BMJ open

Experiences with the Global Trigger Tool in Denmark – a pilot study in a non-English speaking country

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-001324
Article Type:	Research
Date Submitted by the Author:	11-May-2012
Complete List of Authors:	von Plessen, Christian; Hillerød Hospital, Department of pulmonary medicine and infectious diseases Kodal, Anne; Hilleroed Hospital, Department of Anesthesiology Anhøj, Jacob; Danish Society for Patient Safety,
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health services research
Keywords:	AUDIT, Health & safety < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

Title

Experiences with the Global Trigger Tool in Denmark – a pilot study in a non-English speaking country

Authors

Christian von Plessen, MD (corresponding author)
Department of Pulmonary and Infectious Diseases
Hillerød Hospital
Dyrehavevej 29
DK 3400 Hillerød
cple@hih.regionh.dk
Fax,: +45 4829 3935

Anne Marie Kodal, RN
Department of Anesthesiology
Hillerød Hospital
Dyrehavevej 29
DK 3400 Hillerød

anmak@hih.regionh.dk

Jacob Anhøj, MD, DIT
Danish Society for Patient Safety
c/o Hvidovre Hospital
P610 Kettegård Alle 30
2650 Hvidovre

jacob.anhoej@regh.dk

Word count:

Contributions

CvP, JA and AMK designed the study and collected data. CvP drafted the manuscript, JA and AMK revised and approved. JA performed the statistical analyses and produced the graphs.

Copyright

Abstract

Objectives: To describe experiences with the Global Trigger Tool (GTT) in Denmark and suggest ways to improve the performance of GTT review teams

Design: Observational study

Setting: The measurement and monitoring of harms is crucial to campaigns to improve the safety of patients. Increasingly, teams use the GTT to review patient records and measure harms in English and non-English-speaking countries. Meanwhile, it is not clear how the method performs in so diverse settings.

Participants: Review teams from five Danish pilot hospitals of the national Danish patient safety campaign

Primary and secondary outcome measures: We collected harm rates, background and anecdotal information from five pilot hospitals currently participating in the Danish Safer Hospital Programme. An experienced reviewer categorized harms by type. Reported patient safety incidents (PSI) stem from the national Danish Database. We plotted harm rates as run-charts and applied rules for the detection of patterns of non-random variation.

Results: The hospitals differed in size but had similar patient populations and activity. The average harm rate for all hospitals was 60/1000 patient days. The median monthly harm rate ranged from 32 to 91 harms per 1000 patient days. Overall, 96% of harms were temporary. Infections, pressure ulcers procedure related and gastrointestinal problems were common. PSIs varied between 3 and 12/1000 patient days. Teams reported differences in training and review procedures such as the role of the secondary reviewer.

Conclusions: We found substantial variation of harm rates. Differences in training, review procedures and documentation in patient records likely contribute to these variations. Training reviewers as teams, specifying the roles of the different reviewers, training records and a database for findings of reviews may improve the application of the GTT.

Background

Patients run a high risk of being harmed during hospital admissions. Harms occur in up to 10 % of hospitalizations and can cause death, permanent or temporary disability.[1] For patients and health care workers, these harms and the underlying flaws of their health care systems that permit them to happen are deeply upsetting and completely unacceptable.

To improve the safety of patients, national and regional campaigns have been carried through [2-5] or are ongoing.[6] Improvements were achieved in some areas such as reductions of catheter-related blood stream infections.[7] However, system wide progress is slow [8] and improvements are often limited to particular medical conditions or institutions. Indeed, a recent study [9] from the state of North Carolina, an active participant in large scale patient safety initiatives, concluded that overall rates of harm during 2002-2007 were not reduced. Thus, the challenge to improve the safety of patients in hospitals remains and specific and sensitive measures of harms are needed to assess and monitor the effects of changes to make hospitals safer.

In Denmark, the Operation Life campaign during 2006-2008 focused on patient safety in intensive care and during surgery. An estimated 1654 fewer patients died in the Danish population of 5.5 million during the campaign.[10, 11] In 2010, another campaign the Safer Hospital program (www.sikkerpatient.dk/fagfolk/patientsikkert-sygehus.aspx) was launched at five pilot hospitals to reduce mortality by 15 % and harms by 30 % through the implementation of 12 care bundles in the participating hospitals. The hospitals are required to measure and report harms.

Meanwhile, a gold standard for the measurement of harms does not exist. Methods like voluntary reports only detect a small fraction of harms,[12] chart reviews have low interrater reliability [13] and are very time consuming and so are direct observations of health care processes.[14] Studies comparing different methods of harm detection have found very little overlap of the detected harms.[15] Therefore, complete estimates of the incidence of harms probably require the combination of different methods. Meanwhile, such an approach is time consuming and results are often delayed, which is unsuitable for patient safety campaigns in which frequent and regular measurements of harms are needed to evaluate and monitor the effects of interventions and organizational changes.

The Global Trigger Tool (GTT) has been developed for the purpose of monitoring harms at low cost.[16] Harm in this context is defined as an "Unintended physical injury resulting from or contributed to by medical care that requires additional monitoring, treatment or hospitalization, or that results in death."[16] Thus the tool measures factual harm to patients while errors not leading to harm, near errors and errors of omission, are not included. A GTT-review is a trigger based chart audit of closed patient charts. Two reviewers, usually nurses or pharmacists, each review a limited number of randomly chosen charts with a given set of 56 triggers or hints of errors. The finding of such a hint triggers an investigation into whether and, if so, how severely, a patient actually has been harmed. Review time is limited to 20 minutes per admission. Finally, the two reviewers compare their conclusions and eventually a supervisor, usually a physician, judges in cases of disagreement whether the conclusions were appropriate. The number of harms is then expressed as a rate, e.g. harms per 1000 bed days. It has been suggested that GTT teams need a limited amount of training and practice to achieve good levels of reliability to identify harms.[17-19] The feasibility of the method invites for rapid adoption in health care systems around the world where practical ways to measure harms are much in demand. Nevertheless, experiences with the GTT in non-English speaking countries are limited. Thus careful calibration of the instrument and the review team that uses it is warranted to avoid evaluating the safety performance of hospitals with imprecise measurements.

A team of Danish experts translated the GTT to Danish from the English and a Swedish version.[20] The tool was tested in four hospitals in different health regions.[21] The harm rate in these hospitals was around 20/1000 bed days. A recent report of harms to Danish patients with cancer found a rate of 68/1000 bed days.[22] Notwithstanding these variable rates, policy makers advocate the widespread implementation of GTT reviews in Danish hospitals. Meanwhile in our opinion, it is not sufficiently clear, how the tool performs in the hands of Danish review teams.

Therefore, we present in this study our experiences with the GTT in a non-English speaking country and suggest ways to enhance the performance of GTT review teams and contribute to accurate measurements of harms.

Methods

GTT review

A team of two independent reviewers, usually experienced nurses, and one supervisor, usually a physician, reviewed a random sample of 10 admissions twice a month. Closed admissions of patients of at least 18 years of age were and of at least 24 hours duration were eligible. The date of discharge was the index date. The GTT teams reviewed all available information from the admissions, i.e. physician and nurse notes, medication orders and history as well as results of laboratory and other diagnostic tests. Triggers and harms and the severity categorization were recorded on standardized trigger sheets and then transferred to spreadsheets and stored locally. The teams classified the severity of harms according to the National Council for Medication Error Reporting and Prevention Index. There is no shared database for the five hospitals.

Data collection and analysis

The project managers of the Safer Hospital Program and the GTT review teams at the five hospitals supplied background information on the hospitals from hospital administrative systems and the GTT sample populations.

We calculated the monthly harm rate as the total number of harms divided by the total number of patient days multiplied by 1000. The harm rate was then plotted on a run charts and analysed using four run chart rules for the detection of patterns of non-random variation like shifts or trends.[23] We collected and managed the data in Microsoft Excel v. 2003 and produced the statistical analysis and graphs in R Statistical Software v. 2.13.1.

For the purpose of this study, two nurses with experience from over 400 GTT reviews retrospectively categorized harms found at the five hospitals into categories also used by Classen et al.[24] They added gastrointestinal complications and pressure ulcers as categories because these types of harms were common (Figure 4). Each harm was assigned to only one category.

The number and severity of reported patient safety incidents (PSI) stems from the national patient safety incident database (www.dpsd.dk). In Denmark reporting of PSIs is mandatory, confidential and sanction free for health care personnel. Risk managers at

the five hospitals classified the severity of PSIs in the Danish Safety Database into mild, moderate, severe and fatal.

The regional ethical committee deemed an ethical review of the study unnecessary.

Results

Background data of the five pilot hospitals

The five hospitals, one in each Danish health region, have between 142 and 391 beds. All hospitals have departments of internal medicine, orthopaedic and general surgery as well as of obstetrics and gynaecology. All hospitals use electronic patient records, but to a varying degree part of the documentation, such as nursing notes, are on paper. For further information on the hospitals see Table 1. The populations of patients at the five hospitals were similar with regard to age and gender distribution (Table 1). Two hospitals, Næstved and Hillerød, had prior to the Safer Hospitals program been using the GTT for reviewing patient records for one and two years respectively. All teams started reviewing for the safety campaign in May 2010 and reviewed records retrospectively from January 2010.

Table 1: Background information on the five hospitals (2010)

	Hillerød	Horsens	Kolding	Næstved	Thy-Mors
Discharges	60098	30377	27526	28677	11836
Percent females	62	59	61	55	55
Patient days	231978	108060	90710	113353	49711
Outpatient visits	262547	212899	124184	184374	65165
Employees	3163	1367	1507	1668	689
Hospital Standardized Mortality Rate	95	97	96	112	100
Reported safety incidents	2736	365	923	1182	223
Reported patient safety incidents per 1000 patient days	12	3	10	10	4

Experiences with the implementation of the GTT

Reviewing patient files with the GTT was new to three of the five teams. At Hillerød Hospital, the team had reviewed files of deceased patients since 2008.

In May 2010, a GTT team from each of the five hospitals participated in a seven hour training with experts in the method from Denmark and abroad, which included use of the Danish GTT manual, frontal teaching, review of three training records per team and plenary discussions of the findings (Table 2). Only the review team at Hillerød had in 2008 received a similar training. All teams used nurses as primary reviewers and physicians as supervisors. All teams received on-site expert coaching with reviews of 10 or more records, up to three times. Furthermore, all teams participated in two full day network seminars during the study period. All teams started reviewing patient records for the measurement of their baseline in May 2010 and retrospectively reviewed records from January to May 2010.

The compositions of teams changed between zero (Hillerød and Næstved) and three times (Thy) during the study period. Review intervals at the hospitals varied between twice monthly, monthly and irregular. The role of the physician in the review team varied from only judging cases where the primary reviewers were in doubt or disagreed (Hillerød) to identifying all harms based on triggers found by primary reviewers (Horsens). Table 2 shows differences in review procedures at the hospitals.

Table 2: Characteristics and review procedures of GTT teams at five hospitals

	Hillerød	Horsens	Kolding	Næstved	Thy-Mors
Team characteristics					
Number of physicians	2	1	1	2	1
Number of nurses	3	3	4 later 3	4	3
Number of changes in team	0	1 (new physician)	1	0	2 (new nurses)
Review interval	Every fortnight	Monthly	Monthly (two half days)	Monthly	Variable
Training					
Hours of training	14#	7	7	7	7
Site visits by Danish expert (days)	1	3	1	4	2
Complete team present during site visit	-	-	+	-	-
Number of records reviewed together with expert	<10	>10	>10	>10	>10
Review procedures					
Whole team meets for reviews	+	+	Since Jan. 2011	-	-
Physician acts as judge (J) in cases of disagreement or reviewer (R) based on triggers	J	R	J	J	J
Records entirely electronic	-	+	+	-	+
Dedicated person responsible to find records	+	-	+	+	-
Secretary plots triggers and harms	-	-	-	-	+

Team also trained by national expert in 2008

Anecdotal information about GTT reviews

At Næstved, the team sampled 24 records (in case some were incomplete) each month and sorted them in the order of the date of the admission and reviewed the first 20 records. Thus the sample became biased towards admissions in the earlier part of the month. Moreover, the team initially reviewed only admissions to the last department of a hospital admission. Thus they did not find harms that, for example, occurred during an admission to the intensive care unit earlier during the hospital stay. These errors were accidentally discovered during a site visit and the team did a new review for the period. The team at Kolding discovered after three months that their sampling procedure excluded admissions that had an appointment for ambulatory follow up after surgery, and they decided to discard the first three months from their baseline.

GTT findings

In total, 688 adverse events were identified in 11501 patient days, i.e. the overall average harm rate was 60 per 1000 patient days. The median monthly harm rate ranged from 32 to 91 harms per 1000 patient days (Figure 1). One hospital, Hillerød, had a significant upward shift of the harm rate, between September and October 2010, identified by two runs of seven data points below and above the median respectively.[23]

Overall, 96% of harms were temporary (grades E and F). However, the severity distribution varied between hospitals (Figure 2). Notably, the hospital with the highest harm rate (Hillerød) also had the highest proportion of grade E harms. Common types of harm were infections, procedure related, pressure ulcers and gastrointestinal problems. The classification of harms in the category other varied nearly six-fold across hospitals (Figure 4).

Patient safety incidents

The reporting of PSIs among the five hospitals differed between 3 and 12 per thousand patient days. Meanwhile the distributions of PSIs by consequence were almost identical (Figure 3).

Discussion

We were surprised by the wide variation in the detection of harms in the five hospitals. The hospital with the highest harm rate had the highest number of harms of the lowest severity grade E. The hospitals differ in size but the patient populations and activity levels are similar. Also voluntarily reported PSIs were similar in distribution and type. We found differences in the training and the experience of the review teams. Review procedures were standardized for all teams but there were differences in the roles of nurses and physicians in the review process.

Other studies have also found variation of harms across hospitals. Naessens et al.[19] in a study of 1138 admissions to three academic health centres in three states of the US found a variation of harms by hospital between 19,4 % and 37,9 % of admissions. In a study of surgical harms the variation was between 5 % and 35 %,[25] Sharek et al.[26] observed harm rates between 0,18 and 1,28 per patient in 749 admissions to 15 newborn intensive care units in the US and Canada and Resar et al.[27] in 62 intensive care units in the US measured between 3,2 and 27,36 harms per 100 days. Thus significant variations of GTT findings seem to be common.

Several factors could explain the variation in rates of harm. First, they can be caused by real differences in the safety of the clinical processes at the hospitals. However, it seems unlikely that such differences should cause as much as a three-fold variation of harm rates given the similar patient populations at the five hospitals and the homogeneity of the Danish health care system in general. Second, differences in case mix at the hospitals could cause the variations. However, the compositions of the patient populations in the five hospitals are similar. We even found that the hospital with the highest mean age and the highest hospital standardized mortality rate had the lowest rate of harms. Third, the documentation of triggers and harms probably varies across hospitals and the intensified focus on safety could increase the documentation of harms in different ways. Indeed, the hospital with the highest PSI rate also observed a significant increase of harms, which we take as sign of a change in the culture of reporting and documentation at that hospital. Furthermore we observed differences in the way types of harms were coded, especially in the “other” category. Finally, the hospitals use different electronic and paper systems. On the other hand, the training of health care personnel in Denmark is similar in different parts of the country and so are

rules for documentation in the patient records. Concluding on this argument, we assume that the culturally influenced practice of documentation did contribute to the differences in rates of harm across hospitals. Fourth, differences in the training and the experience of the review teams influence the subjective process of judging harms in any record review.[13] The team that found the most harms had attended two training seminars; it also was stable, reviewed regularly with the whole team twice a month and had the longest experience with the GTT. Interestingly, the harm rate at their hospital demonstrated a significant shift in the months after the second training seminar. Fifth, the teams conducted the review processes in slightly different ways. Most importantly, the roles of nurses and physicians varied. The role of the physician in the review team that found the highest harm rate was to judge in cases of disagreement while physicians in the other teams themselves identified harms. We assume that nurses are more prone to register harms of lower severity, while physicians might consider them insignificant. This interpretation is supported by our finding that the variation of harm rates was greatest in the least severe category "E".

Thus in our opinion, the experience of the GTT teams and the way they perform the reviews strongly contributed to the differences in harm rates in the five hospitals. Moreover, differences in the ways harms are documented in the patient records influence the number of harms the GTT team can find. We did not expect these factors to be so important because the GTT reviewing was implemented according to published recommendations [16, 18] and was guided by some of their authors. Moreover, all the teams had attended a GTT network meeting with national and international experts and received site visits by a national expert. However these precautions, it seems, were not sufficient. A recent study from Sweden [28] supports these conclusions. Experienced review teams from five hospitals reviewed 50 patient charts from one of the five hospitals. Harm rates ranged from 27,2 to 99,7 per 1000 patient days and the pair wise interrater reliability of the five teams ranged from a kappa value of 0,26 to 0,77.

Our experiences with the GTT in a non-English speaking country have implications for the implementation of the method in other settings and we suggest the following interventions to improve the implementation of the GTT in new settings:

- Secure that the review team is trained as a team

- Specify of the roles of the reviewers during the reviews to avoid over-/underestimation of especially harms of lesser severity depending on professional background
- Test review teams' abilities to find harms with a set of training charts to estimate their "sensitivity" before routine monitoring is instituted
- Define a minimum number of patient charts that the team should have reviewed before monitoring harms routinely
- Perform reviews with all team members present
- Ensure a structured review process, i.e. a space where the team can work without interruptions, regular time intervals between reviews to keep team "in shape"
- Implement a common database with individual patient data to allow for re-examination of reviewed charts

In conclusion, the GTT is a practical tool to monitor harms that needs careful calibration when using it in new settings. Thus health care staff and policy makers should be aware of the need for sufficient training and retraining of the review teams. Further research should address the training of teams and the evaluation of their performance.

