A potential for further enhancing obstetrical safety

Patient harm measurement with the global trigger tool in the south-east health-care region of Sweden

Raymond Lenrick

Degree project, 30 credits, Master Thesis
Quality improvement and leadership in health and welfare
Jönköping, June 2012

Supervisor: Karin Havemose, assistant professor
Examiner: Boel Andersson-Gäre, professor
Abstract

A decade of heightened awareness concerning safety issues in healthcare since the Institute of Medicine’s awakening call has resulted in a string of counteroffensive measures. The pace of improvement has been slow and not altogether clear. Rates of patient harm are in general now measured by voluntary reporting and indicator measurements. The use of triggers or clues in random nurse-based reviews to enable identification of patient harm is a more effective method for measuring the overall rate of harm in a health care organisation. Measured actual overall rates of patient harm, their variations and patterns during delivery in the south-east health-care region of Sweden are not previously known. Measurement is important to patient safety improvement, as a foundation for accountability, effort selection and keeping track of results. The patient’s voice must also be much clearer in quality and safety improvement efforts in healthcare. The Institute of Healthcare Improvements Global Trigger Tool for measuring adverse events was used to review 1137 deliveries during 2011 in the seven departments (10% of all cases). Mother and newborn were both evaluated. Thirty eight patient harm events per 1000 patient days were identified, correlating to 13% of admissions. Presupposed rates among staff were double this value. Current patient safety indicators are half this value. One third of patient harm events at birth affected the newborn. Twenty different categories of harm were found. This study shows significantly higher rates of patient harm than previously reported. The nurse reviewers defined the method as valuable and a useful method for measuring harm at delivery. Limitations at this stage are no observed changes in health care delivery or clinical outcomes and that value assessment is based entirely on the judgement of the data-abstractors.

Keywords: Quality Improvement, Global Trigger Tool, Patient Safety, Obstetric Care
Introduction

Background knowledge

A decade of heightened awareness concerning safety issues in healthcare since the Institute of Medicine’s awakening call [1] has resulted in a string of counteroffensive measures. The pace of improvement has been slow [2] and not altogether clear [3], although the language of patient safety is now firmly set in healthcare. The Institute of Medicine defines adverse events (AEs) as always associated with ‘unintended harm to the patient by an act of commission or omission rather than by the underlying disease or condition of the patient’ or in summary the reasonable expectation of ‘freedom from accidental injury’ [1, p18]. An AE indicates an undesirable clinical outcome often associated with medical errors, but does not always mean error, negligence or poor quality of care. It may not either be preventable in the present design of healthcare. Examples of commonly understood AEs are wrong site surgery, medication errors, care associated infections, pressure ulcers, falls and faulty diagnoses. There are however several more layers of accidental harm afflicted on patients by healthcare, poorly understood by both beneficiaries and system holder professionals or managers.

Rates of AEs are in general now measured by voluntary reporting and indicator measurements [4]. The use of "triggers", or clues, to enable identification of AEs is an effective method for measuring the overall rate of harm in a health care organisation [5]. The Institute of Healthcare Improvement’s (IHI) Global Trigger Tool for Measuring AEs (GTT) provides a validated audit method that retrospectively identifies AEs using standardised review of randomly selected patient records and, as a result, allows measurement of the actual rate instead of the reported rate [6, 7, 8]. The GTT identifies AEs defined as harm to the patient resulting from medical diagnosis or therapeutics, irrespective of error or preventability. Adverse outcomes caused solely by underlying disease or by the intended consequences of treatment are not considered AEs. GTT is a reliable and practical method for measuring the overall occurrence of AEs to hospitalised patients over time and can be used to assess the effectiveness of interventions to improve patient safety [9]. As GTT is based on random sampling, at least 12 and preferably 24 monthly measurements are needed for a baseline.

A study of methods for identifying AEs in hospitals by the US Department of Health and Human Services [10] showed that nurse reviews using triggers identified 78% of defined AEs. Analysis of indicators present on admission identified 51% of all AEs, whereas beneficiary interviews exposed only 18%. The other commonly used methods studied in that report were hospital incident reports and patient safety indicator (PSI) analysis. These methods identified only 7% of all defined AEs. Similar results shown by Classen [11] compared three methods to detect AEs in hospitalised patients, using the same patient sample set from three leading hospitals. Adverse event detection methods commonly used to track patient safety in the United States today fared very poorly compared to other methods and missed 90% of the AEs. The Institute for Healthcare Improvement’s GTT found at least ten times more confirmed, serious events than these other methods. Only 6.2% of defined PSIs were also identified in the provider-reported AEs system [12]. Indicator measurements such as obstetric trauma risk, is significantly influenced by both patient and hospital characteristics and not a good indicator of patient safety [13].

Results from hospital GTT measurements including all admitted patients commonly produce values between 20 [14] and 50 [15] AEs per 1000 patient days. These studies include an obstetrical subset which is not reported separately. Other reported values are 27 [16], 40 [17] and 41.6 in a recent large scale measurement across a large hospital system [18].
A report [19] from the patient safety office at the County council of Östergötland, Sweden, showed that a structured review of surgical records with GTT identified 45 events, of which only two had been reported in the AE reporting system. This proved to be a strong driver of patient safety improvements. GTT hospital measurements are available for all of the seven hospitals in the south-east health-care region of Sweden since several years. The current range is 10.1 to 18.6 AEs per 1000 patient days (unpublished data).

Obstetrical AE has been measured by a variety of methods and produced notably different values. Forster et al [20] found a rate of 8 AEs per 1000 patient days using 72 triggers in 425 obstetrical patient records. Florea et al [21] reported that 0.4% of deliveries in a self-reported system showed evidence of actual harm. Mann et al [22] found one or more AEs in 9.2% of deliveries. Measurement of medication errors in obstetrical care showed 2% AEs [23] with 3.2% of this causing patient harm. A short test using GTT methodology has been done at the Department of Obstetrics and Gynaecology at the University of Linköping in 2009. This was done as a specialist exam essay (unpublished data). Reviewing 10 deliveries monthly for 6 months showed one AE in four of the months, a rate of 6.6%.

The newly introduced Swedish patient safety law [24] requires in chapter 3, § 2"that the caregiver shall take all measures needed to prevent patients suffering preventable harm”. In order to take all possible measures a deep understanding of local safety rates, variations, patterns and underlying mechanisms is desirable.

**Local problem**

Measured actual overall rates of AE, their variations and patterns during delivery in the south-east health-care region of Sweden with seven departments of Obstetrics and Gynaecology are not known. Rates are now inferred from PSIs and AE reporting systems. This knowledge has been paramount to past improvements but lacks a comprehensive and patient-orientated approach. Apart from increasing the effort in present approaches, the regional board of Heads of Departments (meeting minutes) has expressed the need to add a further dimension to safety improvement efforts. This is attuned to the increasing awareness that the patient’s voice must be much clearer in quality and safety improvement efforts in healthcare.

**Intended improvement**

Measurement and identification of patient harm rates, grades and types is important for patient safety improvement, as a foundation for accountability, effort selection and keeping track of results. The specific aim of this improvement effort is the evaluation of a method for valid comprehensive assessment of obstetrical AE that can be followed over time and to compare its results to presupposed rates, reported levels of PSIs and reported AEs. The general aim is that this will onwards facilitate a continued decrease in harm at delivery departments in the south-east health-care region of Sweden and that the method should supplement present efforts while clearly adding the patient’s perspective.

