during surgery</td>
</tr>
<tr>
<td>9. Other patient complication</td>
</tr>
<tr>
<td>10. Development of neurological deficit not present on admission</td>
</tr>
<tr>
<td>11. Unexpected death</td>
</tr>
<tr>
<td>12. Inappropriate discharge to home.</td>
</tr>
<tr>
<td>13. Cardiac or respiratory</td>
</tr>
<tr>
<td>14. Injury related to abortion or delivery</td>
</tr>
<tr>
<td>15. Healthcare associated infection or sepsis</td>
</tr>
<tr>
<td>16. Dissatisfaction with care documented in the patient's medical record</td>
</tr>
<tr>
<td>17. Documentation or correspondence indicating litigation</td>
</tr>
<tr>
<td>18. Any other undesirable outcome not covered above</td>
</tr>
</tbody>
</table>