Competing interests

CvP leads the current Safer Hospital campaign at one of the participating hospitals. AMK is a member of the GTT review team at the same hospital. JA is an advisor to the patient safety campaign at the national level. None of us has any financial interests related to this study.

Acknowledgements

GTT teams at the five hospitals: Hillerød: Lisbeth Dammegaard, Anne Marie Kodal, Minna Nielsen, Anne Marie Schlüter, Mette Østergaard. Horsens: Inge-Lise Johansen, Tina Kjær Pedersen, Tom Krintel Petersen (anæstesiologisk afdeling), Kiss Ruben Larsen. Kolding: Helle Guldborg Nielsen, Lynge Kirkegaard, Pernille Kølholt Langkilde, Pernille Lennert, Anita Pedersen, Pia Schrøder Petersen. Næstved: Arne Cyron, Helle Fugl, Kim Garde, Annette Knakkergaard, Nina Pedersen, Trine Zachariassen. Thy Mors: Connie Elbeck van der Koij, Mona Kyndi, Mie Landbo, Bente Ringgaard, Hans-Jörg Selter, Lisbeth Thomsen

Project managers at the five hospitals: Anne Balle Larsen, Søren Brogaard, Marianne Frandsen, Mona Kyndi Pedersen, Kirsten Løth Lysdahl, Søren Schousboe Laursen, Maria Staun, Inge Ulriksen

Funding

CvP received a grant from the Danish Society for Patient Safety to write this article.

Legends to figures

Figure 1: Rates of harm

Run charts showing monthly rates of harms measured with the Global Trigger Tool. The curve shows the harm rate expressed as the number of adverse events per 1000 patient days. The horizontal line is the median harm rate. At Hillerød a shift occurs between September and October 2010 identified by two runs of seven data points on the same side of the median.

Figure 2: Harms by Severity

The dot plots show the relative distribution of severity of harms in categories E – I, where E and F are temporary, G – H permanent harms and I death. Overall, 96% of harms were temporary.

Figure 3: Patient Safety Incidents by Consequence

The dot plots show the relative distribution of patient safety incidents reported to the Danish national database by consequence. *Minor* and *moderate* represent no and temporary harms, *major* permanent harms. Overall 96% of the incidents are temporary.

Figure 4: Harms by Type

The dot plots show the relative distribution of harms by type. VTE = venous thromboembolism.

References

1 de Vries EN, Ramrattan MA, Smorenburg SM et al. The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care* 2008 Jun;**17**:216-23.

2 Berwick DM, Calkins DR, McCannon CJ et al. The 100,000 lives campaign: setting a goal and a deadline for improving health care quality. *JAMA* 2006 Jan 18;**295**:324-7.

3 Benning A, Dixon-Woods M, Nwulu U et al. Multiple component patient safety intervention in English hospitals: controlled evaluation of second phase. *BMJ* 2011;**342**:d199.

4 Benning A, Ghaleb M, Suokas A et al. Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation. *BMJ* 2011;**342**:d195.

5 McCannon CJ, Schall MW, Calkins DR et al. Saving 100,000 lives in US hospitals. *BMJ* 2006;**332**:1328-30.

6 NHS Scotland. Scottish Patient Safety Programme. Available at: http://www.patientsafetyalliance.scot.nhs.uk/programme. Accessed Nov. 3rd, 2011.

7 Pronovost P, Needham D, Berenholtz S et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 2006;**355**:2725-32.

8 Vincent C, Aylin P, Franklin BD et al. Is health care getting safer? *BMJ* 2008;**337**:a2426.

9 Landrigan CP, Parry GJ, Bones CB et al. Temporal trends in rates of patient harm resulting from medical care. *N Engl J Med* 2010;**363**:2124-34.

10 Operation Life Danish Society for Patient Safety. Operation Life Slutrapport. 2009.

11 Dansk Selskab for Patientsikkerhed. The Danish Safer Hospital Programme. 2011; Available at: http://www.sikkerpatient.dk/about-sikker-patient/the-danish-safer-hospital-programme.aspx. Accessed Nov. 3rd, 2011.

12 Sari AB, Sheldon TA, Cracknell A et al. Sensitivity of routine system for reporting patient safety incidents in an NHS hospital: retrospective patient case note review. *BMJ* 2007;**334**(7584):79.

13 Thomas EJ, Lipsitz SR, Studdert DM et al. The reliability of medical record review for estimating adverse event rates. *Ann Intern Med* 2002;**136**:812-16.

14 Buckley MS, Erstad BL, Kopp BJ et al. Direct observation approach for detecting medication errors and adverse drug events in a pediatric intensive care unit. *Pediatr Crit Care Med* 2007;**8**:145-52.

15 Naessens JM, Campbell CR, Huddleston JM et al. A comparison of hospital adverse events identified by three widely used detection methods. *Int J Qual Health Care* 2009;**21**:301-7.

16 Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events. IHI Innovation Series White paper. 2009.

17 Classen DC., Lloyd RC., Provost L et al. Development and Evaluation of the Institute for Healthcare Improvement Global Trigger Tool. *J Patient Saf* 2008;**4**:169-77.

18 Adler L, Denham CR, McKeever M et al. Global Trigger Tool: Implementation Basics. *J Patient Saf* 2008;**4**:245-49.

19 Naessens JM, O'Byrne TJ, Johnson MG et al. Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool. *Int J Qual Health Care* 2010;**22**:266-74.

20 Strukturerad journalgranskning för att identifiera och mäta förekomst av skador i vården enligt metoden Global Trigger Tool, Institute for Healthcare Improvement Innovation series 2007, Svensk översättning och anpassning 2008.

21 Rabøl LI, Jølving LR, Bjørn B et al. . Systematisk journalgennemgang med IHI's Global Trigger Tool - Erfaringer fra afprøvning på danske sygehuse. 2008.

22 Patientsikkerhed i kræftbehandlingen - journalgennemgang med Global Trigger Tool. Rapport nr. 3/2010. 2010.

23 Perla RJ, Provost LP, Murray SK. The run chart: a simple analytical tool for learning from variation in healthcare processes. *BMJ Qual Saf* 2011;**20**:46-51.

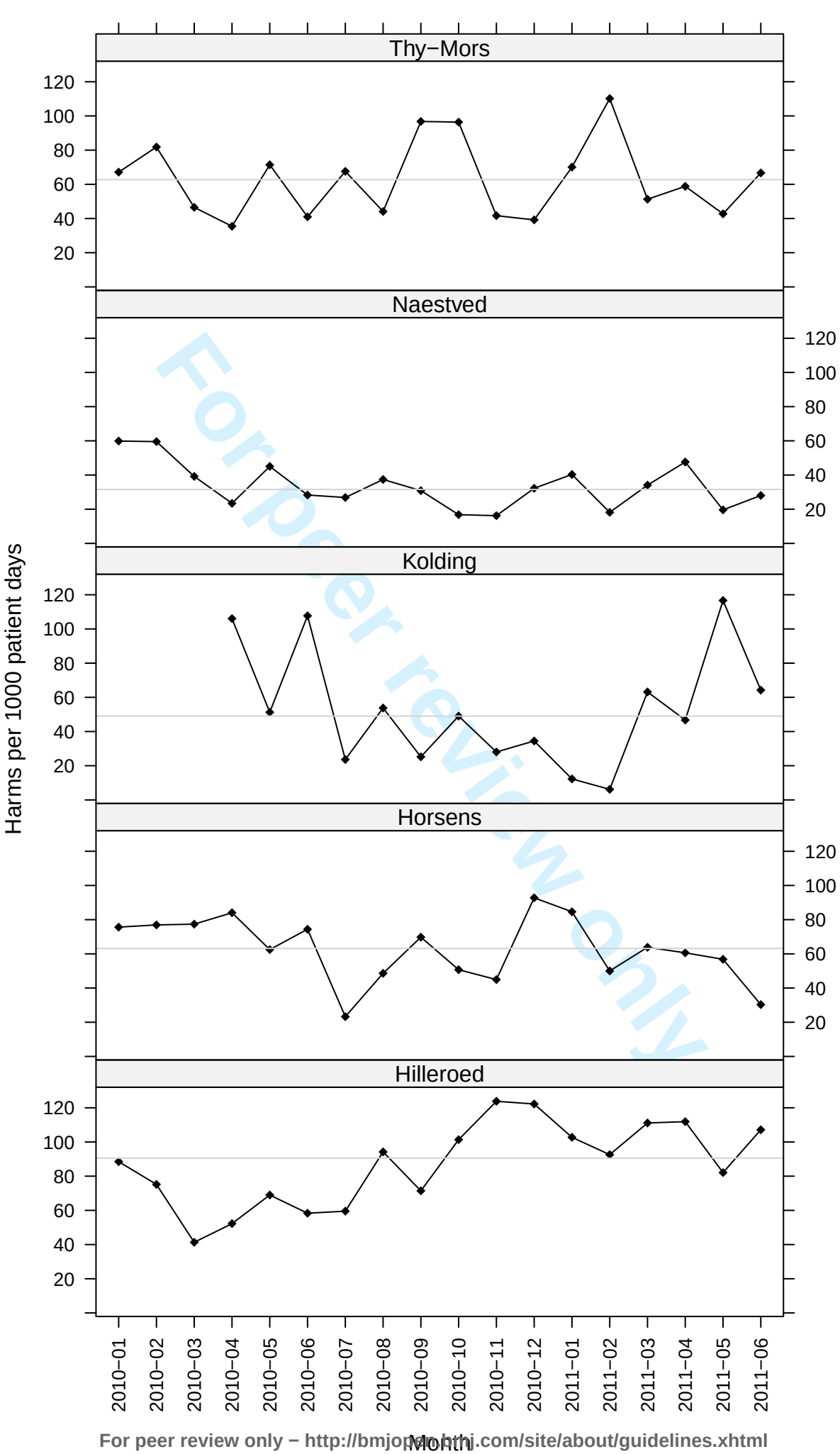
24 Classen DC, Resar R, Griffin F et al. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. *Health Aff (Millwood)* 2011;**30**:581-9.

25 Griffin FA, Classen DC. Detection of adverse events in surgical patients using the Trigger Tool approach. *Qual Saf Health Care* 2008;**17**:253-8.

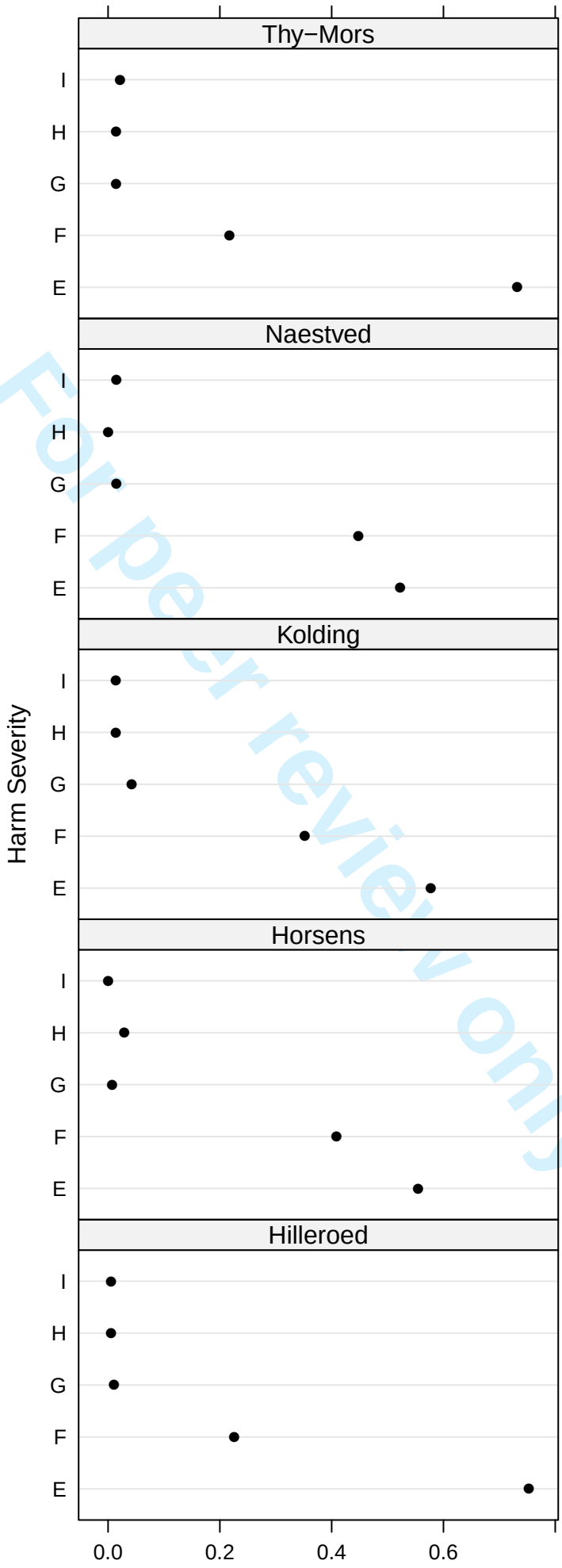
26 Sharek PJ, Horbar JD, Mason W et al. Adverse events in the neonatal intensive care unit: development, testing, and findings of an NICU-focused trigger tool to identify harm in North American NICUs. *Pediatrics* 2006;**118**:1332-40.

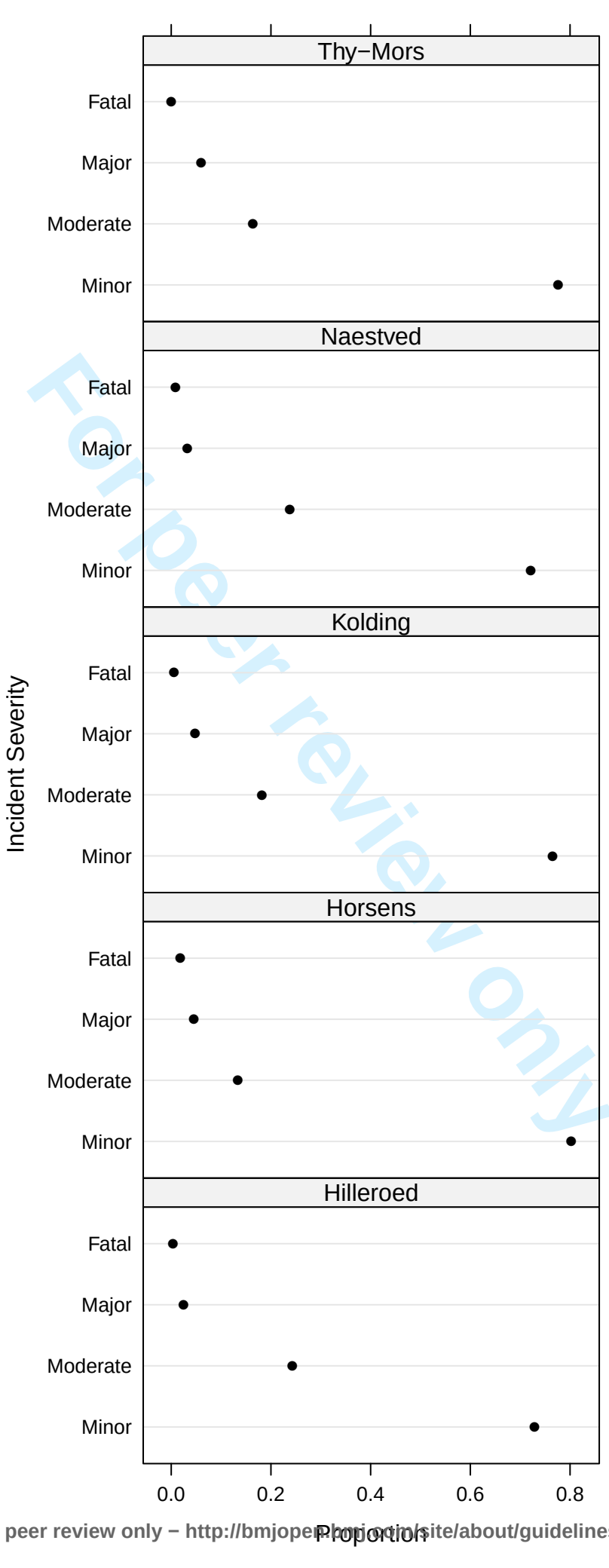
27 Resar RK, Rozich JD, Simmonds T et al. A trigger tool to identify adverse events in the intensive care unit. *Jt Comm J Qual Patient Saf* 2006;**32**:585-90.