Supporters of the quality improvement are the regional board of Heads of Departments, obstetrical leaders and the Development unit at Jönköping County Council. Important drivers are that GTT is an effective method for measuring the overall rate of AE with the patient’s perspective, the presence of local unique knowledge of the GTT method and its accelerating spread across Sweden for measuring patient harm in hospitals and departments.
**Study question**

The primary improvement intervention question is how do measured overall actual rates of AE during delivery in 2011 at all departments in the south-east health-care region of Sweden compare to presupposed rates of AE among co-workers and to reported levels of PSIs and reported AEs. The overall rate is also understood concerning the variation, grades and types of AEs.

The secondary study of the improvement explores the attributed value of measuring AE with GTT as evaluated by participating data-abstraction midwives. Value is here defined as benefit divided by effort. Benefit is defined as confidence in the results as indicative of its possible influence on accountability, effort selection and keeping track of results. Effort is defined as the resources applied.

Knowledge of the triggers identified in this improvement can also be valuable for further development of trigger tools for obstetrics.

The study question was set because of the author’s previous initiation and participation in this regions evolving peer reviewed PSI driven improvement work. There was a need felt by the author for a next step. Exposure to the hospitals GTT measurements gave an indication of where this next step might be found. Participation in the Masters course gave the necessary incentive.

**Methods**

**Ethical issues**

Ethical aspects of implementing the improvement are mainly concerned with the divulgence of potentially embarrassing or harmful information about rates of AE afflicted by the health care services studied. Anonymity of partaking departments in the study is guaranteed as the results are presented as aggregated values for the region. The GTT method is not designed for inter-unit comparison. Implementation of this improvement falls within the scope of the Swedish Health Act which requires continual services improvement.

Privacy concerns raised in studying the improvement are addressed by voluntary and anonymous questionnaires and interviews. There is no risk that the physical well-being of participating persons is endangered. The author has at the time of this study no managerial position. The study of this improvement has been deemed by the local ethics Committee at the Jönköping University, decision number 11-8, to not need a full ethical review. The potential gains in patient safety outweigh potential negative factors.

**Setting**

In the south-east health-care region of Sweden there are seven departments of Obstetrics and Gynaecology with together approximately 10 000 deliveries yearly. These departments are located in the major towns of the three county councils, Östergötland, Kalmar and Jönköping. They have very similar ownership, staff rates, guidelines, medical technology infrastructure, incentives and cultural attributes. Patient safety culture measurements in Östergötland and Jönköping County have showed similar results. Delivery departments are staffed with in-house first call registrars, second call consultants, midwives and junior nurses. Leading each department is a senior obstetrician as medical director together with the head midwife as unit manager. Each department has a process improvement team respondent to the managers.
Peer-reviewed evaluation of obstetrical outcomes (PSI analysis) has been deployed for more than a decade and is an important driver for safety improvement. This has involved the collection of results yearly generated from the same electronic medical record systems that all units use. These are then presented in one spreadsheet at the yearly two day session where leaders meet. Results are analysed, standards discussed and regional goals established. A written report is then distributed within each department. Initially the analysis was based on diagnosis related group codes, length of stay, diagnosis and operation codes. The first years this was done by attending heads of departments. Improvements the first years focused on increased out-patient treatment, less variation in treatment modalities and more similar in-patient length of stay. Since five years the analysis and reporting has branched into more specialised groups using a larger spread of data, including medical outcomes and patient satisfaction results. Some examples of improvement from the obstetrical group are reduced levels of rupture of the sphincter, blood loss at delivery and caesarean operation rates for healthy primigravidae. Other findings are less variation between units with time, for example the use of episiotomy.

An initiative from the Counties Insurance Office for self-assessment and audit of standards for safe deliveries to prevent new-born harm was followed by all units 2007 to 2009. This generated advice on implementation of a number of improvements, with a follow-up after six months. For example, the county hospital of Jönköping was given advice to implement a search tool for the departments 200-odd intranet guidelines, standardised risk grading of patients arriving at the delivery department, the inclusion of midwives in the county perinatal audit group and the training of the hospitals anaesthesiologists in neonatal resuscitation.

The regional board of Heads of Departments has introduced yearly compulsory training for co-workers in simulated practical obstetrical manoeuvres and foetal cardiotocography knowledge including assessment for obstetricians and midwives. All departments are now also implementing root cause analysis of report-required serious AEs.

These factors set the scene for large-scale measurement of AEs at delivery with the GTT method as potential for further enhancing obstetrical safety in south-east health-care region of Sweden. The homogenous nature of the participating delivery departments make it possible to aggregate results from more than 1000 deliveries. This gives the true value of rates, variation, grades and types of AE in obstetrical health care.

**Planning the intervention**

The intervention design was based on Langley’s [25] Change Concepts concerning enhancing the producer/customer relationship by focusing on the outcome to the customer and managing variation by developing operational definitions.

The improvement required the following set of activities:

1. Choice of a method for valid comprehensive assessment of obstetrical AE.
2. Anchoring the improvement with the regional board of Heads of Departments.
3. Limiting the improvements scope.
4. Training the abstractors in the chosen method, initiating and supporting their work.
5. Engineering methods for decision support and physician evaluation.
8. Data collection of reported levels of PSIs and reported AEs.
9. Reporting results to the regional board for decision on further deployment.
The first activity involved a literature search for non PSI patient orientated AE measurement methods in obstetrics. This was discussed initially with a fellow obstetrician knowledgeable in patient’s safety issues after a career as chief medical officer. He also had hands-on experience of the GTT method. A trigger tool nurse review method as opposed to other methods of evaluation was chosen because it is an effective method for measuring the overall rate of AE in a health care organisation [5] and adds the patient’s perspective. A trigger is a small movement that can release a larger action, such as the trigger of a firearm. In this context it implies the identification of easily and quickly found small criteria, such as haemoglobin count fall or change of level of care, that can point to the account of a possible AE more deeply imbedded in the notes. This choice and the projects feasibility was discussed with the chairman of the regional board of Heads of Departments in late 2010. The literature search exposed that a Perinatal Trigger Tool (PTT) for obstetrics had been developed by IHI some years earlier. The PTT method has 22 triggers instead of the 53 triggers defined in the GTT method. Four of the triggers in PTT are specific for obstetrics and not found in GTT. Two of the triggers are not applicable in Sweden. Very few studies have been done using the PTT methodology [26]. The GTT method was chosen because of its greater scope and current local knowledge. Two midwives and one obstetrician in the region are deeply knowledgeable of GTT and apply it now, together with other nurses and physicians, to measuring patient harm rates among admitted patients in their respective hospitals. Easy to follow instructions for training abstractors in the GTT methodology are provided in IHI manuals [6]. A Swedish translation is available [8]. This tool includes a list of known AE triggers as well as instructions for selecting records, training information and appendices with references and common questions. The tool provides instructions and forms for collecting the data needed to track two measures, i.e. AEs per 1000 patient days and AEs per 100 admissions.

The second activity was done in January 2011. There was some discussion if all departments in the region should partake with special uncertainty on behalf of the university department. As a consequence of the regions long standing experience of joint programs it was decided to measure obstetrical AEs with GTT in all departments during 2011.