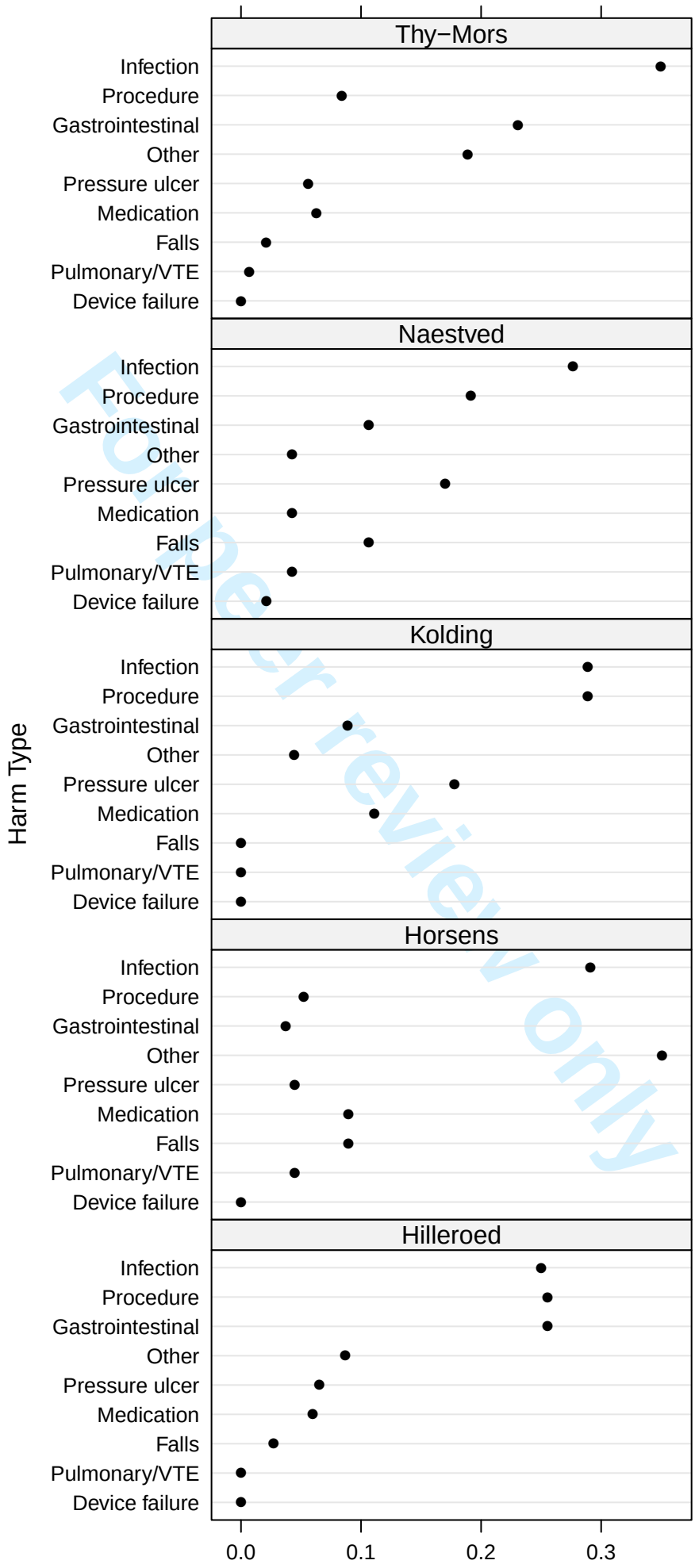
28 Schildmeijer K, Nilsson L, Årestedt K et al. Assessment of adverse events in medical care: lack of consistency between experienced teams using the global trigger tool. *Qual Saf Health Care* 2012;**21**:307-14



Harms by Severity







BMJ open

Experiences with the Global Trigger Tool in Denmark – a pilot study in a non-English speaking country

New Title:
Experiences with Global Trigger Tool reviews in five Danish hospitals – an implementation study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-001324.R1
Article Type:	Research
Date Submitted by the Author:	24-Jul-2012
Complete List of Authors:	von Plessen, Christian; Hillerød Hospital, Department of pulmonary medicine and infectious diseases Kodal, Anne; Hilleroed Hospital, Department of Anesthesiology Anhøj, Jacob; Danish Society for Patient Safety,
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health services research
Keywords:	AUDIT, Health & safety < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

Title

Experiences with the Global Trigger Tool in Denmark – a pilot study in a non-English speaking country

Authors

Christian von Plessen, MD (corresponding author)
Department of Pulmonary and Infectious Diseases
Hillerød Hospital
Dyrehavevej 29
DK 3400 Hillerød
cple@hih.regionh.dk
Fax,: +45 4829 3935

Anne Marie Kodal, RN
Department of Anesthesiology
Hillerød Hospital
Dyrehavevej 29
DK 3400 Hillerød

anmak@hih.regionh.dk

Jacob Anhøj, MD, DIT
Danish Society for Patient Safety
c/o Hvidovre Hospital
P610 Kettegård Alle 30
2650 Hvidovre

jacob.anhoej@regh.dk

Word count:

Contributions

CvP, JA and AMK designed the study and collected data. CvP drafted the manuscript, JA and AMK revised and approved. JA performed the statistical analyses and produced the graphs.

Copyright

Abstract

Objectives: To describe experiences with the Global Trigger Tool (GTT) in Denmark and suggest ways to improve the performance of GTT review teams

Design: Observational study

Setting: The measurement and monitoring of harms is crucial to campaigns to improve the safety of patients. Increasingly, teams use the GTT to review patient records and measure harms in English and non-English-speaking countries. Meanwhile, it is not clear how the method performs in so diverse settings.

Participants: Review teams from five Danish pilot hospitals of the national Danish patient safety campaign

Primary and secondary outcome measures: We collected harm rates, background and anecdotal information from five pilot hospitals currently participating in the Danish Safer Hospital Programme. An experienced reviewer categorized harms by type. Reported patient safety incidents (PSI) stem from the national Danish Database. We plotted harm rates as run-charts and applied rules for the detection of patterns of non-random variation.

Results: The hospitals differed in size but had similar patient populations and activity. The average harm rate for all hospitals was 60/1000 patient days. The median monthly harm rate ranged from 32 to 91 harms per 1000 patient days. Overall, 96% of harms were temporary. Infections, pressure ulcers procedure related and gastrointestinal problems were common. PSIs varied between 3 and 12/1000 patient days. Teams reported differences in training and review procedures such as the role of the secondary reviewer.

Conclusions: We found substantial variation of harm rates. Differences in training, review procedures and documentation in patient records likely contribute to these variations. Training reviewers as teams, specifying the roles of the different reviewers, training records and a database for findings of reviews may improve the application of the GTT.

Background

Patients run a high risk of being harmed during hospital admissions. Harms occur in up to 10 % of hospitalizations and can cause death, permanent or temporary disability.[1] For patients and health care workers, these harms and the underlying flaws of their health care systems that permit them to happen are deeply upsetting and completely unacceptable.

To improve the safety of patients, national and regional campaigns have been carried through [2-5] or are ongoing.[6] Improvements were achieved in some areas such as reductions of catheter-related blood stream infections.[7] However, system wide progress is slow [8] and improvements are often limited to particular medical conditions or institutions. Indeed, a recent study [9] from the state of North Carolina, an active participant in large scale patient safety initiatives, concluded that overall rates of harm during 2002-2007 were not reduced. Thus, the challenge to improve the safety of patients in hospitals remains and specific and sensitive measures of harms are needed to assess and monitor the effects of changes to make hospitals safer.

In Denmark, the Operation Life campaign during 2006-2008 focused on patient safety in intensive care and during surgery. An estimated 1654 fewer patients died in the Danish population of 5.5 million during the campaign.[10, 11] In 2010, another campaign the Safer Hospital program (www.sikkerpatient.dk/fagfolk/patientsikkert-sygehus.aspx) was launched at five pilot hospitals to reduce mortality by 15 % and harms by 30 % through the implementation of 12 care bundles in the participating hospitals. The hospitals are required to measure and report harms.

Meanwhile, a gold standard for the measurement of harms does not exist. Methods like voluntary reports only detect a small fraction of harms,[12] chart reviews have low interrater reliability [13] and are very time consuming and so are direct observations of health care processes.[14] Studies comparing different methods of harm detection have found very little overlap of the detected harms.[15] Therefore, complete estimates of the incidence of harms probably require the combination of different methods. Meanwhile, such an approach is time consuming and results are often delayed, which is unsuitable for patient safety campaigns in which frequent and regular measurements of harms are needed to evaluate and monitor the effects of interventions and organizational changes.

The Global Trigger Tool (GTT) has been developed for the purpose of monitoring harms at low cost.[16] Harm in this context is defined as an "Unintended physical injury resulting from or contributed to by medical care that requires additional monitoring, treatment or hospitalization, or that results in death."[16] Thus the tool measures factual harm to patients while errors not leading to harm, near errors and errors of omission, are not included. A GTT-review is a trigger based chart audit of closed patient charts. Two reviewers, usually nurses or pharmacists, each review a limited number of randomly chosen charts with a given set of 56 triggers or hints of errors. The finding of such a hint triggers an investigation into whether and, if so, how severely, a patient actually has been harmed. Review time is limited to 20 minutes per admission. Finally, the two reviewers compare their conclusions and eventually a supervisor, usually a physician, judges in cases of disagreement whether the conclusions were appropriate. The number of harms is then expressed as a rate, e.g. harms per 1000 bed days. It has been suggested that GTT teams need a limited amount of training and practice to achieve good levels of reliability to identify harms.[17-19] The feasibility of the method invites for rapid adoption in health care systems around the world where practical ways to measure harms are much in demand. Nevertheless, experiences with the GTT in non-English speaking countries are limited. Thus careful calibration of the instrument and the review team that uses it is warranted to avoid evaluating the safety performance of hospitals with imprecise measurements.

A team of Danish experts translated the GTT to Danish from the English and a Swedish version.[20] The tool was tested in four hospitals in different health regions.[21] The harm rate in these hospitals was around 20/1000 bed days. A recent report of harms to Danish patients with cancer found a rate of 68/1000 bed days.[22] Notwithstanding these variable rates, policy makers advocate the widespread implementation of GTT reviews in Danish hospitals. Meanwhile in our opinion, it is not sufficiently clear, how the tool performs in the hands of Danish review teams.

Therefore, we present in this study our experiences with the GTT in a non-English speaking country and suggest ways to enhance the performance of GTT review teams and contribute to accurate measurements of harms.

Methods

GTT review

A team of two independent reviewers, usually experienced nurses, and one supervisor, usually a physician, reviewed a random sample of 10 admissions twice a month. Closed admissions of patients of at least 18 years of age were and of at least 24 hours duration were eligible. The date of discharge was the index date. The GTT teams reviewed all available information from the admissions, i.e. physician and nurse notes, medication orders and history as well as results of laboratory and other diagnostic tests. Triggers and harms and the severity categorization were recorded on standardized trigger sheets and then transferred to spreadsheets and stored locally. The teams classified the severity of harms according to the National Council for Medication Error Reporting and Prevention Index. There is no shared database for the five hospitals.

Data collection and analysis

The project managers of the Safer Hospital Program and the GTT review teams at the five hospitals supplied background information on the hospitals from hospital administrative systems and the GTT sample populations.

We calculated the monthly harm rate as the total number of harms divided by the total number of patient days multiplied by 1000. The harm rate was then plotted on a run charts and analysed using four run chart rules for the detection of patterns of non-random variation like shifts or trends.[23] We collected and managed the data in Microsoft Excel v. 2003 and produced the statistical analysis and graphs in R Statistical Software v. 2.13.1.

For the purpose of this study, two nurses with experience from over 400 GTT reviews retrospectively categorized harms found at the five hospitals into categories also used by Classen et al.[24] They added gastrointestinal complications and pressure ulcers as categories because these types of harms were common (Figure 4). Each harm was assigned to only one category.

The number and severity of reported patient safety incidents (PSI) stems from the national patient safety incident database (www.dpsd.dk). In Denmark reporting of PSIs is mandatory, confidential and sanction free for health care personnel. Risk managers at

the five hospitals classified the severity of PSIs in the Danish Safety Database into mild, moderate, severe and fatal.

The regional ethical committee deemed an ethical review of the study unnecessary.

Results

Background data of the five pilot hospitals

The five hospitals, one in each Danish health region, have between 142 and 391 beds. All hospitals have departments of internal medicine, orthopaedic and general surgery as well as of obstetrics and gynaecology. All hospitals use electronic patient records, but to a varying degree part of the documentation, such as nursing notes, are on paper. For further information on the hospitals see Table 1. The populations of patients at the five hospitals were similar with regard to age and gender distribution (Table 1). Two hospitals, Næstved and Hillerød, had prior to the Safer Hospitals program been using the GTT for reviewing patient records for one and two years respectively. All teams started reviewing for the safety campaign in May 2010 and reviewed records retrospectively from January 2010.

Table 1: Background information on the five hospitals (2010)

	Hillerød	Horsens	Kolding	Næstved	Thy-Mors
Discharges	60098	30377	27526	28677	11836
Percent females	62	59	61	55	55
Patient days	231978	108060	90710	113353	49711
Outpatient visits	262547	212899	124184	184374	65165
Employees	3163	1367	1507	1668	689
Hospital Standardized Mortality Rate	95	97	96	112	100
Reported safety incidents	2736	365	923	1182	223
Reported patient safety incidents per 1000 patient days	12	3	10	10	4

Experiences with the implementation of the GTT

Reviewing patient files with the GTT was new to three of the five teams. At Hillerød Hospital, the team had reviewed files of deceased patients since 2008.

In May 2010, a GTT team from each of the five hospitals participated in a seven hour training with experts in the method from Denmark and abroad, which included use of the Danish GTT manual, frontal teaching, review of three training records per team and plenary discussions of the findings (Table 2). Only the review team at Hillerød had in 2008 received a similar training. All teams used nurses as primary reviewers and physicians as supervisors. All teams received on-site expert coaching with reviews of 10 or more records, up to three times. Furthermore, all teams participated in two full day network seminars during the study period. All teams started reviewing patient records for the measurement of their baseline in May 2010 and retrospectively reviewed records from January to May 2010.

The compositions of teams changed between zero (Hillerød and Næstved) and three times (Thy) during the study period. Review intervals at the hospitals varied between twice monthly, monthly and irregular. The role of the physician in the review team varied from only judging cases where the primary reviewers were in doubt or disagreed (Hillerød) to identifying all harms based on triggers found by primary reviewers (Horsens). Table 2 shows differences in review procedures at the hospitals.

Table 2: Characteristics and review procedures of GTT teams at five hospitals

	Hillerød	Horsens	Kolding	Næstved	Thy-Mors
Team characteristics					
Number of physicians	2	1	1	2	1
Number of nurses	3	3	4 later 3	4	3
Number of changes in team	0	1 (new physician)	1	0	2 (new nurses)
Review interval	Every fortnight	Monthly	Monthly (two half days)	Monthly	Variable
Training					
Hours of training	14#	7	7	7	7
Site visits by Danish expert (days)	1	3	1	4	2
Complete team present during site visit	-	-	+	-	-
Number of records reviewed together with expert	<10	>10	>10	>10	>10
Review procedures					
Whole team meets for reviews	+	+	Since Jan. 2011	-	-
Physician acts as judge (J) in cases of disagreement or reviewer (R) based on triggers	J	R	J	J	J
Records entirely electronic	-	+	+	-	+
Dedicated person responsible to find records	+	-	+	+	-
Secretary plots triggers and harms	-	-	-	-	+

Team also trained by national expert in 2008

Anecdotal information about GTT reviews

At Næstved, the team sampled 24 records (in case some were incomplete) each month and sorted them in the order of the date of the admission and reviewed the first 20 records. Thus the sample became biased towards admissions in the earlier part of the month. Moreover, the team initially reviewed only admissions to the last department of a hospital admission. Thus they did not find harms that, for example, occurred during an admission to the intensive care unit earlier during the hospital stay. These errors were accidentally discovered during a site visit and the team did a new review for the period. The team at Kolding discovered after three months that their sampling procedure excluded admissions that had an appointment for ambulatory follow up after surgery, and they decided to discard the first three months from their baseline.

GTT findings

In total, 688 adverse events were identified in 11501 patient days, i.e. the overall average harm rate was 60 per 1000 patient days. The median monthly harm rate ranged from 32 to 91 harms per 1000 patient days (Figure 1). One hospital, Hillerød, had a significant upward shift of the harm rate, between September and October 2010, identified by two runs of seven data points below and above the median respectively.[23]

Overall, 96% of harms were temporary (grades E and F). However, the severity distribution varied between hospitals (Figure 2). Notably, the hospital with the highest harm rate (Hillerød) also had the highest proportion of grade E harms. Common types of harm were infections, procedure related, pressure ulcers and gastrointestinal problems. The classification of harms in the category other varied nearly six-fold across hospitals (Figure 4).

Patient safety incidents

The reporting of PSIs among the five hospitals differed between 3 and 12 per thousand patient days. Meanwhile the distributions of PSIs by consequence were almost identical (Figure 3).

Discussion

We were surprised by the wide variation in the detection of harms in the five hospitals. The hospital with the highest harm rate had the highest number of harms of the lowest severity grade E. The hospitals differ in size but the patient populations and activity levels are similar. Also voluntarily reported PSIs were similar in distribution and type. We found differences in the training and the experience of the review teams. Review procedures were standardized for all teams but there were differences in the roles of nurses and physicians in the review process.

Other studies have also found variation of harms across hospitals. Naessens et al.[19] in a study of 1138 admissions to three academic health centres in three states of the US found a variation of harms by hospital between 19,4 % and 37,9 % of admissions. In a study of surgical harms the variation was between 5 % and 35 %,[25] Sharek et al.[26] observed harm rates between 0,18 and 1,28 per patient in 749 admissions to 15 newborn intensive care units in the US and Canada and Resar et al.[27] in 62 intensive care units in the US measured between 3,2 and 27,36 harms per 100 days. Thus significant variations of GTT findings seem to be common.

Several factors could explain the variation in rates of harm. First, they can be caused by real differences in the safety of the clinical processes at the hospitals. However, it seems unlikely that such differences should cause as much as a three-fold variation of harm rates given the similar patient populations at the five hospitals and the homogeneity of the Danish health care system in general. Second, differences in case mix at the hospitals could cause the variations. However, the compositions of the patient populations in the five hospitals are similar. We even found that the hospital with the highest mean age and the highest hospital standardized mortality rate had the lowest rate of harms. Third, the documentation of triggers and harms probably varies across hospitals and the intensified focus on safety could increase the documentation of harms in different ways. Indeed, the hospital with the highest PSI rate also observed a significant increase of harms, which we take as sign of a change in the culture of reporting and documentation at that hospital. Furthermore we observed differences in the way types of harms were coded, especially in the “other” category. Finally, the hospitals use different electronic and paper systems. On the other hand, the training of health care personnel in Denmark is similar in different parts of the country and so are

rules for documentation in the patient records. Concluding on this argument, we assume that the culturally influenced practice of documentation did contribute to the differences in rates of harm across hospitals. Fourth, differences in the training and the experience of the review teams influence the subjective process of judging harms in any record review.[13] The team that found the most harms had attended two training seminars; it also was stable, reviewed regularly with the whole team twice a month and had the longest experience with the GTT. Interestingly, the harm rate at their hospital demonstrated a significant shift in the months after the second training seminar. Fifth, the teams conducted the review processes in slightly different ways. Most importantly, the roles of nurses and physicians varied. The role of the physician in the review team that found the highest harm rate was to judge in cases of disagreement while physicians in the other teams themselves identified harms. We assume that nurses are more prone to register harms of lower severity, while physicians might consider them insignificant. This interpretation is supported by our finding that the variation of harm rates was greatest in the least severe category "E".

Thus in our opinion, the experience of the GTT teams and the way they perform the reviews strongly contributed to the differences in harm rates in the five hospitals. Moreover, differences in the ways harms are documented in the patient records influence the number of harms the GTT team can find. We did not expect these factors to be so important because the GTT reviewing was implemented according to published recommendations [16, 18] and was guided by some of their authors. Moreover, all the teams had attended a GTT network meeting with national and international experts and received site visits by a national expert. However these precautions, it seems, were not sufficient. A recent study from Sweden [28] supports these conclusions. Experienced review teams from five hospitals reviewed 50 patient charts from one of the five hospitals. Harm rates ranged from 27,2 to 99,7 per 1000 patient days and the pair wise interrater reliability of the five teams ranged from a kappa value of 0,26 to 0,77.

Our experiences with the GTT in a non-English speaking country have implications for the implementation of the method in other settings and we suggest the following interventions to improve the implementation of the GTT in new settings:

- Secure that the review team is trained as a team

- Specify of the roles of the reviewers during the reviews to avoid over-/underestimation of especially harms of lesser severity depending on professional background
- Test review teams' abilities to find harms with a set of training charts to estimate their "sensitivity" before routine monitoring is instituted
- Define a minimum number of patient charts that the team should have reviewed before monitoring harms routinely
- Perform reviews with all team members present
- Ensure a structured review process, i.e. a space where the team can work without interruptions, regular time intervals between reviews to keep team "in shape"
- Implement a common database with individual patient data to allow for re-examination of reviewed charts

In conclusion, the GTT is a practical tool to monitor harms that needs careful calibration when using it in new settings. Thus health care staff and policy makers should be aware of the need for sufficient training and retraining of the review teams. Further research should address the training of teams and the evaluation of their performance.

Competing interests

CvP leads the current Safer Hospital campaign at one of the participating hospitals. AMK is a member of the GTT review team at the same hospital. JA is an advisor to the patient safety campaign at the national level. None of us has any financial interests related to this study.

Acknowledgements

GTT teams at the five hospitals: Hillerød: Lisbeth Dammegaard, Anne Marie Kodal, Minna Nielsen, Anne Marie Schlüter, Mette Østergaard. Horsens: Inge-Lise Johansen, Tina Kjær Pedersen, Tom Krintel Petersen (anæstesiologisk afdeling), Kiss Ruben Larsen. Kolding: Helle Guldborg Nielsen, Lynge Kirkegaard, Pernille Kølholt Langkilde, Pernille Lennert, Anita Pedersen, Pia Schrøder Petersen. Næstved: Arne Cyron, Helle Fugl, Kim Garde, Annette Knakkergaard, Nina Pedersen, Trine Zachariassen. Thy Mors: Connie Elbeck van der Koij, Mona Kyndi, Mie Landbo, Bente Ringgaard, Hans-Jörg Selter, Lisbeth Thomsen

Project managers at the five hospitals: Anne Balle Larsen, Søren Brogaard, Marianne Frandsen, Mona Kyndi Pedersen, Kirsten Løth Lysdahl, Søren Schousboe Laursen, Maria Staun, Inge Ulriksen

Funding

CvP received a grant from the Danish Society for Patient Safety to write this article.

Legends to figures

Figure 1: Rates of harm

Run charts showing monthly rates of harms measured with the Global Trigger Tool. The curve shows the harm rate expressed as the number of adverse events per 1000 patient days. The horizontal line is the median harm rate. At Hillerød a shift occurs between September and October 2010 identified by two runs of seven data points on the same side of the median.

Figure 2: Harms by Severity

The dot plots show the relative distribution of severity of harms in categories E – I, where E and F are temporary, G – H permanent harms and I death. Overall, 96% of harms were temporary.

Figure 3: Patient Safety Incidents by Consequence

The dot plots show the relative distribution of patient safety incidents reported to the Danish national database by consequence. *Minor* and *moderate* represent no and temporary harms, *major* permanent harms. Overall 96% of the incidents are temporary.

Figure 4: Harms by Type

The dot plots show the relative distribution of harms by type. VTE = venous thromboembolism.

References

1 de Vries EN, Ramrattan MA, Smorenburg SM et al. The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care* 2008 Jun;**17**:216-23.

2 Berwick DM, Calkins DR, McCannon CJ et al. The 100,000 lives campaign: setting a goal and a deadline for improving health care quality. *JAMA* 2006 Jan 18;**295**:324-7.

3 Benning A, Dixon-Woods M, Nwulu U et al. Multiple component patient safety intervention in English hospitals: controlled evaluation of second phase. *BMJ* 2011;**342**:d199.

4 Benning A, Ghaleb M, Suokas A et al. Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation. *BMJ* 2011;**342**:d195.

5 McCannon CJ, Schall MW, Calkins DR et al. Saving 100,000 lives in US hospitals. *BMJ* 2006;**332**:1328-30.

6 NHS Scotland. Scottish Patient Safety Programme. Available at: http://www.patientsafetyalliance.scot.nhs.uk/programme. Accessed Nov. 3rd, 2011.

7 Pronovost P, Needham D, Berenholtz S et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 2006;**355**:2725-32.

8 Vincent C, Aylin P, Franklin BD et al. Is health care getting safer? *BMJ* 2008;**337**:a2426.

9 Landrigan CP, Parry GJ, Bones CB et al. Temporal trends in rates of patient harm resulting from medical care. *N Engl J Med* 2010;**363**:2124-34.

10 Operation Life Danish Society for Patient Safety. Operation Life Slutrapport. 2009.

11 Dansk Selskab for Patientsikkerhed. The Danish Safer Hospital Programme. 2011; Available at: http://www.sikkerpatient.dk/about-sikker-patient/the-danish-safer-hospital-programme.aspx. Accessed Nov. 3rd, 2011.

12 Sari AB, Sheldon TA, Cracknell A et al. Sensitivity of routine system for reporting patient safety incidents in an NHS hospital: retrospective patient case note review. *BMJ* 2007;**334**(7584):79.

13 Thomas EJ, Lipsitz SR, Studdert DM et al. The reliability of medical record review for estimating adverse event rates. *Ann Intern Med* 2002;**136**:812-16.

14 Buckley MS, Erstad BL, Kopp BJ et al. Direct observation approach for detecting medication errors and adverse drug events in a pediatric intensive care unit. *Pediatr Crit Care Med* 2007;**8**:145-52.

15 Naessens JM, Campbell CR, Huddleston JM et al. A comparison of hospital adverse events identified by three widely used detection methods. *Int J Qual Health Care* 2009;**21**:301-7.

16 Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events. IHI Innovation Series White paper. 2009.

17 Classen DC., Lloyd RC., Provost L et al. Development and Evaluation of the Institute for Healthcare Improvement Global Trigger Tool. *J Patient Saf* 2008;**4**:169-77.

18 Adler L, Denham CR, McKeever M et al. Global Trigger Tool: Implementation Basics. *J Patient Saf* 2008;**4**:245-49.

19 Naessens JM, O'Byrne TJ, Johnson MG et al. Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool. *Int J Qual Health Care* 2010;**22**:266-74.

20 Strukturerad journalgranskning för att identifiera och mäta förekomst av skador i vården enligt metoden Global Trigger Tool, Institute for Healthcare Improvement Innovation series 2007, Svensk översättning och anpassning 2008.

21 Rabøl LI, Jølving LR, Bjørn B et al. . Systematisk journalgennemgang med IHI's Global Trigger Tool - Erfaringer fra afprøvning på danske sygehuse. 2008.

22 Patientsikkerhed i kræftbehandlingen - journalgennemgang med Global Trigger Tool. Rapport nr. 3/2010. 2010.

23 Perla RJ, Provost LP, Murray SK. The run chart: a simple analytical tool for learning from variation in healthcare processes. *BMJ Qual Saf* 2011;**20**:46-51.

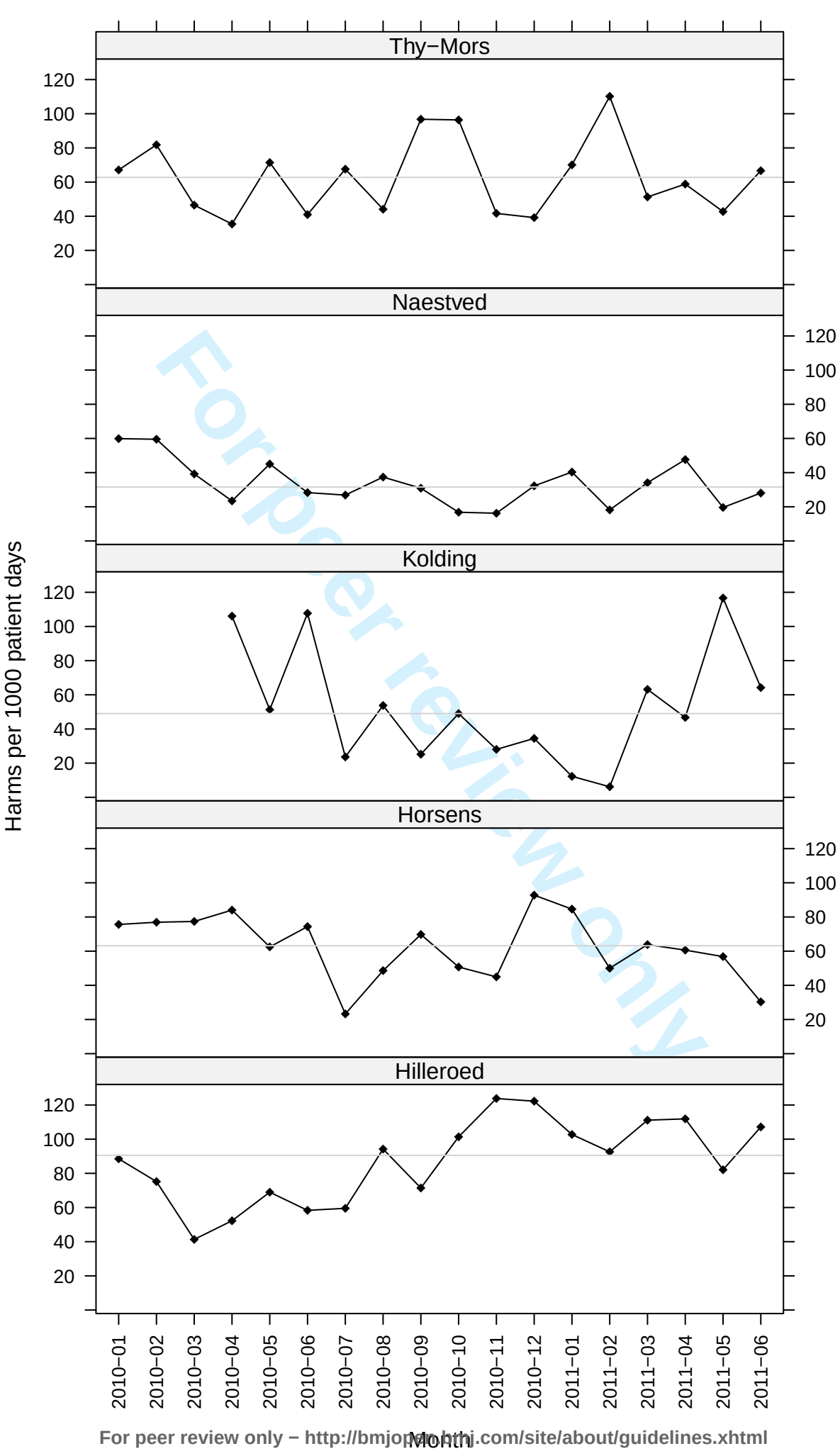
24 Classen DC, Resar R, Griffin F et al. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. *Health Aff (Millwood)* 2011;**30**:581-9.

25 Griffin FA, Classen DC. Detection of adverse events in surgical patients using the Trigger Tool approach. *Qual Saf Health Care* 2008;**17**:253-8.

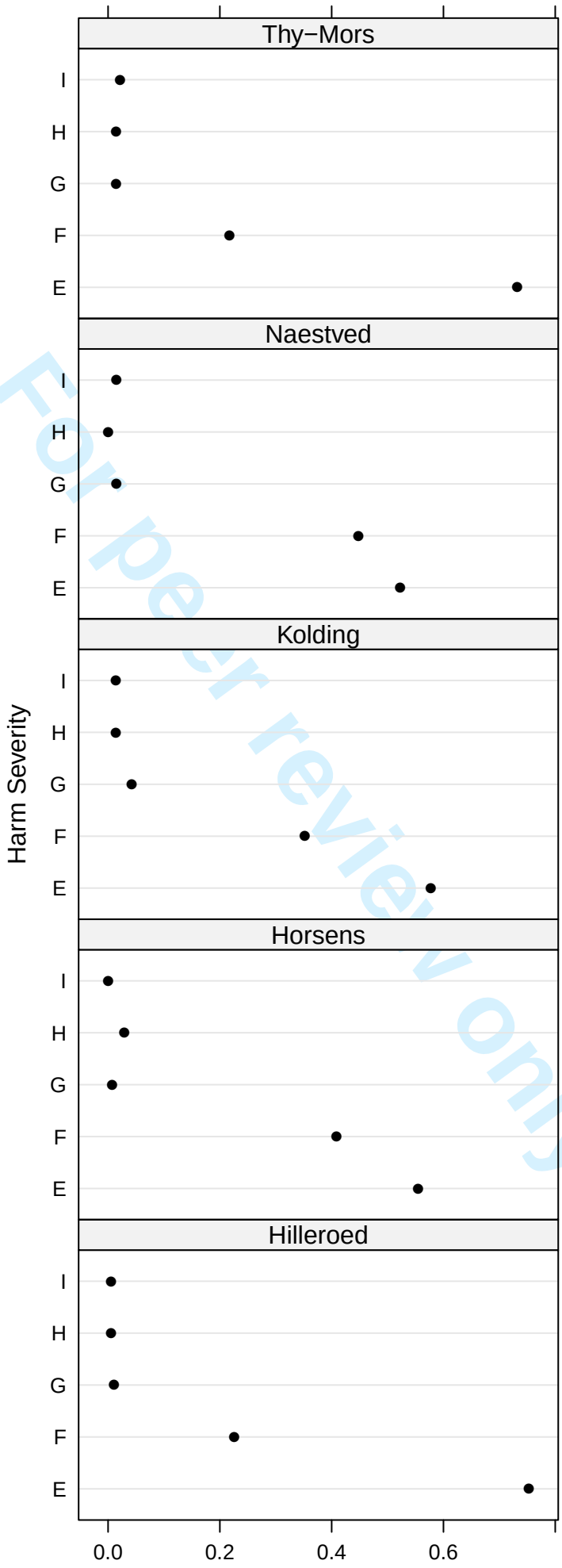
26 Sharek PJ, Horbar JD, Mason W et al. Adverse events in the neonatal intensive care unit: development, testing, and findings of an NICU-focused trigger tool to identify harm in North American NICUs. *Pediatrics* 2006;**118**:1332-40.

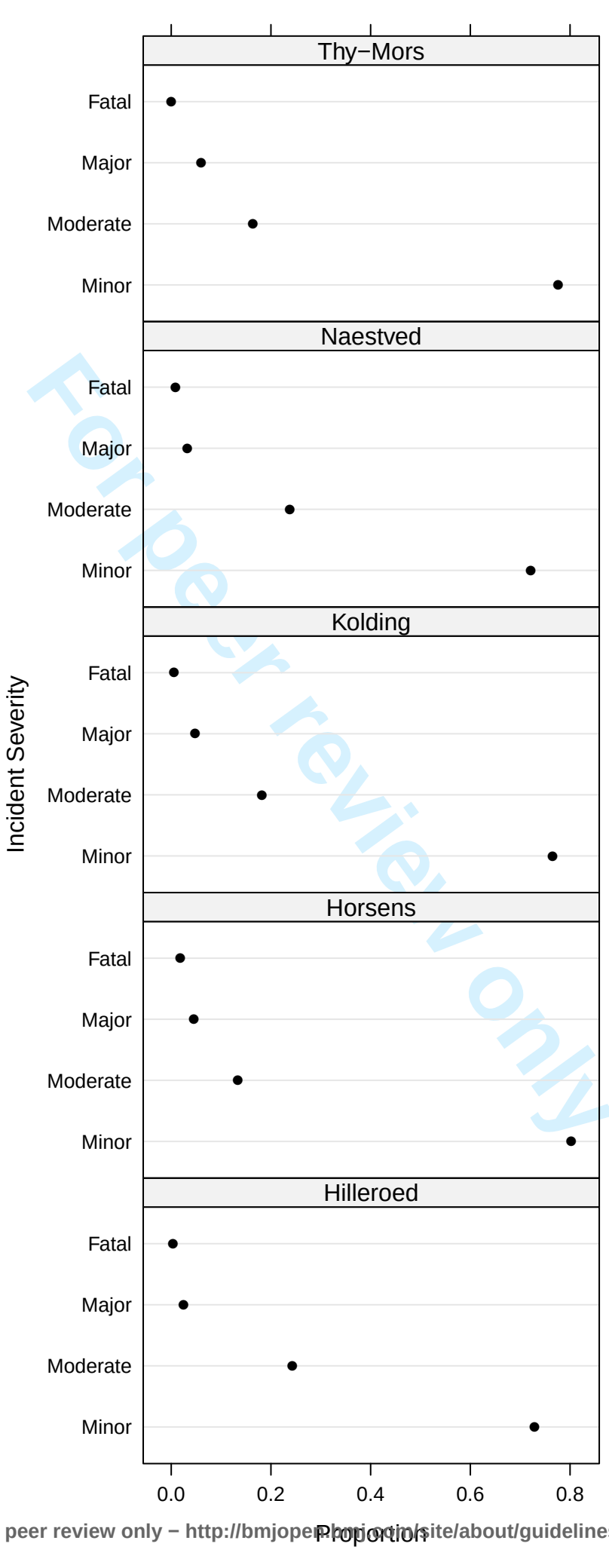
27 Resar RK, Rozich JD, Simmonds T et al. A trigger tool to identify adverse events in the intensive care unit. *Jt Comm J Qual Patient Saf* 2006;**32**:585-90.