The third activity entailed limiting the scope. The motive for limiting this project to delivered patients is the similar structure for these patients among participating departments, allowing aggregated results. The structure differs for antenatal and gynaecology patients. This would however give a local competency base in reviewing in each department that could subsequently be spread to gynaecology and antenatal care.

For the fourth activity it was necessary for two to three midwives from each department to be chosen as the abstractors. They were then trained in the GTT method by the knowledgeable obstetrician and midwives at a two day session on the 7th to 8th of February 2011. They had previously been responsible for training of hospital abstractors in the region. This session also included guidance to:

- Extracting the whole population from the electronic medical record system that all departments had been using for minimum eight years.
- Randomisation procedures using an easily applicable method in Excel.
- Using a joint website on www.qreflex.se to openly collect all data with a forum for debate and questions with mutual and expert answers.
- Using an Excel template with auto-sums for noting each reviewed cases results concerning length of stay, triggers and grade and type of AEs found. These where subsequently uploaded monthly to the joint website.

After training the midwives practiced on randomly selected cases for the last quarter of 2010 following the GTT protocol. After that they moved on to monthly reviewing of the months of 2011. This was done from April 2011 until March 2012. Two follow-up one-day sessions in Sep-
tember 2011 and March 2012 for repetition of elements of the GTT method, on-going analysis of incoming results and discussion of difficult cases. The GTT method entails deploying the patient’s perspective when reviewing for harm and not evaluating for organisational or human errors. Abstractors review each case separately for maximum 20 minutes. If they reach similar conclusions this is recorded. If not, a consensus discussion follows with referral to the physician if this is not resolved.

The fifth activity implied designing the joint website for suitable support. Cases with uncertain assessment and a random sample of cases from all units were reviewed by the author and when needed by the fellow consultant obstetrician responsible for the introduction training. A document of this physician assessment for each unit was generated and exchanged between the author and the units abstractors for verification.

Awareness of reality as opposed to opinion is a strong driver of improvement. The presupposed knowledge and assumptions of involved co-workers is therefore valuable to know in order to plan relevant AE reducing activities [9]. Before their introduction to the GTT method the abstractor midwives were requested, as the sixth activity, to anonymously estimate monthly rates of AE during delivery for 2010. They were also requested to seek the same information from their leaders, obstetricians and peers on returning home after the training session.

The period examined is all months of 2011 in the seven delivery departments in the south-east health-care region of Sweden. As the number of yearly deliveries at the seven units range from 900 to 2600, 10% of each month’s deliveries were randomly selected for trigger identification and AE assessment including grades and types. This implied that between 7 and 25 cases were reviewed monthly at each unit and the results aggregated to regional rates, variation, grades and types of patient harm. In order to equate evaluation of AE assessment and triggers in the region, a joint website on www.qreflex.se was set up to openly collect all data and including a forum for questions and answers. This was the seventh activity.

The eighth activity entailed that regional levels of PSIs and reported AEs were obtained from the regional board of Heads of Departments, to which the results and points of learning were reported on the 23rd March 2012 as this interventions final activity.

**Planning the study of the intervention**

There is no empirical data on the value of GTT measurements at the delivery department level. Continued GTT measurements and implementation of relevant improvements can be a patient-experience orientated force for further change. This study elucidated the attributed future value as evaluated by participating data-abstractor midwives after 10 months of reviewing. Value is here defined as benefit divided by effort. Benefit is defined as confidence in the results from important members of their departments as indicative of its power for accountability, effort selection and keeping track of results. Effort is defined as the resources applied with their perception of how reasonable the time spent reviewing and collecting data was, when considering their local context and parallel claims on time for improvement work.

The study design was an anonymous electronic quantitative questionnaire to all abstractor midwives supplemented with a semi structured interview of one abstractor midwife from each hospital. The interview questions were based on the results of the electronic questionnaire in order to further clarify uncertainties in these results.
Methods of evaluation

The web-based questionnaire to the abstractor midwives consisted of 14 questions with answers given as a visual analogue scale (VAS) [27]. The VAS is a psychometric response scale for measuring subjective characteristics or attitudes where the respondent indicates their level of agreement by indicating a position along a continuous line between two endpoints each consisting of a proposition. These were followed by a free text field titled "What comments would you like to add?" The introductory text informed the participating abstractors that the value of all improvement work is the benefit in proportion to the effort. The purpose of this study was to explore the attributed value of measuring AE with GTT as evaluated by participating data-abstractor midwives. They were also informed that answering was voluntary and the answers anonymous. The questionnaire was designed together with the academic supervisor and improvement coach and pilot tested in three cycles with a midwife, fellow students and the survey designer. The questionnaire consists of four categories, background knowledge, the GTT method, resources used and confidence in results (all questions and scales listed in the appendix).

The results of the electronic questionnaire showed that a deeper understanding concerning project design and value was necessary. This resulted in a semi structured telephone interview with four questions as guidelines for the interview. These were:

- How can the project design be improved?
- How sure are you of deploying the patient’s perspective during reviewing?
- How reasonable is the time spent?
- How confident are you with the results?

Eight interviews were done, at least one from each department. The answers were notated as text during the interview. This was then summarily analysed by the interviewer and author to develop a sense of what had been said. The author then analysed the text in more depth to discover repetitive statements that build categories [28, 29].

Analysis

The improvement intervention analysis starts with calculation of averages of estimates, with the standard deviation, for anonymous estimates of monthly rates of AE during delivery for 2010 for four groups of respondents. Inferences are also drawn directly from the line charts. The actual monthly patient harm statistics for AEs per 1000 patient days and AEs per 100 admissions are then presented with statistical process control charts, giving measures of the mean value, upper and lower control limits and numbers beyond limits. A Pareto diagram explains the frequency of different types of AE and a histogram gives the monthly frequency of grade E – I.

A table of the frequency and type of triggers identified is added for possible future evolvement of nurse review tools of AEs in obstetrical care.

The study of the improvement uses quantitative VAS results from respondents for each of the 14 questions are presented as a summary figure. A crude total value for the GTT method for measurement of obstetrical patient harm at this stage of implementation can also be attributed by dividing the average of questions 10 to 14 by the inverse average of questions 8 and 9.

The interview statements were analysed for qualitative content and resulted in a set of claims relating to project design improvements, certainty of deploying the patient’s perspective and the abstractor’s judgement of value.
Results

Improvement intervention outcomes

The primary improvement intervention question is how do measured overall actual rates of AE during delivery in 2011 at all departments in the south-east health-care region of Sweden compare to presupposed rates of AE among co-workers and to reported levels of PSIs and reported AEs and what are the variation, grades and types of AEs.

The results are presented in the following three categories:
- Presupposed rates of AE among co-workers as table 1 and figures 1 to 4.
- Actual rates with monthly variation, grades and types of AE during delivery in 2011 in figures 5 to 10, including the triggers found and their frequency in table 2.
- Reported levels of PSIs in table 3 and reported AEs in the accompanying text.

The presupposed rates of AE among co-workers are numerically summarised in table 1 and presented as monthly line charts for estimated AE rates 2010 for each of the groups of responders in diagram 1 to 4. The instructions given when collecting the data was for the responders to first consider their personal estimate of the highest percentage occurrence of AEs at delivery per patient day. This was then set at the top of a blank y axle on a blank line chart with each month of the year on the x axle. They then scaled the y axle evenly and added their estimates for all the other months of the year, creating a line chart of estimates.