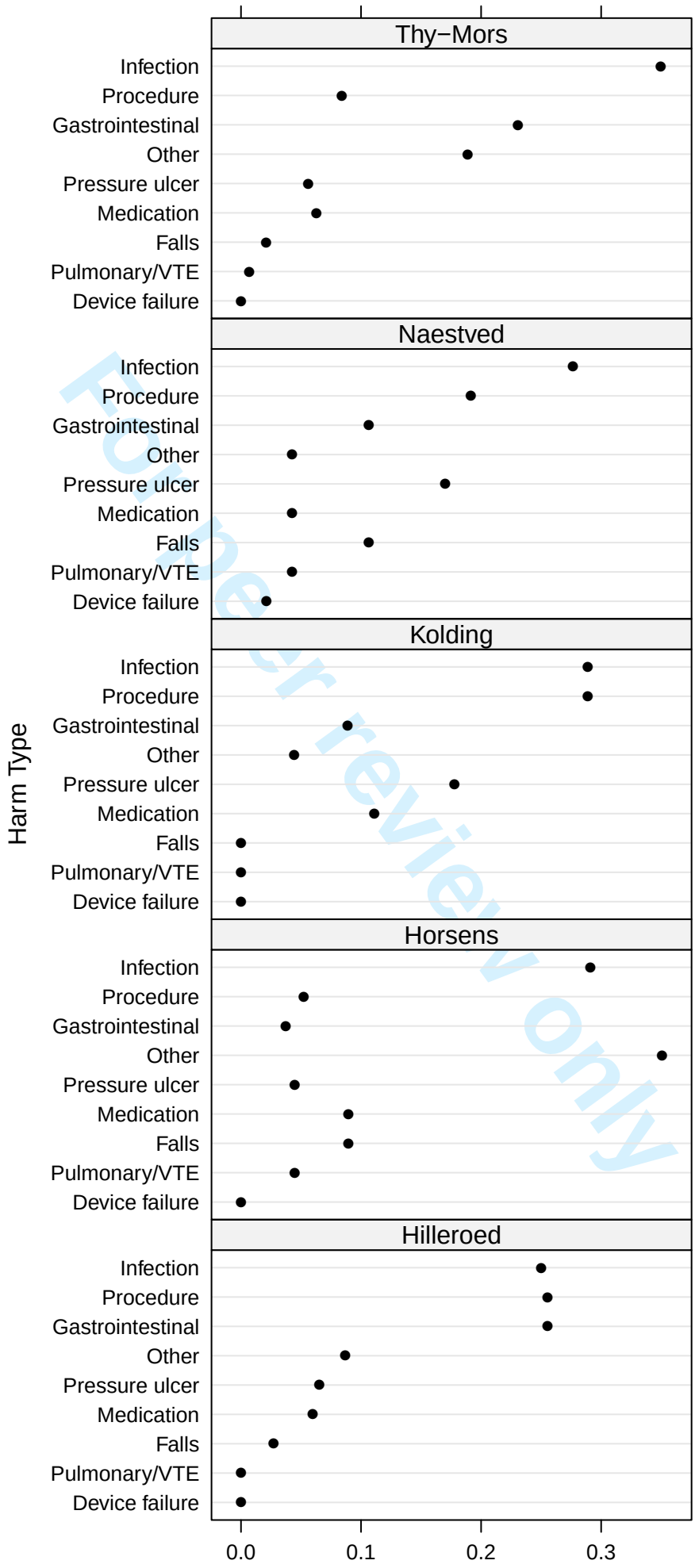
28 Schildmeijer K, Nilsson L, Årestedt K et al. Assessment of adverse events in medical care: lack of consistency between experienced teams using the global trigger tool. *Qual Saf Health Care* 2012;**21**:307-14



Harms by Severity







Title

Experiences with Global Trigger Tool reviews in five Danish hospitals – an implementation study

Authors

Christian von Plessen, MD (corresponding author)
Department of Pulmonary and Infectious Diseases
Hillerød Hospital
Dyrehavevej 29
DK 3400 Hillerød
cple@hih.regionh.dk
Fax: +45 4829 3935

Anne Marie Kodal, RN
Department of Anaesthesiology
Hillerød Hospital
Dyrehavevej 29
DK 3400 Hillerød

anmak@hih.regionh.dk

Jacob Anhøj, MD, DIT
Danish Society for Patient Safety
c/o Hvidovre Hospital
P610 Kettegård Alle 30
2650 Hvidovre

jacob.anhoej@regh.dk

Contributions

CvP, JA and AMK designed the study and collected data. CvP drafted the manuscript, JA and AMK revised and approved. JA performed the statistical analyses and produced the graphs.

Copyright

Abstract

Objectives: To describe experiences with the implementation of Global Trigger Tool (GTT) reviews in five Danish hospitals and to suggest ways to improve the performance of GTT review teams

Design: Retrospective observational study

Setting: The measurement and monitoring of harms is crucial to campaigns to improve the safety of patients. Increasingly, teams use the GTT to review patient records and measure harms in English and non-English-speaking countries. Meanwhile, it is not clear how the method performs in so diverse settings.

Participants: Review teams from five Danish pilot hospitals of the national Danish Safer Hospital Program

Primary and secondary outcome measures: We collected harm rates, background and anecdotal information and reported patient safety incidents (PSIs) from five pilot hospitals currently participating in the Danish Safer Hospital Programme. Experienced reviewers categorized harms by type. We plotted harm rates as run-charts and applied rules for the detection of patterns of non-random variation.

Results: The hospitals differed in size but had similar patient populations and activity. PSIs varied between 3 and 12/1000 patient days. The average harm rate for all hospitals was 60/1000 patient days. The median monthly harm rate ranged from 32 to 91 harms per 1000 patient days. Overall, 96% of harms were temporary. Infections, pressure ulcers procedure related and gastrointestinal problems were common. Teams reported differences in training and review procedures such as the role of the secondary reviewer.

Conclusions: We found substantial variation of harm rates. Differences in training, review procedures and documentation in patient records likely contributed to these variations. Training reviewers as teams, specifying the roles of the different reviewers, training records and a database for findings of reviews may improve the application of the GTT.

Background

Patients run a high risk of being harmed during hospital admissions. Harms occur in up to 10 % of hospitalizations and can cause death, permanent or temporary disability.[1] For patients and health care workers, these harms and the underlying flaws of their health care systems that permit them to happen are deeply upsetting and completely unacceptable.

To improve the safety of patients, national and regional campaigns have been carried through [2-5] or are ongoing.[6] Improvements were achieved in some areas such as reductions of catheter-related blood stream infections.[7] However, system wide progress is slow [8] and improvements are often limited to particular medical conditions or institutions. Indeed, a recent study [9] from the state of North Carolina, an active participant in large scale patient safety initiatives, concluded that overall rates of harm during 2002-2007 were not reduced. Thus, the challenge to improve the safety of patients in hospitals remains and specific and sensitive measures of harms are needed to assess and monitor the effects of changes to make hospitals safer.

In Denmark, the Operation Life campaign during 2006-2008 focused on patient safety in intensive care and during surgery. An estimated 1654 fewer patients died in the Danish population of 5.5 million during the campaign.[10] In 2010, another campaign the Safer Hospital Program (www.sikkerpatient.dk/fagfolk/patientsikkert-sygehus.aspx) was launched at five pilot hospitals to reduce mortality by 15 % and harms by 30 % through the implementation of 12 care bundles. The hospitals are required to measure and report harms.

Meanwhile, a gold standard for the measurement of harms does not exist. Methods like voluntary reports only detect a small fraction of harms,[11] chart reviews have low interrater reliability [12] and are very time consuming and so are direct observations of health care processes.[13] Studies comparing different methods of harm detection have found very little overlap of the detected harms.[14] Therefore, complete estimates of the incidence of harms probably require the combination of different methods. Meanwhile, such an approach is time consuming and results are often delayed, which is unsuitable for patient safety campaigns in which frequent and regular measurements of

harms are needed to evaluate and monitor the effects of interventions and organizational changes.

The Global Trigger Tool (GTT) has been developed for the purpose of monitoring harms at low cost.[15] Harm in this context is defined as an "Unintended physical injury resulting from or contributed to by medical care that requires additional monitoring, treatment or hospitalization, or that results in death."[16] Thus the tool measures factual harm to patients while errors not leading to harm, near errors and errors of omission, are not included. A GTT-review is a trigger based chart audit of closed patient charts. Two reviewers, usually nurses or pharmacists, each review a limited number of randomly chosen charts with a given set of 56 triggers or hints of errors. The finding of such a hint triggers an investigation into whether and, if so, how severely, a patient actually has been harmed. Review time is limited to 20 minutes per admission. Finally, the two reviewers compare their conclusions and a supervisor, usually a physician, qualifies the number and severity of harms and decides in cases of disagreement. The number of harms is then expressed as a rate, e.g. harms per 1000 bed days. It has been suggested that GTT teams need a limited amount of training and practice to achieve good levels of reliability to identify harms.[16-18] The feasibility of the method invites for rapid adoption in health care systems around the world where practical ways to measure harms are much in demand. Nevertheless, experiences with the GTT in non-English speaking countries are limited. Thus careful calibration of the instrument and the review team that uses it is warranted to avoid evaluating the safety performance of hospitals with imprecise measurements.

A team of Danish experts translated the GTT to Danish [19] from the English and a Swedish version.[20] The tool was tested in four hospitals in different health regions.[21] The harm rate in these hospitals was around 20/1000 bed days. A recent report of harms to Danish patients with cancer found a rate of 68/1000 bed days.[22] Notwithstanding these variable rates, policy makers advocate the widespread [23] implementation of GTT reviews in Danish hospitals. Meanwhile in our opinion, it is not sufficiently clear, how the tool performs in the hands of Danish review teams.

The aim of this study was to describe experiences with the GTT in five Danish hospitals and suggest ways to improve the performance of GTT review teams and thus contribute to the accurate measurements of harms.

Methods

In this retrospective observational study, we present harm rates as measured by the GTT at the five hospitals participating in the Safer Hospital Program in Denmark. The hospitals had used the GTT for 18 months from January 2010 until June 2011. The harm rates are registered at common database. The project managers at the hospitals supplied tabulated data on the size, activity and patient populations of the five hospitals from the year 2010. All data were collected between August and December 2011.

CvP, the project manager at one of the five hospitals and a consultant in pulmonary medicine and AMK, a registered nurse and experienced GTT reviewer interviewed members of the GTT teams and project managers on the telephone. The interviews comprised questions on the training of the GTT teams, team composition, roles of the team members and review processes. Moreover, they asked open questions about unexpected observations and changes. The GTT teams supplied lists of the recorded harms from their reviews.

To give an impression of the safety culture at the participating hospitals, we present the number and severity of reported patient safety incidents (PSIs) in 2010. JA, a physician by training who works in the Danish Society for Patient Safety, gathered these data from the Danish Patient Safety Database (www.dpsd.dk) for voluntary reporting of PSIs. Risk managers at the five hospitals classify the severity of PSIs in the Danish Safety Database into mild, moderate, severe and fatal. In Denmark, reporting of PSIs is mandatory, confidential and sanction free for health care personnel.

GTT review

The Danish translation of the GTT toolkit was the reference for the review teams.[19] Teams were to review a random sample of 10 admissions twice a month. Closed admissions of patients of at least 18 years of age and of at least 24 hours duration were eligible. The date of discharge was the index date. The GTT teams should review all available information from the admissions, i.e. physician and nurse notes, medication orders and history as well as results of diagnostic tests. Each primary reviewer should

review each record. Afterwards the primary reviewers should meet to discuss harms and come to consensus. Finally the primary reviewers should meet with the supervising physician to present their findings for approval and severity classification. Triggers, harms and the severity of harms should be recorded on standardised work sheets, then transferred to spreadsheets and stored locally. The team should classify the severity of harms according to the National Council for Medication Error Reporting and Prevention Index. The five hospitals do not have a shared database for their GTT findings.

Data analysis

We calculated the monthly harm rate as the total number of harms divided by the total number of patient days multiplied by 1000. The harm rate was then plotted on a run chart and analysed using four run chart rules for the detection of patterns of non-random variation such as shifts or trends.[24] We collected and managed the data in Microsoft Excel v. 2003 and produced the statistical analysis and graphs in R Statistical Software v. 2.13.1.

For the purpose of this study, two nurses from the GTT team of one of the hospitals who had done over 400 GTT reviews each retrospectively categorized harms found at all five hospitals into categories also used by Classen et al.[25] They added gastrointestinal complications and pressure ulcers as categories because these were common types of harm (Figure 4). Each harm was assigned to only one category.

The regional ethical committee deemed an ethical review of the study unnecessary.

Results

Background data of the five pilot hospitals

The five hospitals, one in each Danish health region vary in size but all hospitals have departments of internal medicine, orthopaedic and general surgery as well as of obstetrics and gynaecology. All hospitals use electronic patient records, but to varying degree parts of the documentation, such as nursing notes, are on paper. The populations of patients at the five hospitals were similar with regard to age and gender distribution. There was a four-fold difference in reporting of PSIs among hospitals (Table 1). Meanwhile, the distributions of PSIs by consequence were almost identical (Figure 3). All teams started reviewing for the safety program in May 2010 and reviewed records retrospectively from January 2010.

Table 1: Background information on the five hospitals (2010)

	Hillerød	Horsens	Kolding	Næstved	Thy-Mors
Discharges	60098	30377	27526	28677	11836
Average patient age (years)	55	57	53	59	58
Percent females	62	59	61	55	55
Patient days	231978	108060	90710	113353	49711
Outpatient visits	262547	212899	124184	184374	65165
Employees	3163	1367	1507	1668	689
Hospital Standardized Mortality Rate	95	97	96	112	100

Reported safety incidents	2736	365	923	1182	223
Reported patient safety incidents per 1000 patient days	12	3	10	10	4

Experiences with the implementation of the GTT

Reviewing patient files with the GTT was new to three of the five teams. Prior to the Safer Hospitals programme, Næstved and Hillerød Hospital had been using the GTT for reviewing patient records for one and two years respectively.

In May 2010, the GTT team from each of the five hospitals participated in a seven hour training session with experts in the method from Denmark and the Institute of Health Care Improvement (IHI). The session included an introduction to the Danish GTT manual, frontal teaching, review of three training records per team and plenary discussions of the findings (Table 2). Only the review team at Hillerød had in 2008 received a similar training. All teams used nurses as primary reviewers and physicians as supervisors. All teams received on-site expert coaching with reviews of 10 or more records, up to three times. The expert coach was a physician who was trained by experts from the IHI. Furthermore, all teams participated in two full day network seminars during the study period. All teams started reviewing patient records for the measurement of their baseline in May 2010 and retrospectively reviewed records from January to May 2010.

The compositions of teams changed between zero (Hillerød and Næstved) and three times (Thy) during the study period. Review intervals at the hospitals varied between two weeks, one monthly and irregular. Complete teams reviewed together and compared findings at two, later three hospitals. The role of the physician in the review team varied from only judging cases where the primary reviewers were in doubt or

disagreed (Hillerød) to identifying harms based on triggers found by primary reviewers (Horsens). Table 2 shows differences in review procedures at the hospitals.

Table 2: Characteristics and review procedures of GTT teams at five hospitals

	Hillerød	Horsens	Kolding	Næstved	Thy-Mors
Team characteristics					
Number of physicians	2	1	1	2	1
Number of nurses	3	3	4 later 3	4	3
Number of changes in team	0	1 (physician)	1	0	2 (nurses)
Review intervals	Twice per month	Monthly	Monthly (two half days)	Monthly	Variable
Training					
Hours of training	14[#]	7	7	7	7
Site visits by Danish expert (days)	1	3	1	4	2
Complete team present during site visit	-	-	+	-	-f
Number of records reviewed together with expert	<10	>10	>10	>10	>10
Review procedures					
Whole team meets for reviews	+	+	Since Jan. 2011	-	-
Physician acts as judge (J) in cases of disagreement or reviewer (R) based on triggers	J	R	J	J	J

Records entirely electronic	-	+	+	-	+
Dedicated person responsible to find records	+	-	+	+	-
Secretary plots triggers and harms	-	-	-	-	+

Team also trained by national expert in 2008

Anecdotal information about GTT reviews

At Næstved, the team sampled 24 records (in case some were incomplete) each month and sorted them in the order of the date of the admission and reviewed the first 20 records. Thus the sample became biased towards admissions in the earlier part of the month. Moreover, the team initially reviewed only admissions to the last department of a hospital admission. Thus they did not find harms that, for example, occurred during an admission to the intensive care unit earlier during the hospital stay. These errors were accidentally discovered during a site visit and the team did a new review for the period. The team at Kolding discovered after three months that their sampling procedure excluded admissions that had an appointment for ambulatory follow up after surgery, and they decided to discard the first three months from their baseline.

GTT findings

In total, 688 adverse events were identified in 11501 patient days, i.e. the overall average harm rate was 60 per 1000 patient days. The median monthly harm rate ranged from 32 to 91 harms per 1000 patient days (Figure 1). The number of harms per 100 admissions were Hillerød 55,0, Horsens 41,9, Kolding 30,0, Næstved 23,9 and Thy Mors 45,3. One hospital, Hillerød, had a significant upward shift of the harm rate, between September and October 2010, identified by two runs of seven data points below and above the median respectively.

Overall, 96% of harms were temporary (grades E and F). However, the severity distribution varied between hospitals (Figure 2). Notably, the hospital with the highest harm rate (Hillerød) also had the highest proportion of grade E harms. Common types of harm were infections, procedure related, pressure ulcers and gastrointestinal problems. The classification of harms in the category other varied nearly six-fold across hospitals (Figure 4).

Discussion

We observed marked differences in the harm rates and types identified by GTT review teams in five Danish public hospitals. The GTT is not designed to compare hospitals but we were surprised by the magnitude of the variations. Therefore we designed this study to identify factors that contributed to the differences. The hospitals, their patient populations, structures and activity levels were similar but we found differences in the training, the review procedures and the experience of the review teams.

Other studies have also found variation of harms across hospitals. Naessens et al.[18] in a study of 1138 admissions to three academic health centres in three states of the US found a variation of harms by hospital between 23,1 % and 37,9 % of admissions. In a study of surgical harms the variation was between 5 % and 35 %,[26] Sharek et al.[27] observed harm rates between 0,18 and 1,28 per patient in 749 admissions to 15 newborn intensive care units in the US and Canada and Resar et al.[28] in 62 intensive care units in the US measured between 3,2 and 27,36 harms per 100 days. Thus significant variations of GTT findings seem to be common.

Several factors could explain the variation in rates of harm. First, they can be caused by real differences in the safety of the clinical processes at the hospitals. However, it seems unlikely that such differences should cause as much as a three-fold variation of harm rates given the similar patient populations at the five hospitals and the homogeneity of the Danish health care system in general. Second, differences in case mix at the hospitals could cause the variations. However, the compositions of the patient populations in the five hospitals were similar. We even found that the hospital with the highest mean age and the highest hospital standardized mortality rate had the lowest rate of harms. Third, the documentation of triggers and harms probably varies across hospitals. Interestingly, the hospital with the highest PSI rate also observed a significant increase of harm rates. High PSI rates can be a sign of a mature safety culture rather than of poor safety and one could assume that staff documents more harms in a hospital with such a culture. Also different types of patient records (electronic and paper) and differences in layout and presentation could influence the results of the reviews. We do not have data to explore these questions but we certainly cannot exclude an influence on the different harm rates across hospitals. Fourth, differences in the training and the experience of the review teams influence the subjective process of judging harms in any record review.[12] The team that found the

most harms had also used the GTT longer and attended two training seminars; it also was stable, reviewed regularly with the whole team twice a month. Interestingly, the harm rate at their hospital demonstrated a significant shift in the months after the second training seminar. Finally, the teams conducted the review processes in slightly different ways. Most importantly, the roles of nurses and physicians varied. The role of the physician in the review team that found the highest harm rate was to judge in cases of disagreement while physicians in the other teams themselves identified harms. We assume that nurses are more prone to register harms of lower severity, while physicians might consider them insignificant. This interpretation is supported by our finding that the variation of harm rates was greatest in the least severe category "E".

Thus in our opinion, the experience of the GTT teams and the way they perform the reviews strongly contributed to the differences in harm rates in the five hospitals. Notably, four of the five teams found rates between Moreover, differences in the way harms are documented in the patient records seem to influence the number of harms the GTT team can find. We did not expect these factors to be so important because the GTT was implemented according to current recommendations [15, 17] and was guided by some of their authors. Moreover, all the teams had attended a GTT network meeting with national and international experts and received site visits by a national expert. However these precautions, it seems, were not sufficient to secure at standardized reviewing process at the hospitals. Thus users of the GTT and its results, health care personnel, administrators, payers or the public, should be aware of the challenges of the implementation of the method and allow for sufficient training and evaluation of the results. Our findings also stress that GTT findings should guide hospitals in their efforts to improve patient safety but the results should not be used to compare hospitals.

Strengths, limitations and further research

The strengths of this study are its relevance for the implementation of the GTT that increasingly is being used to monitor the safety in hospitals. Our contextual data is detailed and thus practical. The limitations are inherent to the observational nature of the study that prevents conclusions on causal links between the variations of harm rates and the observed differences in team training and review processes. For the same reason, we cannot quantify the contribution of the different factors. Nevertheless, the

findings are plausible and fit with the recommendations for the use of the method [REF WHITE PAPER].

Further research should address how teams' reviewing experience and training influence team performance and how team training can be optimised. Moreover, studies should investigate the influence of changes of documentation and presentation of information in patient charts and the use of the types of harm for improving patient safety.

In conclusion, differences in training, review processes, and documentation contributed to variations in rates of harm as measured by the GTT. Thus health care staff and policy makers should be aware of the need for systematic training of the review teams and standardisation of the review process when implementing the GTT in new settings. These factors are related to the implementation of the GTT reviews and are not inherent to the method as such. Our findings have implications for the implementation of the method in other settings and we suggest considering the following interventions to improve the implementation of the GTT in new settings:

- Secure that the review team is trained as a team
- Specify of the roles of the reviewers during the reviews to avoid over-/underestimation of especially harms of lesser severity depending on professional background
- Test review teams' abilities to find harms with a set of training charts to estimate their "sensitivity" before routine monitoring is instituted
- Define a minimum number of patient charts that the team should have reviewed before monitoring harms routinely
- Perform reviews with all team members present
- Ensure a structured review process, i.e. a space where the team can work without interruptions, regular time intervals between reviews to keep team "in shape"
- Implement a common database with individual patient data to allow for re-examination of reviewed charts to avoid problems such as sampling errors

Competing interests

CvP leads the current Safer Hospital campaign at one of the participating hospitals. AMK is a member of the GTT review team at the same hospital. JA is an advisor to the patient safety campaign at the national level. None of us has any financial interests related to this study.

Acknowledgements

GTT teams at the five hospitals: Hillerød: Lisbeth Dammegaard, Anne Marie Kodal, Minna Nielsen, Anne Marie Schlüter, Mette Østergaard. Horsens: Inge-Lise Johansen, Tina Kjær Pedersen, Tom Krintel Petersen (anæstesiologisk afdeling), Kiss Ruben Larsen. Kolding: Helle Guldborg Nielsen, Lynge Kirkegaard, Pernille Kølholt Langkilde, Pernille Lennert, Anita Pedersen, Pia Schrøder Petersen. Næstved: Arne Cyron, Helle Fugl, Kim Garde, Annette Knakkergaard, Nina Pedersen, Trine Zachariassen. Thy Mors: Connie Elbeck van der Koij, Mona Kyndi, Mie Landbo, Bente Ringgaard, Hans-Jörg Selter, Lisbeth Thomsen

Project managers at the five hospitals: Anne Balle Larsen, Søren Brogaard, Marianne Frandsen, Mona Kyndi Pedersen, Kirsten Løth Lysdahl, Søren Schousboe Laursen, Maria Staun, Inge Ulriksen

Funding

CvP received a grant from the Danish Society for Patient Safety to write this article.

Legends to figures

Figure 1: Rates of harm

The run charts show monthly rates of harms measured with the Global Trigger Tool. The curve shows the harm rate expressed as the number of adverse events per 1000 patient days. The horizontal line is the median harm rate. At Hillerød a shift occurs between September and October 2010 identified by two runs of seven data points on the same side of the median.

Figure 2: Harms by Severity

The dot plots show the relative distribution of severity of harms in categories E – I, where E and F are temporary, G – H permanent harms and I death. Overall, 96% of harms were temporary.

Figure 3: Patient Safety Incidents by Consequence

The dot plots show the relative distribution of patient safety incidents reported to the Danish national database by consequence. *Minor* and *moderate* represent no and temporary harms, *major* permanent harms. Overall 96% of the incidents are temporary.

Figure 4: Harms by Type

The dot plots show the relative distribution of harms by type. VTE = venous thromboembolism.

References

1 de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. Qual Saf Health Care 2008 Jun;**17**(3):216-223.

2 Berwick DM, Calkins DR, McCannon CJ, Hackbarth AD. The 100,000 lives campaign: setting a goal and a deadline for improving health care quality. JAMA 2006 Jan 18;**295**(3):324-327.

3 Benning A, Dixon-Woods M, Nwulu U, Ghaleb M, Dawson J, Barber N, et al. Multiple component patient safety intervention in English hospitals: controlled evaluation of second phase. BMJ 2011 Feb 3;**342**:d199.

4 Benning A, Ghaleb M, Suokas A, Dixon-Woods M, Dawson J, Barber N, et al. Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation. BMJ 2011 Feb 3;**342**:d195.

5 McCannon CJ, Schall MW, Calkins DR, Nazem AG. Saving 100,000 lives in US hospitals. BMJ 2006 Jun 3;**332**(7553):1328-1330.

6 NHS Scotland. Scottish Patient Safety Programme. Available at: http://www.patientsafetyalliance.scot.nhs.uk/programme. Accessed 11/03, 2011.

7 Pronovost P, Needham D, Berenholtz S, Sinopoli D, Chu H, Cosgrove S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. N Engl J Med 2006 Dec 28;**355**(26):2725-2732.

8 Vincent C, Aylin P, Franklin BD, Holmes A, Iskander S, Jacklin A, et al. Is health care getting safer? BMJ 2008 Nov 13**;337**:a2426.

9 Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal trends in rates of patient harm resulting from medical care. N Engl J Med 2010 Nov 25**;363**(22):2124-2134.

10 Dansk Selskab for Patientsikkerhed. The Danish Safer Hospital Programme. 2011; Available at: http://www.sikkerpatient.dk/about-sikker-patient/the-danish-safer-hospital-programme.aspx. Accessed 11/03, 2011.

11 Sari AB, Sheldon TA, Cracknell A, Turnbull A. Sensitivity of routine system for reporting patient safety incidents in an NHS hospital: retrospective patient case note review. BMJ 2007 Jan 13**;334**(7584):79.

12 Thomas EJ, Lipsitz SR, Studdert DM, Brennan TA. The reliability of medical record review for estimating adverse event rates. Ann Intern Med 2002 Jun 4**;136**(11):812-816.

13 Buckley MS, Erstad BL, Kopp BJ, Theodorou AA, Priestley G. Direct observation approach for detecting medication errors and adverse drug events in a pediatric intensive care unit. Pediatr Crit Care Med 2007 Mar**;8**(2):145-152.

14 Naessens JM, Campbell CR, Huddleston JM, Berg BP, Lefante JJ, Williams AR, et al. A comparison of hospital adverse events identified by three widely used detection methods. Int J Qual Health Care 2009 Aug**;21**(4):301-307.

15 Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events. IHI Innovation Series White paper. 2009.

16 Classen DC., Lloyd RC., Provost L., Griffin FA., Resar R. Development and Evaluation of the Institute for Healthcare Improvement Global Trigger Tool. Journal of Patient Safety 2008**;4**(3):169-177.

17 Adler L, Denham CR, McKeever M, Purinton R, Guilloteau F, Moorhead JD, et al. Global Trigger Tool: Implementation Basics. J Patient Saf 2008**;4**:245-249.

18 Naessens JM, O'Byrne TJ, Johnson MG, Vansuch MB, McGlone CM, Huddleston JM. Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool. Int J Qual Health Care 2010 Aug**;22**(4):266-274.

19 Anonymous. Global Trigger Tool. En oversættelse af IHI's værktøj tilpasset danske forhold. Center for Kvalitet. Region Syddanmark 2008.

20 Anonymous. Strukturerat journalgranskning för att identifiera och mätta förekomst av skador i vården enligt metoden Global Trigger Tool. Institute for Health Care Improvement Innovation series 2007, Svensk översätning och anpassning. Stockholm. Kommentus Förlag 2008.

21 Rabøl LI, Jølving LR, Bjørn B and Viskum B. Systematisk journalgennemgang med IHI's Global Trigger Tool - Erfaringer fra afprøvning i danske sygehuse. Copenhagen. Dansk Selskap for Patientsikkerhed 2008.

22 Anonymous. Patientsikkerhed i kræftbehandlingen - journalgennemgang med Global Trigger Tool. Rapport nr 3/2010. Copenhagen. Kræftens Bekæmpelse 2010.

23 Auerbach AD, Landefeld CS, Shojania KG. The tension between needing to improve care and knowing how to do it. N Engl J Med 2007 Aug 9**;357**(6):608-613.

24 Perla RJ, Provost LP, Murray SK. The run chart: a simple analytical tool for learning from variation in healthcare processes. BMJ Qual Saf 2011 Jan;**20**(1):46-51.

25 Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. Health Aff (Millwood) 2011 Apr;**30**(4):581-589.

26 Griffin FA, Classen DC. Detection of adverse events in surgical patients using the Trigger Tool approach. Qual Saf Health Care 2008 Aug;**17**(4):253-258.

27 Sharek PJ, Horbar JD, Mason W, Bisarya H, Thurm CW, Suresh G, et al. Adverse events in the neonatal intensive care unit: development, testing, and findings of an NICU-focused trigger tool to identify harm in North American NICUs. Pediatrics 2006 Oct;**118**(4):1332-1340.

28 Resar RK, Rozich JD, Simmonds T, Haraden CR. A trigger tool to identify adverse events in the intensive care unit. Jt Comm J Qual Patient Saf 2006 Oct;**32**(10):585-590.

Christian von Plessen, MD
Department of Pulmonary and Infectious Diseases
Hillerød Hospital
Dyrehavevej 29
DK 3400 Hillerød
cple@hih.regionh.dk
Fax,: +45 4829 3935

7/24/2012

Dear Sir

Thank you for reviewing our manuscript *Experiences with the Global Trigger Tool in Denmark – a pilot study in a non-English speaking country.*

We would like to change the title to *Experiences with Global Trigger Tool reviews in five Danish hospitals – an implementation study* because it fits better with the content of the paper.

We have extensively revised the text according to the reviewers' comments and marked changes with "track changes". Furthermore, we explain the changes below.

Best regards

Christian von Plessen

Reviewer 1

Research question:
The research question/aim is clearly stated in the abstract, but should be clarified further in the background section
###
We have rephrased the last paragraph of the introduction to clarify this point.
###

Study design:
The study is mainly descriptive. A more structured approach towards data collection would have enhanced the quality – the anecdotal information could advantageously have been replaced by more systematic data collection using e.g. interview or questionnaire.
###
We agree, a more systematic collection of contextual/anecdotal data, e.g. with a questionnaire, would have yielded more reliable results. However at this stage, we found an explorative approach adequate to collect contextual information on the training and review processes of the GTT teams. We have added some detail on how we collected the data in the method section to clarify the point.
###

Description of methods:
The methods section commences with a description of the GTT reviews, however, a precise description of the study described in the paper would have been preferred, e.g. there is lack of information on the authors and their role in the study – who are they and how did they collect the information? This is key since the study is observational.
###
Thank you for this comment. We have rearranged the paragraphs in the methods section and give more detailed information regarding these relevant questions.
###

Data on collection of background information is provided, but how was data on experiences with the implementation collected, by whom and when?
###
We have added this information in the methods section.
###

Experienced nurses carried out the GTT review. Did they have experience in reviewing charts or did they have certain clinical experiences?
###
The authors know most of the nurses doing the reviews but we did not systematically collect specific information on how experienced they were and have thus deleted the word “experienced” from the paragraph.
###

A description of the consensus processes involving the two primary reviewers as well as the whole team (primary reviewers and physician) would contribute to the understanding of differences between the compared review teams, e.g. did the primary reviewers reach consensus in a face to face discussion, and did the three of them meet for the final decisions?
###

There were differences in how the teams reviewed met. We have expanded the description of the process in the text otherwise the data are presented in Table 2. Further detail regarding this question would require new interviews.
###

The nurses who categorized harms – what clinical experience did they have? Were they affiliated with the hospitals where the harms were identified?
###
The two nurses are experienced clinicians; they work as quality coordinators at two departments of one of the hospitals participating in the Safer Hospital program and have extensive experience with GTT reviews. We have added the latter information in the methods section of the manuscript.
###

The use of DPSD data in this study is not clear. The aim is to describe experiences with GTT – how does event reporting contribute to this? If used, it is not clear what period the DPSD data originates from.
###
We present the PSI reporting rates as background information that should be considered a measure of safety culture at the hospitals. We have moved this part to the background in the methods section, added the time period (2010) and an explanation to make this point clearer.
###

Discussion:
A discussion of the strengths and limitations of the study is missing. Unanswered questions and specific proposals for future research are not described – what elements of the GTT's measurement properties need further investigation?
###
We agree, and we have added paragraphs with strengths/limitations and future research.
###

The study is – in spite of its scientific limitations – highly relevant in that it provides insights of the variation of harm rates related to the use of GTT and calls for further research of a widely implemented tool that is used locally, nationally and internationally.
###
☺
###