<table>
<thead>
<tr>
<th></th>
<th>Average/100 pd</th>
<th>Standard deviation</th>
<th>Median/100 pd</th>
<th>Average/1000 pd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstractors</td>
<td>7.32</td>
<td>6.64</td>
<td>6</td>
<td>73.2</td>
</tr>
<tr>
<td>Obstetricians</td>
<td>9.21</td>
<td>6.25</td>
<td>8</td>
<td>92.1</td>
</tr>
<tr>
<td>Midwives</td>
<td>10.47</td>
<td>11.28</td>
<td>6</td>
<td>104.7</td>
</tr>
<tr>
<td>Leaders</td>
<td>11.28</td>
<td>11.14</td>
<td>8</td>
<td>112.8</td>
</tr>
</tbody>
</table>

The line charts (figure 1-4) illustrate one person’s estimate for each line. Inferences that can be drawn are:
- There is a relative consensus concerning rates.
- There are a small number of outliers in each group. The extreme values of some answers with up to 50% AE rates can possibly reflect on a full understanding of the question being answered.
- There is a tendency among midwives for an assumption that AE rates increase during the summer months, correlating to the yearly vacation period.
- Most responders visualise only a small variation between months. Some responders visualise an absolute flat rate of AEs.
Figure 1: Estimated AEs per 100 patient days 2010 for the 16 abstractor midwives - one person’s estimate for each line.

Figure 2: Estimated AEs per 100 patient days for 21 obstetricians at all participating units - one person’s estimate for each line.
Figure 3: Estimated AEs per 100 patient days for 24 midwives at all participating units - one person’s estimate for each line.

Figure 4: Estimated AEs per 100 patient days for 9 leaders of participating units (obstetricians and midwives) - one person’s estimate for each line.
The most common and important assumption illustrated in figures 1-4 is the lack of monthly variation. This is in contrast to the actual results for each department. Due to the effect of aggregation for regional rates the individual department’s variations are not visible in the reported results (figures 5 and 6). This is discussed in more detail under Limitations.

Control charts (U-type) with actual aggregated monthly rates of AEs 2011 for the region presented as AEs per 1000 patient days in diagram 5 and AEs per 100 admissions in diagram 6. The results are based on midwife reviews of 10% of all deliveries in seven departments in the region, totalling 1137 patients with 3819 patient days. There were 148 AEs in total. The mother and her child were considered one unit as they were cared for together. Length of stay was defined as the mothers discharge date minus admission date plus one. Review of new-born AE was only done when cared for at the delivery department or postnatal unit.

Figure 5: SPC control chart with AEs per 1000 patient days in the south-east health-care region of Sweden for mother and new-born at delivery.
There is contention as to whether mother and new-born should instead be considered as two separate admissions with their own length of stay and AEs. A subset of the reviews consisting of Jönköping County with stratified data for mother and new-born is presented in figure 7 and 8. AEs were counted separately for both groups; length of stay for the new-born is defined as discharge date home or to the neonatal unit minus date of birth plus one. Twins were considered two admissions and intraterine foetal deaths considered as no admission. In Jönköping County’s three delivery departments there were in total 37 AEs in 417 mother admissions with 1356 patient days and 16 AEs in 425 new-born admissions with 1202 patient days. Average AEs for these mothers in 2011 are 27.3 per 1000 patient days and 8.9 per 100 admissions. Average AEs for the new-borns in 2011 are 13.3 per 1000 patient days and 3.8 per 100 admissions. Approximately one third of AEs experienced at birth affected the new-born.
Figure 7: Line chart with AEs per 1000 patient days at delivery stratified for mother and new-born in Jönköping County of Sweden.

Figure 8: Line chart with AEs per 100 admissions at delivery stratified for mother and new-born in Jönköping County of Sweden.
Monthly frequency histogram (figure 9) of grades of AEs differentiated in categories E - I. These are based on the NCC MERP index [30] according to the following definitions. Categories A - D describe errors not resulting in harm and therefore not measured in the GTT method:

- Category E = AE contributing to temporary harm requiring treatment.
- Category F = AE contributing to temporary harm requiring hospital care or lengthened hospitalisation.
- Category G = AE contributing to permanent harm.
- Category H = AE requiring life support measures.
- Category I = AE contributing to death.

![Figure 9: AE grades with monthly frequency of categories E – I.](image)

A Pareto diagram (figure 10) shows the frequency of different types of AE aggregated for 2011 differentiated for AEs affecting the mother or the new-born. As 1137 cases were reviewed, corresponding to 10% of all deliveries in the region, the extrapolated values in the region are, for example, 300 cases of bleeding needing treatment and 10 cases of transfusion reaction.
26 of the GTT methods 53 triggers were used in this study. The type and frequency with which they were identified is listed in table 2. The triggers in themselves are not of fundamental interest, as these are only pointers in the process of measuring AEs. However as this tool has been previously not been used for describing delivery care this can be of value when training new groups of abstractors or developing new nurse review methods for obstetrical care. The triggers are in one of the six following categories.

- C = Care
- E = Emergency
- I = Intensive care
- M = Medication
- P = Perinatal
- S = Surgical
Table 2: Triggers identified by abstractors during reviewing of obstetrical care cases.

<table>
<thead>
<tr>
<th>Trigger</th>
<th>Description</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Transfusion</td>
<td>25</td>
</tr>
<tr>
<td>C2</td>
<td>Fall in haemoglobin count</td>
<td>37</td>
</tr>
<tr>
<td>C4</td>
<td>Heart failure or arrest</td>
<td>5</td>
</tr>
<tr>
<td>C8</td>
<td>Fall</td>
<td>1</td>
</tr>
<tr>
<td>C10</td>
<td>Readmission within 30 days</td>
<td>11</td>
</tr>
<tr>
<td>C11</td>
<td>Healthcare related infection</td>
<td>33</td>
</tr>
<tr>
<td>C12</td>
<td>Transfer to higher care level</td>
<td>17</td>
</tr>
<tr>
<td>C13</td>
<td>Treatment</td>
<td>20</td>
</tr>
<tr>
<td>C14</td>
<td>Other care trigger</td>
<td>19</td>
</tr>
<tr>
<td>E1</td>
<td>Return till Casualty &lt;48h</td>
<td>1</td>
</tr>
<tr>
<td>I4</td>
<td>Intubation/reintubation</td>
<td>1</td>
</tr>
<tr>
<td>M4</td>
<td>Glucose &gt; 3 mmol/l</td>
<td>2</td>
</tr>
<tr>
<td>M7</td>
<td>Antihistamine</td>
<td>2</td>
</tr>
<tr>
<td>M10</td>
<td>Antiemetic's</td>
<td>2</td>
</tr>
<tr>
<td>M12</td>
<td>Sudden cessation of drug administration</td>
<td>3</td>
</tr>
<tr>
<td>P1</td>
<td>Apgar score &lt;7 at 5 minutes</td>
<td>8</td>
</tr>
<tr>
<td>P2</td>
<td>Transfer mother/child</td>
<td>20</td>
</tr>
<tr>
<td>P3</td>
<td>Magnesium sulphate or terbutaline</td>
<td>9</td>
</tr>
<tr>
<td>P4</td>
<td>Serious lacerations</td>
<td>34</td>
</tr>
<tr>
<td>P5</td>
<td>Induced labour</td>
<td>23</td>
</tr>
<tr>
<td>S1</td>
<td>Return to operation</td>
<td>1</td>
</tr>
<tr>
<td>S2</td>
<td>Change of procedure</td>
<td>1</td>
</tr>
<tr>
<td>S4</td>
<td>Intubation/reintubation/CPAP in OR</td>
<td>1</td>
</tr>
<tr>
<td>S10</td>
<td>Change of anaesthesia mode</td>
<td>1</td>
</tr>
<tr>
<td>S11</td>
<td>Consultation in recovery area</td>
<td>1</td>
</tr>
<tr>
<td>S12</td>
<td>Post-operative complication</td>
<td>1</td>
</tr>
</tbody>
</table>
Reported levels of PSIs for 2011 as followed by the regional board of Heads of Departments relating to patient harm are listed in table 3.