IHI state that in determining whether an adverse event occurred, it should be considered that an AE is defined as unintended harm to a patient from the viewpoint to the patient. How was this managed? Did it lead to any discussions between nurses and physicians that support the conclusion that nurses are more inclusive.
###
An interesting question, but unfortunately we do not have this information.
###

It is stated that each AE/harm was assigned to only one type category. Does this indicate that sometimes a decision had to be made as to which of two or more categories to choose? If yes, what is the implication of this?
###
Yes, the two nurses discussed these cases and usually came to an agreement. When they could not the case was discussed with one of the authors, CvP. So, an element of judgment is inherent to this approach. On the other hand the same nurses categorized all harms from all hospitals. Moreover we used categories of harms that make sense clinically and are documented in the literature.
###

It is stated in the discussion that the hospital with the highest PSI rate observed a significant increase in harms. This is interpreted as a result of a change in the culture of reporting and documentation. Are there any other possible explanations? Has the potential change in reviewer's attention once they have learned that specific event types occur frequently (e.g. pressure ulcers and gastrointestinal complications) been investigated?
###
One would assume that reviewers can become biased towards frequent or otherwise "prominent" harms. On the other hand, the use of triggers should at least partly prevent such a development. This would be interesting idea for further research. We have changed the paragraph slightly to make it clearer.
###

The differences in documentations systems are described, though not conferred any special significance. Is it possible that the layout/the presentation of data in the different systems affects the findings – that some data is more eyecatching in some systems than in another – or are layout properties the same across systems?
###
We have not studied this aspect of the reviews but it would be an interesting question for further research. The point has been added in the discussion section of the manuscript.
###

If DPSD data are to be retained in the paper, a discussion of differences in chart review and event reporting would be of value; one would not expect the same types of events to be identified with the two methods – results must be interpreted in light of this.
###
As mentioned above, we use the DPSD data as background and did not intend to compare the two approaches.
###

The discussion mainly concerns the variation in harm rates. It would also be interesting to learn more about the differences in the assessment of harm types and consequences, e.g. what would be the practical implications of these variations? And as a curiosum: how does a hospital use the knowledge that most harms are in the 'others' category for safety improvement?
###
We agree, that it would be very interesting to study the use of harm types or, as some of the hospitals in the Safer Hospital Program call it, the harm profiles but we do not have data to explore this question. As you point out the 'others' category is not useful for preventing adverse events. However, for the purpose of this manuscript we found it necessary to limit the number of categories.
###

Although not stated directly in the conclusion, it is indicated that the authors do not interpret the results in a way that affect the decision to use the GTT. The authors recommend that health care staff and policy makers should be aware of the 'variation-problem'/the need for sufficient training and retraining of review teams – is it possible to specify this awareness; given the results of this study – what can GTT be used for – what shouldn't it be used for? What would be the 'dangers' if the limitations are not taken into account in the use of the GTT results? Also, is training and retraining of teams sufficient to consider GTT a valid tool and to recommend further use, even before further scientific evaluation of the measurement properties of the tool?
###
In our opinion the study indicates caveats that should be considered when implementing the GTT in a new setting rather than limitations of the method as such. Therefore a discussion about the properties of the GTT as a method was not included in the paper. However, our findings support the recommendation that the GTT should not be used to compare hospitals. We have added a sentence on this in the concluding paragraph.
###

A discussion of the idea of a 'global' measure of safety would be interesting – can a tool be 'global' if it does not measure omissions and is based on what is registered in the medical record (i.e does not include documentation errors and administrative processes leading to harm and don't take into account the patients' experience of patient safety)? Is it possible at all to develop a true global measure?
###
An interesting discussion, but the aim of our paper was to describe the experiences with the GTT method as it is. We feel that a critique of the method as such is beyond the scope of the paper.
###

Do any alternatives to using GTT exist and does electronic capture of triggers have the potential to reduce the problems described in this study?
###
We find that alternative methods to identify harms and electronic trigger capture, although important and relevant areas, reach beyond the scope of this paper which aims to support teams in implementing the manual GTT reviews.
###

2. Reviewer: Ellen Tveter Deilkås MD PhD
Clinical consultant/ Senior advisor,
Akershus University Hospital/The Norwegian Knowledge Centre

I have no competing interests

1. A) The aim of the study is to describe experiences with the Global Trigger Tool (GTT) in Denmark in order to identify ways to improve the performance of the GTT review teams. Since the study only presents characteristics and procedures of GTT teams at five hospitals and their quantitative results, and some anecdotal information, the objective of the study seems more to be to "present experiences with...." rather than to “describe “them”. I suggest that the objective is adjusted accordingly.
###
We agree, and have changed the wording of the aim of the study.
###

B) The authors argue that it is necessary to calibrate the GTT instrument and GTT teams before the instrument is adopted to evaluate safety performance in hospitals across health systems. It does not mention that the GTT manual (page 29) warns against using the instrument to compare results between hospitals: “The IHI Global Trigger Tool is meant to be used as a mechanism to track your organization’s progress over time. Although efforts are made to maintain a standard of training and process for the IHI Global Trigger Tool, organizations will vary in the skill of reviewers and other aspects of the IHI Global Trigger Tool process. We assume this bias is relatively stable over time in a given organization. The stability over time allows comparison to your own organization over time, but is not as useful in comparing between organizations. You can use national data to determine if your rates are in the general range of others. Organizations that have decreased adverse event rates should also be contacted to learn how this was achieved, even if the data is not exactly the same as yours.”
To prevent the article from evaluating the instrument for a purpose it is not made to fulfill, it should be precise about what the purpose of the instrument is.
###
We agree that GTT is not a benchmarking tool. However, we were truly surprised how large the differences in harm rates in the five hospitals were. We have changed the discussion section to clarify this aspect.
###

2. A) The conclusion draws support from a recent study from Sweden which has studied interrater reliability between five teams from different hospitals. Harm rates between these teams ranged from 27,2 to 99,7 per 1000 patient days, with a pairwise interrater reliability ranging from a kappa value of 0,26 to 0,77. The article does not mention that the team in the Swedish study, with the highest harm rate, team IV, used a different definition for harm (The Swedish National Board of Health and Welfare’s definition of AE: ‘Any suffering, discomfort, bodily or mental injury, illness or death caused by healthcare and which is not an inevitable consequence of the patient’s condition or an expected effect of the treatment received by the patient because of her/his condition’), than the other four teams, which used the GTT definition of harm. The harm rates between the four teams which did use the same definition for harm, ranged from 27,2 to 33,2 AE’s per 1000 patient days, with a pairwise interrater reliability kappa value estimate of 0,62 ranging from 0,38 to 0,81. These results are not that bad, and should be taken into account in the discussion.

###
Thank you for pointing this out. As you mention, the Swedish definition of an AE is somewhat broader than the GTT definition of harm which could explain some of the difference between team IV and the other teams. Notably, the team with the highest harm rate had not attended the network meetings with the other teams. The reference has been omitted from the manuscript.
###

B) The article with reference number 19, referred to on page 12, is imprecisely cited (variation of harm by hospital was between 23, 1% and 37, 9%, and not 19,4 5 and 37,9%).
###
Thank you, we mistook the lower CI for the value.
###

1 A) The results in the study are presented with only one of the measures that the instrument provides: Patient safety incidence per 1000 patient days. Since the intention of the article is to present experiences with the Global Trigger Tool, it should also present the results for the two other measurements, which the GTT provides: Adverse events per 100 admissions; and Percent of admissions with an adverse event.
###
We used the most widely known measure of harms per 1000 bed days because the focus of the paper was on the variation of harms across hospitals. We have added the harms per 100 admissions for each hospital under results. In our view the percentage of harmed patients, while useful in quality improvement, does not add much information with regard to the topic of this article. However, we can calculate these rates but will need more time than the editor permitted because of the summer holiday in Denmark.
###

B) The study concludes that the way the GTT teams perform the reviews, strongly contribute to the differences in harm rates between the hospitals and suggests measures to improve and standardize the conditions for the GTT teams. It should be clear about that the differences in results between hospitals, and the reasons for them, do not contradict the purpose of the instrument.
###
We agree and have changed parts of the discussion accordingly.
###

BMJ open

Experiences with Global Trigger Tool reviews in five Danish hospitals – an implementation study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-001324.R2
Article Type:	Research
Date Submitted by the Author:	05-Sep-2012
Complete List of Authors:	von Plessen, Christian; Hillerød Hospital, Department of pulmonary medicine and infectious diseases Kodal, Anne; Hilleroed Hospital, Department of Anesthesiology Anhøj, Jacob; Danish Society for Patient Safety,
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Health services research
Keywords:	AUDIT, Health & safety < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Risk management < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

Title

Experiences with Global Trigger Tool reviews in five Danish hospitals – an implementation study

Authors

Christian von Plessen, MD (corresponding author)
Department of Pulmonary and Infectious Diseases
Hillerød Hospital
Dyrehavevej 29
DK 3400 Hillerød
cple@hih.regionh.dk
Fax: +45 4829 3935

Anne Marie Kodal, RN
Department of Anaesthesiology
Hillerød Hospital
Dyrehavevej 29
DK 3400 Hillerød

anmak@hih.regionh.dk

Jacob Anhøj, MD, DIT
Danish Society for Patient Safety
c/o Hvidovre Hospital
P610 Kettegård Alle 30
2650 Hvidovre

jacob.anhoej@regh.dk

Contributions

CvP, JA and AMK designed the study and collected data. CvP drafted the manuscript, JA and AMK revised and approved. JA performed the statistical analyses and produced the graphs.

Copyright

Abstract

Objectives: To describe experiences with the implementation of Global Trigger Tool (GTT) reviews in five Danish hospitals and to suggest ways to improve the performance of GTT review teams

Design: Retrospective **o**bservational study

Setting: The measurement and monitoring of harms is crucial to campaigns to improve the safety of patients. Increasingly, teams use the GTT to review patient records and measure harms in English and non-English-speaking countries. Meanwhile, it is not clear how the method performs in so diverse settings.

Participants: Review teams from five Danish pilot hospitals of the national Danish Safer Hospital Program

Primary and secondary outcome measures: We collected harm rates, background and anecdotal information and reported patient safety incidents (PSIs) from five pilot hospitals currently participating in the Danish Safer Hospital Programme. Experienced reviewers categorised harms by type. We plotted harm rates as run-charts and applied rules for the detection of patterns of non-random variation.

Results: The hospitals differed in size but had similar patient populations and activity. PSIs varied between 3 and 12 per 1000 patient days. The average harm rate for all hospitals was 60 per 1000 patient days ranging from 34 to 84. The percentage of harmed patients was 25 and ranged from 18 to 33. Overall, 96% of harms were temporary. Infections, pressure ulcers procedure related and gastrointestinal problems were common. Teams reported differences in training and review procedures such as the role of the secondary reviewer.

Conclusions: We found substantial variation of harm rates. Differences in training, review procedures and documentation in patient records probably contributed to these variations. Training reviewers as teams, specifying the roles of the different reviewers, training records and a database for findings of reviews may improve the application of the GTT.

Background

Patients run a high risk of being harmed during hospital admissions. Adverse events occur in up to 10 % of hospitalisations and can cause death, permanent or temporary disability.[1] For patients and health care workers, these harms and the underlying flaws of their health care systems that permit them to happen are deeply upsetting and completely unacceptable.

To improve the safety of patients, national and regional campaigns have been carried through [2-5] or are ongoing.[6] Improvements have been achieved in some areas such as reductions of catheter-related blood stream infections.[7] However, system wide progress is slow [8] and improvements are often limited to particular medical conditions or institutions. Indeed, a recent study [9] from the state of North Carolina, an active participant in large scale patient safety initiatives, concluded that overall rates of harm during 2002-2007 were not reduced. Thus, the challenge to improve the safety of patients in hospitals remains and specific and sensitive measures of harms are needed to assess and monitor the effects of changes to make hospitals safer.

In Denmark, the Operation Life campaign during 2006-2008 focused on patient safety in intensive care and during surgery. An estimated 1654 fewer patients died in the Danish population of 5.5 million during the campaign.[10] In 2010, another campaign the Safer Hospital Program (www.sikkerpatient.dk/fagfolk/patientsikkert-sygehus.aspx) was launched at five pilot hospitals to reduce mortality by 15 % and harms by 30 % through the implementation of 12 care bundles. The hospitals are required to measure and report harms.

Meanwhile, a gold standard for the measurement of harms does not exist. Methods like voluntary reports only detect a small fraction of harms,[11] chart reviews have low interrater reliability [12] and are very time consuming and so are direct observations of health care processes.[13] Studies comparing different methods of harm detection have found very little overlap of the detected harms.[14] Therefore, complete estimates of the incidence of harms probably require the combination of different methods. Meanwhile, such an approach is time consuming and results are often delayed, which is unsuitable for patient safety campaigns in which frequent and regular measurements of

harms are needed to evaluate and monitor the effects of interventions and organisational changes.

The Global Trigger Tool (GTT) has been developed for the purpose of monitoring harms at low cost.[15] Harm in this context is defined as an "Unintended physical injury resulting from or contributed to by medical care that requires additional monitoring, treatment or hospitalisation, or that results in death."[16] Thus the tool measures factual harm to patients while errors not leading to harm, near errors and errors of omission, are not included. A GTT-review is a trigger based chart audit of closed patient charts. Two reviewers, usually nurses or pharmacists, each review a limited number of randomly chosen charts with a given set of 56 triggers or hints of errors. The finding of such a hint triggers an investigation into whether and, if so, how severely, a patient actually has been harmed. Review time is limited to 20 minutes per admission. Finally, the two reviewers compare their conclusions and a supervisor, usually a physician, qualifies the number and severity of harms and decides in cases of disagreement. The number of harms is then expressed as a rate, e.g. harms per 1000 bed days. It has been suggested that GTT teams need a limited amount of training and practice to achieve good levels of reliability to identify harms.[16-18] The feasibility of the method invites for rapid adoption in health care systems around the world where practical ways to measure harms are much in demand. Nevertheless, experiences with the GTT in non-English speaking countries are limited. Thus careful calibration of the instrument and the review team that uses it is warranted to avoid evaluating the safety performance of hospitals with imprecise measurements.

A team of Danish experts translated the GTT to Danish [19] from the English and a Swedish version.[20] The tool was tested in four hospitals in different health regions.[21] The harm rate in these hospitals was around 20 per 1000 bed days. A recent report of harms to Danish patients with cancer found a rate of 68 per 1000 bed days.[22] Notwithstanding these variable rates, policy makers advocate the widespread [23] implementation of GTT reviews in Danish hospitals. Meanwhile in our opinion, it is not sufficiently clear, how the tool performs in the hands of Danish review teams.

The aim of this study was to describe experiences with the GTT in five Danish hospitals and suggest ways to improve the performance of GTT review teams and thus contribute to the accurate measurements of harms.

Methods

In this retrospective observational study, we present harm rates as measured by the GTT at the five hospitals participating in the Safer Hospital Program in Denmark. The hospitals had used the GTT for 18 months from January 2010 until June 2011. The harm rates are registered in a common database. The project managers at the hospitals supplied tabulated data on the size, activity and patient populations of the five hospitals from the year 2010. All data were collected between August and December 2011.

CvP, the project manager at one of the five hospitals and a consultant in pulmonary medicine and AMK, a registered nurse and experienced GTT reviewer interviewed members of the GTT teams and project managers on the telephone. The interviews comprised questions on the training of the GTT teams, team composition, roles of the team members and review processes. Moreover, they asked open questions about unexpected observations and changes. The GTT teams supplied lists of the recorded harms from their reviews.

To give an impression of the safety culture at the participating hospitals, we present the number and severity of reported patient safety incidents (PSIs) in 2010. JA, a physician by training who works in the Danish Society for Patient Safety, gathered these data from the Danish Patient Safety Database (www.dpsd.dk) for voluntary reporting of PSIs. Risk managers at the five hospitals classify the severity of PSIs in the Danish Safety Database into mild, moderate, severe and fatal. In Denmark, reporting of PSIs is mandatory, confidential and sanction free for health care personnel.

GTT review

The Danish translation of the GTT toolkit was the reference for the review teams.[19] Teams were to review a random sample of 10 admissions twice a month. Closed admissions of patients of at least 18 years of age and of at least 24 hours duration were eligible. The date of discharge was the index date. The GTT teams should review all available information from the admissions, i.e. physician and nurse notes, medication orders and history as well as results of diagnostic tests. Each primary reviewer should

review each record. Afterwards the primary reviewers should meet to discuss harms and come to consensus. Finally the primary reviewers should meet with the supervising physician to present their findings for approval and severity classification. Triggers, harms and the severity of harms should be recorded on standardised work sheets, then transferred to spreadsheets and stored locally. The team should classify the severity of harms according to the National Council for Medication Error Reporting and Prevention Index. The five hospitals do not have a shared database for their GTT findings.

Data analysis

We calculated the monthly harm rate as the total number of harms divided by the total number of patient days multiplied by 1000. The harm rate was then plotted on a run chart and analysed using four run chart rules for the detection of patterns of non-random variation such as shifts or trends.[24] We collected and managed the data in Microsoft Excel v. 2003 and produced the statistical analysis and graphs in R Statistical Software v. 2.13.1.

For the purpose of this study, two nurses from the GTT team of one of the hospitals who had done over 400 GTT reviews each retrospectively categorised harms found at all five hospitals into categories also used by Classen et al.[25] They added gastrointestinal complications and pressure ulcers as categories because these were common types of harm (Figure 1). Each harm was assigned to only one category.

The regional ethical committee deemed an ethical review of the study unnecessary.