Table 3: Reported levels of PSI in the region 2011.

<table>
<thead>
<tr>
<th>PSI</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perineal rupture</td>
<td>2.7% of admissions</td>
</tr>
<tr>
<td>Bleeding &gt;1000ml</td>
<td>4.2% of admissions</td>
</tr>
</tbody>
</table>

Reported AEs at delivery in the departments reporting systems are sporadic and so few in number that no comparison is reasonable. Leaders recognise that AE reporting is not an established method in delivery care in the south-east health-care region of Sweden.

Study of the improvement outcomes

The secondary study of the improvement explores the attributed value of measuring AE with GTT as evaluated by participating data-abstractor midwives. Value is here defined as benefit divided by effort. Benefit is defined as confidence in the results as indicative of its possible influence on accountability, effort selection and keeping track of results. Effort is defined as the resources applied.

The web questionnaire link was sent via email to fifteen abstractors as one had not taken part in reviewing after training. The response rate was 91%, with 10 questions answered by fourteen responders and an additional 4 questions answered by thirteen responders. The first question in background information was the number of years of employment. The abstractors were evenly spaced between 5 and 34 years of experience. The other thirteen questionnaire results are presented in figure 11. The bold numbers represent the number of respondents choosing a point on a VAS scale from 0-10 within the propositions defined at each end of the scale and rounded off to the nearest whole number. The propositions vary depending on the question.

<table>
<thead>
<tr>
<th>Background knowledge:</th>
</tr>
</thead>
<tbody>
<tr>
<td>2: How is your interest in/commitment to patient safety?</td>
</tr>
<tr>
<td>Very low                                0 1 2 3 4 5 6 7 8 9 10 Very high</td>
</tr>
<tr>
<td>1 2 5 6</td>
</tr>
<tr>
<td>3: What was your general knowledge of patient harm in childbirth before this project?</td>
</tr>
<tr>
<td>Very low                                0 1 2 3 4 5 6 7 8 9 10 Very high</td>
</tr>
<tr>
<td>1 1 4 2 2 4</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>The GTT method:</td>
</tr>
<tr>
<td>4: How do you assess this project's implementation?</td>
</tr>
<tr>
<td>Very poor                               0 1 2 3 4 5 6 7 8 9 10 Very satisfactory</td>
</tr>
<tr>
<td>1 1 4 4 1 2</td>
</tr>
<tr>
<td>5: How is it to find relevant information in the medical records, such as triggers?</td>
</tr>
<tr>
<td>Very difficult                          0 1 2 3 4 5 6 7 8 9 10 Very easy</td>
</tr>
<tr>
<td>1 2 4 3 3 1</td>
</tr>
<tr>
<td>6: How do you rate your knowledge of patient harm at birth now?</td>
</tr>
<tr>
<td>Very low                                0 1 2 3 4 5 6 7 8 9 10 Very high</td>
</tr>
<tr>
<td>1 1 3 4 3 1</td>
</tr>
<tr>
<td>7: How is it to decide whether there is patient harm?</td>
</tr>
<tr>
<td>Very difficult                          0 1 2 3 4 5 6 7 8 9 10 Very easy</td>
</tr>
<tr>
<td>1 3 2 6 1 1</td>
</tr>
</tbody>
</table>
Resources used (Effort):

8: How reasonable is the time you have spent reviewing?

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unreasonable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very reasonable</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9: What is it like for you to secure the time needed for reviewing?

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very difficult</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very easy</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Confidence in results (Benefit):

10: What is your confidence in the results? (ability to show the true rate of patient harm)

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very high</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11: What confidence do you estimate that improvement teams will have to the results?

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very high</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

12: What confidence do you estimate that leaders/managers will have to the results?

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very high</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

13: How important is your estimate that the results will have for future improvements to reduce patient harm at birth?

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very high</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

14: Which level of diffusion do you estimate that the GTT method to measure patient harm at birth will gain?

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very high</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 11: Results from abstractor questionnaire.

The abstractors are a group selected by their leaders for their interest in patient safety. This is confirmed by the results in question 2 with high results (the sum of all values is 128 out of a possible 140). Their prior knowledge of patient harm at childbirth is however more average (sum 82 of 140) as shown in question 3. This knowledge increases to sum 101 of a possible 130 after 10 months of partaking in this improvement, the result in question 6. Question 2 was however answered at the same time as all other questions, and it is difficult to clearly remember one’s previous status once exposed to the information in this improvement.

This improvement’s design was based on experience from training nurses for and collecting data from hospital GTT measurement in the region. Despite this being the first project in obstetrical care, the project design was acceptable with a score of 100 of possible 130 in question 4. Possible improvements are further explored in the interview.

The regions electronic medical records are conducive to finding relevant information, such as triggers. This is shown by a sum of 105 of 140 for question 5. The GTT method entails a decision on the presence of AE by the nurse reviewer. This is an experiential skill confirmed at the face-to-face collaborative meetings and evaluated at 10 months to be 78 of 140 in question 7. A distinct scope for further learning of this skill is present and reflects on the validity of these improvement results. The experience of previous GTT introductions is that there is a more than one year period before decisions on AE occurrence are stable, and thereafter are a real reflection of improvement and not only measurement deficits.
The assessment of the reasonability of resources used (the effort) by the responders themselves (question 8) or indirectly on their leader’s assessment (question 9) shows values of 111 of 140 and 107 of 140. The average score given is 7.9 (8.2; 7.6) for these two questions. This is a score indicating a reasonable use of resources. The inverse can be said to represent the cost, i.e. a score of 2.1.

Confidence in the results (the benefit) is illustrated by the average of questions 10 to 14 which is 7.0 (7.6; 7.1; 7.0; 6.2; 6.1). The score declines some from high confidence in the results as the focus moves further from the respondent themselves. Their score (question 10) is 107 of 140. For improvement teams in their departments (question 11) the score is 99 and for their leaders/managers (question 12) the score is 98. A measure of their assessment of potential spread is given in questions 13 and 14 with scores 80 and 75.

A crude total value for the GTT method for measurement of obstetrical patient harm at this stage of implementation can be attributed by dividing the average of questions 10 to 14 (benefit) by the inverse average of questions 8 and 9 (cost). The crude total value may be said to be 3.3 (7.0/2.1) implying a possible threefold return on resources applied, as evaluated by the midwife reviewers. No attempt to investigate other sources of value judgement has been made.

The free text commentary from the abstractor midwives is quoted as follows:

"It's fun and interesting to work with investigation and am amazed at how good the delivery care is in relation to the few harm we have discovered."

"Feels like what we got was not so revolutionary. Much of the harm is already included in our statistics."

"During the examination only found some harm which has not been diagnosed before. Do not know if the time of work weighs up the harm we find."

"Some questions are almost impossible to answer. Among other things, what I think other people think and how will I be able to predict the future of the instrument."

"Our leadership has not taken note of our measurements to date, it is difficult to judge how they will interpret it and what they think. A stronger interest here would be desirable to know what they think. I myself believe strongly in the method and hope that management will want to continue the measurements."

"It is interesting to examine the records, you learn a lot yourself. Also important for all of our colleagues to share the results."

"Fun to participate and get more insight into the process of patient injury. Have become more interested in patient safety. Comment to the survey; difficult to get the mark on the line to stay."

"Would have liked more structure to the organizing. QReflex is not an easy system. We have been given a task that takes time and burdens us extra in an already difficult situation."

"The answers only representative of inspection of obstetrical records. Simple method, but are the results useful? How? What about irrelevant "injuries" at the expense of the mental stress (the damage of a psychologically traumatic childbirth)? Hard to guess how wide spread/how useful the method will be."

The free text commentary reflects an interest in further work with patient safety, but an uncertainty of the value of GTT measures. This may be in part due to the large variation in cases reviewed (range 7 to 25) per month. Some smaller departments had long intervals between AE. An analysis from the free text commentary and also discussions at the face-to-face collaborative meetings is that the new and unique nature of the GTT measures is not fully understood or accepted. GTT measures are the first attempt in this region to analyse clinical results wholly from the patient’s perspective. PSI measures and AE reporting systems use an inside to outside approach. GTT starts with no preconception of what to find, observes for AE in a random sample
of cases from the patients viewpoint and is therefore an outside to inside approach. However, measuring itself does not make things better; it is only an incentive for effort selection and a means of keeping track of results. It is also a valid source for accountability.

The results of the electronic questionnaire showed that a deeper understanding concerning project design and value was necessary. This resulted in a semi structured interview exploring how the project design could be improved and how sure the respondents were of deploying the patient's perspective. Further knowledge of how reasonable the time spent was and how confident they were of the results was sought as an assessment of the abstractor's judgement of value.

A telephone interview by a third party interviewer was done. This person had no affiliation to the project or worksites.

Respondents identified an initial uncertainty and sense of difficulty with the project, particularly with practical issues in the beginning. Repetitive interview statements concerning project design improvement was the wish for more support on using the results template at the start. This could have been addressed with more instructions on how to use the web facilities and practising online at the training session. Even a second training session sooner after the first would be helpful. A greater clarity on the method came after the first follow-up when 3-4 months had been reviewed. The website was planned as a collaborative. This was not clear from the start and some wished for quicker answers on the website, especially by the experts partaking. All respondents confirmed the value of initial and repetitive face-to-face collaborative meetings as complement to the website.

The respondents were aware of the difficulty of deploying the patient's perspective based on their long experience as midwives, and somewhat of a challenge at the beginning. An attitude of “who could do better" would creep into the conversations from time to time. A midwife could see what happened as a natural part of her work and not as the patient’s expectation of freedom from injury. Some were aware of their status as co-workers when reviewing their colleagues. In the beginning there was a tendency to read the notes as a midwife looking for clinical information, instead of a reviewer looking for AEs. The methodology with paired review and active iteration of whose perspective to take, built confidence with time. The abstractors felt strongly that they could in time deploy the patient’s perspective while reviewing and refrained from the temptation of shielding their organisations.

Repetitive interview statements from the reviewers concerning concordant results when reviewing was that this was a very minor exception, and when this occurred, a quick discussion resulted in a common viewpoint.

The abstractors judgement of value was based on balancing confidence in results against time spent. A few respondents questioned the additional information brought to the table by this approach. The majority found that new perspectives were added but full value only realisable when relevant improvements have been implemented. The majority considered time spent to be reasonable and in same responses even limited. All respondents acknowledged general confidence in the validity of the results. Statements given included that this was considered a modern method. Other statements were that it is spreading and has strengths because of a clear structure. A few doubts concerning lack of psychological harm evaluation in the method were given (as this is not generally recorded and therefore not discovered by the GTT method). The abstractors interviewed responded that the GTT method is in general valuable to delivery departments.
Discussion

Summary
Awareness of reality as opposed to opinion is a strong driver of improvement. The presupposed assumptions of involved co-workers are therefore valuable to know to compare with reality. This improvements success is that it measured the actual overall rates of AE for a large subset of patients, their variations and patterns during delivery in the south-east health-care region of Sweden using a method applying the patient’s perspective. It has measured AEs at delivery for both mother and new-born in 10% of all deliveries in the region. It has the largest number of reviewed cases reported to date in this field. The measured results were found to be half of the presupposed values. One third of AEs affected the new-born. The measured results were double presently reported PSIs. Reporting of AEs at delivery has not been used in this region.

This improvements difficulty is that there are at this stage no observed changes in health care delivery or clinical outcomes and no active plan to effectuate this either.

The secondary study of the improvement explores the attributed value of measuring AE with GTT as evaluated by participating data-abstractor midwives. Value is here defined as benefit divided by effort. Benefit is defined as confidence in the results as indicative of its possible influence on accountability, effort selection and keeping track of results. Effort is defined as the resources applied. A crude total value extracted from a questionnaire answered by abstractors implied a possible threefold return on resources applied. The abstractors interviewed responded that the GTT method is beneficial to delivery departments when considering confidence in the results contra resources applied.

This study of the improvements difficulty is that it relies entirely on the viewpoint of the abstractors themselves and has no data on the views of leaders or improvement teams. The group selected for reviewing was done on basis of their interest in patient safety and probably biased towards the improvements value.

Relation to other evidence
There is a small body of other evidence relating to this improvement. Measuring AEs with the GTT method in obstetrical care has not been published. One other nurse review study is available (Forster), but it used other triggers and definitions and was a prospective cohort study. It included harm to mother and the new-born. AEs in this study where infrequent, but identified several common important quality problems. Others methods for AE identification are less effective as shown by Classen [11]. They are included in table 4 for the sake of comparison. PSI measures are more easily extracted from the databases but attached with limitations mainly concerning interpreting clinical outcomes from the patient’s perspective and probably some trust issues when regarded as a source of accountability. The PSI measures followed by the regional board of heads of department are listed in table 3.

The level of perineal rupture is identical in both the regional report and this improvement (2.7% of admissions). The level of bleeding differs with a level of 4.2% in the regional report and 3.0% in this study. This is due the regional report measuring this as a value in a database and this study reviewing if the bleeding was related to patient harm and requiring treatment. Self-reported AEs show very low rates and are deemed not reliable.
Table 4: Comparison with other results (pd = patient days, adm = admissions)

<table>
<thead>
<tr>
<th>Study</th>
<th>Method</th>
<th>AE rate</th>
<th>Number of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forster [20]</td>
<td>Nurse review</td>
<td>8/1000 pd</td>
<td>425</td>
</tr>
<tr>
<td>This study</td>
<td>Nurse review</td>
<td>38/1000 pd</td>
<td>1137</td>
</tr>
<tr>
<td>Mann [22]</td>
<td>PSI measure</td>
<td>9.2/100 adm</td>
<td>28000</td>
</tr>
<tr>
<td>Regional report</td>
<td>PSI measure</td>
<td>6.9/100 adm</td>
<td>11000</td>
</tr>
<tr>
<td>Linköping (unpubl)</td>
<td>Nurse review</td>
<td>6.6/100 adm</td>
<td>60</td>
</tr>
<tr>
<td>This study</td>
<td>Nurse review</td>
<td>13/100 adm</td>
<td>1137</td>
</tr>
<tr>
<td>Florea [21]</td>
<td>Self-reported</td>
<td>0.4/100 adm</td>
<td>6752</td>
</tr>
</tbody>
</table>

Limitations

Internal validity of the improvement

The process of estimating AE rates 2010 that was done by the four groups of respondents is fraught with uncertainty. Apart from the innate uncertainty of guessing, other design factors that added to this were:

- The values were set as per 100 patient days and not 1000 patient days as all other results were graded against.
- It was not clearly stated when giving instructions before estimating, that AEs affecting both mother and child should be included.

The process for selecting cases through randomisation, and identification of standardised triggers is fairly clear and not liable to much misinterpretation. There is no interdepartmental variation in correct number of cases selected and reviewed. There is little interdepartmental variation in the number of triggers identified, even if there is a tendency for departments with lower rates of AE to also have lower rates of trigger identification.

Possible or even probable sources of bias or imprecision are in the GTT methods design, with the need for an abstractor decision on the presence of an AE. This decision has three important sources of bias or imprecision:

- The quality of the medical records and ease of finding relevant information. Information can be added to the records in different places by different co-workers.
- The level of honesty among co-workers when recording details concerning abnormalities in the health-care provided.
- The willingness for abstractors to review cases conscientiously in order to reveal AEs.

No effort to thoroughly test internal validity for the regional rates of AE has been done. As this is an aggregated value from seven different departments this also includes the departmental biases and imprecisions. The process to standardise AE assessment was:

- The use of five cases that all abstractors evaluated during the training session.
- The joint website for collaborative questions and answers with physician resolution when needed.
- A second physician review of unclear cases and also of a random selection of cases presented in the results from each department.

These efforts to minimize the effect of stated limitations are what were possible within the scope of this Master’s thesis. A simple method to increase the internal validity would have been to send
a number of the same cases to all departments for their evaluation and follow the variation in decision on presence of AE. The certainty of deployment of the patient’s perspective while reviewing could probably have been increased by assuring some measure of standard methods. These could have included, for example that the reviewing room was a distance from the workplace, that the abstractors should not have been involved in patient care the same day before commencing reviewing and that a picture of a mother and new-born should hang above the computer. For this improvement there were no instructions concerning the external format for reviewing, only for the internal content of reviewing.

The actual rates of AEs in all participating departments (figure 12) illustrates a variation with department C showing higher values than average and department E showing lower values than average. How much this is related to the sources of bias or imprecision mentioned above is not possible to know within the scope of this improvement. The improvement design took this in to account by discussing only the aggregated rate for the region. This variation could also be pure coincidence as the GTT method stipulates minimum 12 months of reviewing, but preferably 24 months, before drawing conclusions.

Figure 12: The monthly AE rates per 1000 patient days for all seven participating departments.

**Internal validity of the study of the improvement**

The study of the improvement is fully based on the assessment of value by the involved abstractors. No effort to solicit the views of leaders or improvement teams has been made. The group selected for reviewing was done on basis of their interest in patient safety and probably biased towards the improvements value. The variation with few AEs in some units (figure 12) in 12 months of reviewing may have also reduced the sense of value. If there had been an availability of more resources, a focus group interview of leaders and improvement teams would have added more insights to the value of the improvement.

The questionnaire used is not validated; it was designed for this study because no specific validated questionnaire was available. The design process was done using a series of PDSA cycles described in Methods of evaluation.

The interviews were not transcribed and the content was analysed by means of an exploratory discussion on content between interviewer and author, followed by a deeper seeking of repetitive statements by the author in the notes made by the interviewer with subsequent categorisation.
External validity of the improvement
Factors that could affect generalizability are differences in settings of care in this region compared to other regions. However this data is extracted with the aspect that the delivery departments in the region are similar to many around the world, especially in developed nations. Comparisons must after all be made taking context into account. The knowledge gained is how do measured overall actual rates of AE during delivery in this region compare to presupposed rates of AE among co-workers and to reported levels of PSIs and reported AEs and what are the variation, grades and types of AEs. As the departments in the region are also very similar in nature, they are suitable for aggregating results, and comparing to other overall baseline measures. A limitation to other users of this improvement is that there is no measure on effectiveness as the improvement included no plan for continued use at the start.

External validity of the study of the improvement
The generalizability of the attributed value of measuring AE with GTT as evaluated by participating data-abstractor midwives is probably more limited to a Scandinavian setting. As value is here defined as benefit divided by effort, the local care and human resource factors are influential on the result given. Other health care systems have different decision hierarchies and different financial imperatives and incentives.

Interpretation
At the time of reporting the regional board of Heads of Departments of Obstetrics and Gynaecology has decided to cease further collective measurement of delivery care and continue with some unit based more limited measurements of gynaecological care. Their motivation was that the results were not in proportion to the effort spent considering present conditions. The Greek philosopher Heraclitus said that “Change is the only constant” and this improvements implementation is interpreted considering John Kotter’s eight-step change process. A professor at Harvard Business School and world-renowned change expert, he introduced his eight-step change process in the book "Leading Change" from 1995. These are:

- Step 1: Create urgency
- Step 2: Form a powerful coalition
- Step 3: Create a vision for change
- Step 4: Communicate the vision
- Step 5: Remove obstacles
- Step 6: Create short-term wins
- Step 7: Build on the change
- Step 8: Anchor the changes in corporate culture

The author’s expected outcome of this improvement was that building on the regions long experience of peer-reviewed PSI obstetrical care analysis, the next natural step in increasing patient safety and clinical quality was to:

1. Measure baseline regional overall AE rates using the patient’s perspective.
2. Identify a wider range of patient safety defects.
3. Implement specific improvement efforts based on this knowledge.
4. Implement the GTT method for keeping track of results.
5. Allow this to provide the necessary data for external accountability.

The last three of the expected outcomes did not materialise. Possible reasons for this are discussed below for each of Kotter’s steps:

Create Urgency: The initial implementation and design was based on dialogue, in part by e-mail, between the author and the regional board chairman. A more convincing dialogue with all mem-
bers of the board would have revealed some issues of insecurity about the real need for this improvement work. Significant more time and energy should have been spent on this. Redesign of the improvement could have been undertaken early on, instead of pressing on to find an absolute regional rate with this large number of patients. A smaller set of concurrent departments would have probably increased motivation. The impetus from national agencies to increase departmental patient harm measures could have been more clearly voiced. There was little interest in the regional board in discussing the method or results. Perhaps there was an apprehension of the implied consequences of the results and as yet not readiness for the reality that service organisations must be deeply inclined to improvement driven by customer experience.

Form a powerful coalition: The need for a powerful coalition was not identified. The idea was thought to have enough impact in itself. There are several influential people in the region that this improvement has been loosely discussed with during the improvement, but not enough energy was put into forming a defined coalition and building that team. The strong formal leadership that can lead change was not in place, and it did not suffice to just manage change. A smaller set of concurrent departments would have allowed a stronger leadership to have influence. One department started using a follow up system after 6 months. This included a template with the following five categories Month, Patient Injury, Action, Improvement project, Results. This template has provided an additional basis for the Patients Safety Groups work in that department. This hospital is also characterised by strong patient safety orientated leaders. The spread in the whole region of information concerning the stocktaking of possible improvement projects inspired by the evolving knowledge of rates of AE would have been supportive.

Create a vision for change: The GTT method is the first system in this context that starts with a blank slate and examines health-care results from the patient’s perspective. That this is the direction that service organisations are headed could have been elucidated much more clearly with examples from other industries. A clear vision could have helped those involved with what this improvement was asking them to do. Some of the free text commentaries in the questionnaires and comments from the data-abstractor midwives in the interview indicate that this vision was not clear or shared. They were not responsible for decisions about step 3-5 of the expected outcomes but obviously influential in discussions with their respective leaders. The patient orientated and outside in values in this improvement were not explicit and easily grasped.

Communicate the vision: Peter Senge [31] identified that an innovators focus is often glued on the innovation, losing sensitivity to the surrounding resistance. The worsening economic reality started biting during the year that the improvement was done and is a powerful competitor to more long-term goals. The author’s occupation with auditing removed focus from the importance of “walking the talk”. An initial plan to visit all sites regularly for auditing in situ was shelved due to change of work obligations. This plan would have given opportunity to communicate the vision more vigorously.

Remove obstacles: The scope of this improvement in this time frame left little space for identification of obstacles, both human and other. The author’s sense is that resistance grew during the year and that an engagement with the regional board after one year was too long a time frame. An attempt to get involved with the board earlier was declined, but this should have been seen as a call for re-evaluation of the improvement, instead of persevering. Keeping the leaders in the know as to progress along the way would have been helpful. A short status report could have been mailed. Now resistance grew to a point where reasoning was not the option and investments in point of view had already been made. A more comprehensive introduction in how to use the joint website would have reduced some initial frustration. Such an introduction could have included printed instructions, number to a help-desk and personal contact with prior users.
Create short-term wins: This improvement and its study design should have calculated on the possibility of no extension and designed in some activities early in the process that gave a very visible sense of value. Apart from the long-term goal set in the improvement, some sure-fire inexpensive short-term targets designed together with those involved and a change coalition team would have probably made things different. Examples are a simple certificate for the training that the data-abstractors took part in, one or two quick improvement projects on the first interesting and quick-fix AEs that were discovered. There are a few such numbers in the results, such as accidental incision of the foetus at caesarean section. There are identified stakeholders that could have been torch-bearers for a project to reduce accidental foetal incision. A myth arose in some circles that this improvement was just about the authors Master’s thesis and not about the long-term goal of introducing a next step in the health-care regions since many years evolving quality improvement mission. This could have been addressed with some short-term successes.

Build on the change: A number of small improvement cycles could have been designed from the start. The regional approach was probably too big as a first step. A kaizen design with continuous, incremental improvement could build on a commitment from closer more affiliated departments. Kaizen has as its base the philosophical belief that everything can be improved. It originated in Japan and the word translates to mean change (kai) for the good (zen). That this belief existed for AEs in obstetrical care should probably have been secured beforehand. Giving council that has not been asked for is often a sure road to disaster.

Anchor the changes in corporate culture: Corporate culture change is dependent on the success of previous steps. The mechanisms for this could have been thought through in the initial design. An early training of heads of departments and obstetrical leaders in understanding the GTT results in their own departments could have been part of that core culture change.

In this health-care region more than 500 midwives are employed. The financial cost for this improvement is the time spent by the abstractors. This is in total about 1 year’s salary for 1 fulltime midwife, divided between:
- 4 days for training and follow up collaborative meetings for 15 abstractors.
- Between 6 and 12 days for 2 (one hospital 3) abstractors for reviewing.

Infrastructural costs are for computer time and facilities. These are part of a hospital’s normal setup and not considered a cost here.

Quality improvement is about managing complexity, balancing aspects that are relational, motivational and power political. This improvement and its study have provided ample hands-on field study opportunities. The most important lessons learned are:
- Be careful about one’s own assumptions and analyse these carefully, preferably together with a “critical friend”. Don’t assume that other people’s values are in line with your own. Consider the hidden agendas that might exist.
- Think clearly about the whole process, including pre-work and long term implications.
- Start small.
- Study your improvement without advocacy.
- Engage experienced people early in the process (with the downside that you won’t learn as much the hard way).
- Calculate the time needed, double that, and double again.
- Perseverance sorts out most issues and hardships make experience. Good sailors aren’t made in calm waters.
- Read extensively, but keep notes of your reflections while reading.
- Don’t switch jobs in the middle of your thesis.
Conclusions
The overall practical usefulness of the improvement is that it has the potential for enhancing obstetrical safety in the south-east health-care region of Sweden by providing reliable baseline data on measured overall rates of AE seen from the patient’s perspective and also how these compare to presupposed rates of AE among co-workers and to reported levels of PSIs and reported AEs. Knowledge of the variation, different grades and types of AEs is certainly influential on selection of future improvement efforts.
Knowledge of using the GTT method in obstetrics and triggers found can influence further development of nurse review methods in obstetrics.
The study of the improvement based on value assessment by participating data-abstractor midwives provides a foundation for decisions by organisational leaders.
This improvement also supplies reliable data for accountability to the public, patients and healthcare financiers.
Further studies can be based on how selected improvement efforts affect rates of patient harm in obstetrical care, now that baseline rates, types and grades of AEs are known.

Funding
The Jönköping County Council has funded my participation in the Master’s program.

Acknowledgements
Aside from the Jönköping Academy’s academic support from teachers and other students, supervisors Karin Havemose and Göran Henriks, enablers for this thesis are three groups of people:

- The heads of departments for the seven obstetrical units in the region for arranging that the midwife abstractors (2-3 per unit) have the necessary time for training and reviewing.
- The dedicated midwives for their diligence and dedication to a delicate issue.
- Qulturum for statistical support, coaching and the use of QReflex. Urban Jurgensen for teaching the GTT method.
References


Appendix

The questionnaire consisted of the following questions and a VAS scale 0-10 for answers with the stated propositions as limits at each end.

Background knowledge:
1: How many years have you worked as a midwife?
   0 _________________________________ 40
2: How is your interest in/commitment to patient safety?
   Very low _________________________________ Very high
3: What was your general knowledge of patient harm in childbirth before this project?
   Very low _________________________________ Very high

The GTT method:
4: How do you assess this project's implementation?
   Very poor _________________________________ Very satisfactory
5: How is it to find relevant information in the medical records, such as triggers?
   Very difficult _________________________________ Very easy
6: How do you rate your knowledge of patient harm at birth now?
   Very low _________________________________ Very high
7: How is it to decide whether there is patient harm?
   Very difficult _________________________________ Very easy

Resources used (Effort):
8: How reasonable is the time you have spent reviewing?
   Unreasonable _________________________________ Very reasonable
9: What is it like for you to secure the time needed for reviewing?
   Very difficult _________________________________ Very easy

Confidence in results (Benefit)
10: What is your confidence in the results? (ability to show the true rate of patient harm)
    Very low _________________________________ Very high
11: What confidence do you estimate that improvement teams will have to the results?
    Very low _________________________________ Very high
12: What confidence do you estimate that leaders/managers will have to the results?
    Very low _________________________________ Very high
13: How important is your estimate that the results will have for future improvements to reduce patient harm at birth?
    Very low _________________________________ Very high
14: Which level of diffusion do you estimate that the GTT method to measure patient harm at birth will gain?
    Very low _________________________________ Very high