Results

Background data of the five pilot hospitals

The five hospitals, one in each Danish health region vary in size but all hospitals have departments of internal medicine, orthopaedic and general surgery as well as of obstetrics and gynaecology. All hospitals use electronic patient records, but to varying degree parts of the documentation, such as nursing notes, are on paper. The populations of patients at the five hospitals were similar with regard to age and gender distribution. There was a four-fold difference in reporting of PSIs among hospitals (Table 1). Meanwhile, the distributions of PSIs by consequence were almost identical (Figure 2). All teams started reviewing for the safety program in May 2010 and reviewed records retrospectively from January 2010.

Table 1: Background information on the five hospitals (2010)

	Hillerød	Horsens	Kolding	Næstved	Thy-Mors
Discharges	60098	30377	27526	28677	11836
Average patient age (years)	55	57	53	59	58
Percent females	62	59	61	55	55
Patient days	231978	108060	90710	113353	49711
Outpatient visits	262547	212899	124184	184374	65165
Employees	3163	1367	1507	1668	689
Hospital Standardised Mortality Rate	95	97	96	112	100

Reported safety incidents	2736	365	923	1182	223
Reported patient safety incidents per 1000 patient days	12	3	10	10	4

Experiences with the implementation of the GTT

Reviewing patient files with the GTT was new to three of the five teams. Prior to the Safer Hospitals programme, Næstved and Hillerød Hospital had been using the GTT for reviewing patient records for one and two years respectively.

In May 2010, the GTT team from each of the five hospitals participated in a seven hour training session with experts in the method from Denmark and the Institute of Health Care Improvement (IHI). The session included an introduction to the Danish GTT manual, frontal teaching, review of three training records per team and plenary discussions of the findings (Table 2). Only the review team at Hillerød had in 2008 received a similar training. All teams used nurses as primary reviewers and physicians as supervisors. All teams received on-site expert coaching with reviews of 10 or more records, up to three times. The expert coach was a physician who was trained by experts from the IHI. Furthermore, all teams participated in two full day network seminars during the study period. All teams started reviewing patient records for the measurement of their baseline in May 2010 and retrospectively reviewed records from January to May 2010.

The compositions of teams changed between zero (Hillerød and Næstved) and three times (Thy) during the study period. Review intervals at the hospitals varied between two weeks, one monthly and irregular. Complete teams reviewed together and compared findings at two, later three hospitals. The role of the physician in the review team varied from only judging cases where the primary reviewers were in doubt or

disagreed (Hillerød) to identifying harms based on triggers found by primary reviewers (Horsens). Table 2 shows differences in review procedures at the hospitals.

Table 2: Characteristics and review procedures of GTT teams at five hospitals

	Hillerød	Horsens	Kolding	Næstved	Thy-Mors
Team characteristics					
Number of physicians	2	1	1	2	1
Number of nurses	3	3	4 later 3	4	3
Number of changes in team	0	1 (physician)	1	0	2 (nurses)
Review intervals	Twice per month	Monthly	Monthly (two half days)	Monthly	Variable
Training					
Hours of training	14[#]	7	7	7	7
Site visits by Danish expert (days)	1	3	1	4	2
Complete team present during site visit	-	-	+	-	-f
Number of records reviewed together with expert	<10	>10	>10	>10	>10
Review procedures					
Whole team meets for reviews	+	+	Since Jan. 2011	-	-
Physician acts as judge (J) in cases of disagreement or reviewer (R) based on triggers	J	R	J	J	J

Records entirely electronic	-	+	+	-	+
Dedicated person responsible to find records	+	-	+	+	-
Secretary plots triggers and harms	-	-	-	-	+

Team also trained by national expert in 2008

Anecdotal information about GTT reviews

At Næstved, the team sampled 24 records (in case some were incomplete) each month and sorted them in the order of the date of the admission and reviewed the first 20 records. Thus the sample became biased towards admissions in the earlier part of the month. Moreover, the team initially reviewed only admissions to the last department of a hospital admission. Thus they did not find harms that, for example, occurred during an admission to the intensive care unit earlier during the hospital stay. These errors were accidentally discovered during a site visit and the team did a new review for the period. The team at Kolding discovered after three months that their sampling procedure excluded admissions that had an appointment for ambulatory follow up after surgery, and they decided to discard the first three months from their baseline.

GTT findings

In total, 687 adverse events were identified in 11487 patient days, i.e. the overall average harm rate was 60 per 1000 patient days. The monthly harm rate ranged from 34 to 84 harms per 1000 patient days (Figure 3). The harm rates at all five hospitals showed only random variation.[23] The overall numbers of harms per 100 admissions were at Thy Mors 45, Næstved 24, Kolding 30, Horsens 43 and Hillerød 54. The percentage of harmed patients was 25, ranging from 18 (Horsens) to 33 (Hillerød) (Figure 4).

Overall, 96% of harms were temporary (grades E and F). However, the severity distribution varied between hospitals (Figure 5). Notably, the hospital with the highest harm rate (Hillerød) also had the highest proportion of grade E harms.

553 harms (80%) were recorded with sufficient detail in the hospital datasheets to categorise them by type. The proportion of harms without description varied: 5% (Kolding and Hillerød), 12% (Thy-Mors and Horsens) and 45% (Næstved). Common types of harm were infections, procedure related, pressure ulcers and gastrointestinal problems (Figure 1).

Discussion

We observed marked differences in the harm rates and types identified by GTT review teams in five Danish public hospitals. The GTT is not designed to compare hospitals but we were surprised by the magnitude of the variations. Therefore we designed this study to identify factors that contributed to the differences. The hospitals, their patient populations, structures and activity levels were similar but we found differences in the training, the review procedures and the experience of the review teams.

Other studies have also found variation of harms across hospitals. Naessens et al.[18] in a study of 1138 admissions to three academic health centres in three states of the US found a variation of harms by hospital between 23,1 % and 37,9 % of admissions. In a study of surgical harms the variation was between 5 % and 35 %,[26] Sharek et al.[27] observed harm rates between 0,18 and 1,28 per patient in 749 admissions to 15 newborn intensive care units in the US and Canada and Resar et al.[28] in 62 intensive care units in the US measured between 3,2 and 27,36 harms per 100 days. Thus significant variations of GTT findings seem to be common.

Several factors could explain the variation in rates of harm. First, there can be differences in the safety of the clinical processes at the hospitals. However, it seems unlikely that such differences should cause as much as a 2,5 fold variation of harm rates given the similar patient populations at the five hospitals and the homogeneity of the Danish health care system in general. We even found that the hospital with the highest mean age and the highest hospital standardised mortality rate had the lowest rate of harms. Second, the documentation of triggers and harms probably varies across hospitals. High PSI reporting rates are generally considered a sign of a mature safety culture rather than of poor safety and one could speculate that staff is more likely to document harms in a hospital with such a culture. Interestingly, the hospital with the highest PSI rate also observed a significant increase of harm rates. Also different types of patient records (electronic and paper) and differences in layout and presentation could influence the results of the reviews. We do not have data to explore these questions but we certainly cannot exclude an influence on the different harm rates across hospitals. Third, differences in the training and the experience of the review teams influence the subjective process of judging harms in any record review.[12] The team that found the most harms was the most experienced review team and had attended two training seminars. This team was also stable, reviewed regularly with the

whole team present twice a month. Finally, the teams conducted the review processes in slightly different ways. Most importantly, the roles of nurses and physicians varied. The role of the physician in the review team that found the highest harm rate was to judge in cases of disagreement while physicians in the other teams themselves identified harms. We assume that nurses are more prone to register harms of lower severity, while physicians might consider them insignificant. This interpretation is supported by our finding that the variation of harm rates was greatest in the least severe category "E".

Thus in our opinion, the experience of the GTT teams and the way they perform the reviews strongly contributed to the differences in harm rates in the five hospitals. Moreover, differences in the documentation of harms in the patient records seem to influence the number of harms the GTT team can find. We did not expect these factors to be so important because the GTT was implemented according to current recommendations [15, 17] and was guided by some of their authors. Moreover, all the teams had attended a GTT network meeting with national and international experts and received site visits by a national expert. However these precautions, it seems, were not sufficient to secure at standardised reviewing process at the hospitals. Thus users of the GTT and its results, health care personnel, administrators, payers or the public, should be aware of the challenges of the implementation of the method and allow for sufficient training and evaluation of the results. Our findings also stress that GTT findings should guide hospitals in their efforts to improve patient safety but the results should not be used to compare hospitals.

Strengths, limitations and further research

The strength of this study is its relevance for the implementation of the GTT that increasingly is being used to monitor the safety in hospitals. Our contextual data are detailed and thus practical. The limitations are inherent to the observational nature of the study that prevents conclusions on causal links between the variations of harm rates and the observed differences in team training and review processes. For the same reason, we cannot quantify the contribution of the different factors. Nevertheless, the findings are plausible and fit with the recommendations for the use of the method.[15]

Further research should address how teams' reviewing experience and training influence team performance and how team training can be optimised. Moreover, studies should investigate the influence of changes of documentation and presentation of information in patient charts and the use of the types of harm for improving patient safety.

In conclusion, differences in training, review processes, and documentation probably have contributed to variations in rates of harm as measured by the GTT. Thus health care staff and policy makers should be aware of the need for systematic training of the review teams and standardisation of the review process when implementing the GTT in new settings. These factors are related to the implementation of the GTT reviews and are not inherent to the method as such. Our study has implications for the implementation of the method in other settings and we suggest considering the following interventions to improve the implementation of the GTT in new settings:

- Secure that the review team is trained as a team
- Specify of the roles of the reviewers during the reviews to avoid over-/underestimation of especially harms of lesser severity depending on professional background
- Test review teams' abilities to find harms with a set of training charts to estimate their "sensitivity" before routine monitoring is instituted
- Define a minimum number of patient charts that the team should have reviewed before monitoring harms routinely
- Perform reviews with all team members present
- Ensure a structured review process, i.e. a space where the team can work without interruptions, regular time intervals between reviews to keep team "in shape"
- Implement a common database with individual patient data to allow for re-examination of reviewed charts to avoid problems such as sampling errors

Competing interests

CvP leads the current Safer Hospital campaign at one of the participating hospitals. AMK is a member of the GTT review team at the same hospital. JA is an advisor to the patient safety campaign at the national level. None of us has any financial interests related to this study.

Acknowledgements

GTT teams at the five hospitals: Hillerød: Lisbeth Dammegaard, Anne Marie Kodal, Minna Nielsen, Anne Marie Schlüter, Mette Østergaard. Horsens: Inge-Lise Johansen, Tina Kjær Pedersen, Tom Krintel Petersen (anæstesiologisk afdeling), Kiss Ruben Larsen. Kolding: Helle Guldborg Nielsen, Lynge Kirkegaard, Pernille Kølholt Langkilde, Pernille Lennert, Anita Pedersen, Pia Schrøder Petersen. Næstved: Arne Cyron, Helle Fugl, Kim Garde, Annette Knakkergaard, Nina Pedersen, Trine Zachariassen. Thy Mors: Connie Elbeck van der Koij, Mona Kyndi, Mie Landbo, Bente Ringgaard, Hans-Jörg Selter, Lisbeth Thomsen

Project managers at the five hospitals: Anne Balle Larsen, Søren Brogaard, Marianne Frandsen, Mona Kyndi Pedersen, Kirsten Løth Lysdahl, Søren Schousboe Laursen, Maria Staun, Inge Ulriksen

Funding

CvP received a grant from the Danish Society for Patient Safety to write this article.

Data Sharing Statement

The five hospitals participating in the study measure their harms continuously. These data might also in the future be used for further research into the global trigger tool or other aspects of patient safety.

Legends to figures

Figure 1: Harms by Type

The dot plots show the relative distribution of harms by type. VTE = venous thromboembolism.

Figure 2: Patient Safety Incidents by Consequence

The dot plots show the relative distribution of patient safety incidents by consequence as reported to the Danish national database. Categories *minor* and *moderate* represent no and temporary harms, *major* permanent harms. Overall 96% of the incidents were temporary.

Figure 3: Rates of Harm

The run charts show monthly rates of harms measured with the Global Trigger Tool. The curve shows the harm rate expressed as the number of adverse events per 1000 patient days. The horizontal line is the median harm rate.

Figure 4: Harmed Patients

The run charts show the percentages of harmed patients measured with the Global Trigger Tool. The horizontal line is the median percentage of harms.

Figure 5: Harms by Severity

The dot plots show the relative distribution of severity of harms in categories *E* – *I*, where *E* and *F* are temporary, *G* – *H* permanent harms and *I* death. Overall, 96% of harms were temporary.

References

1 de Vries EN, Ramrattan MA, Smorenburg SM, et al. The incidence and nature of in-hospital adverse events: a systematic review. Qual Saf Health Care 2008 Jun**;17**(3):216-223.

2 Berwick DM, Calkins DR, McCannon CJ, et al. The 100,000 lives campaign: setting a goal and a deadline for improving health care quality. JAMA 2006 Jan 18**;295**(3):324-327.

3 Benning A, Dixon-Woods M, Nwulu U, et al. Multiple component patient safety intervention in English hospitals: controlled evaluation of second phase. BMJ 2011 Feb 3**;342**:d199.

4 Benning A, Ghaleb M, Suokas A, et al. Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation. BMJ 2011 Feb 3**;342**:d195.

5 McCannon CJ, Schall MW, Calkins DR, et al. Saving 100,000 lives in US hospitals. BMJ 2006 Jun 3**;332**(7553):1328-1330.

6 NHS Scotland. Scottish Patient Safety Programme. Available at: http://www.patientsafetyalliance.scot.nhs.uk/programme. Accessed 11/03, 2011.

7 Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. N Engl J Med 2006 Dec 28**;355**(26):2725-2732.

8 Vincent C, Aylin P, Franklin BD, et al. Is health care getting safer? BMJ 2008 Nov 13**;337**:a2426.

9 Landrigan CP, Parry GJ, Bones CB, et al. Temporal trends in rates of patient harm resulting from medical care. N Engl J Med 2010 Nov 25**;363**(22):2124-2134.

10 Dansk Selskab for Patientsikkerhed. The Danish Safer Hospital Programme. 2011; Available at: http://www.sikkerpatient.dk/about-sikker-patient/the-danish-safer-hospital-programme.aspx. Accessed 11/03, 2011.

11 Sari AB, Sheldon TA, Cracknell A, Turnbull A. Sensitivity of routine system for reporting patient safety incidents in an NHS hospital: retrospective patient case note review. BMJ 2007 Jan 13**;334**(7584):79.

12 Thomas EJ, Lipsitz SR, Studdert DM, et al. The reliability of medical record review for estimating adverse event rates. Ann Intern Med 2002 Jun 4**;136**(11):812-816.

13 Buckley MS, Erstad BL, Kopp BJ, et al. Direct observation approach for detecting medication errors and adverse drug events in a pediatric intensive care unit. Pediatr Crit Care Med 2007 Mar**;8**(2):145-152.

14 Naessens JM, Campbell CR, Huddleston JM, et al. A comparison of hospital adverse events identified by three widely used detection methods. Int J Qual Health Care 2009 Aug**;21**(4):301-307.

15 Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events. IHI Innovation Series White paper. 2009.

16 Classen DC., Lloyd RC., Provost L., et al. Development and Evaluation of the Institute for Healthcare Improvement Global Trigger Tool. Journal of Patient Safety 2008**;4**(3):169-177.

17 Adler L, Denham CR, McKeever M, et al. Global Trigger Tool: Implementation Basics. J Patient Saf 2008**;4**:245-249.

18 Naessens JM, O'Byrne TJ, Johnson MG, et al. Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool. Int J Qual Health Care 2010 Aug**;22**(4):266-274.

19 Anonymous. Global Trigger Tool. En oversættelse af IHI's værktøj tilpasset danske forhold. Center for Kvalitet. Region Syddanmark 2008.

20 Anonymous. Strukturerat journalgranskning för att identifiera och mätta förekomst av skador i vården enligt metoden Global Trigger Tool. Institute for Health Care Improvement Innovation series 2007, Svensk översätning och anpassning. Stockholm. Kommentus Förlag 2008.

21 Rabøl LI, Jølving LR, Bjørn B et al. Systematisk journalgennemgang med IHI's Global Trigger Tool - Erfaringer fra afprøvning i danske sygehuse. Copenhagen. Dansk Selskap for Patientsikkerhed 2008.

22 Anonymous. Patientsikkerhed i kræftbehandlingen - journalgennemgang med Global Trigger Tool. Rapport nr 3/2010. Copenhagen. Kræftens Bekæmpelse 2010.

23 Auerbach AD, Landefeld CS, Shojania KG. The tension between needing to improve care and knowing how to do it. N Engl J Med 2007 Aug 9**;357**(6):608-613.

24 Perla RJ, Provost LP, Murray SK. The run chart: a simple analytical tool for learning from variation in healthcare processes. BMJ Qual Saf 2011 Jan**;20**(1):46-51.

25 Classen DC, Resar R, Griffin F, et al. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. Health Aff (Millwood) 2011 Apr**;30**(4):581-589.

26 Griffin FA, Classen DC. Detection of adverse events in surgical patients using the Trigger Tool approach. Qual Saf Health Care 2008 Aug**;17**(4):253-258.

27 Sharek PJ, Horbar JD, Mason W, et al. Adverse events in the neonatal intensive care unit: development, testing, and findings of an NICU-focused trigger tool to identify harm in North American NICUs. Pediatrics 2006 Oct**;118**(4):1332-1340.

28 Resar RK, Rozich JD, Simmonds T, et al. A trigger tool to identify adverse events in the intensive care unit. Jt Comm J Qual Patient Saf 2006 Oct**;32**(10):585-590.

Title

Experiences with Global Trigger Tool reviews in five Danish hospitals – an implementation study

Authors

Christian von Plessen, MD (corresponding author)
Department of Pulmonary and Infectious Diseases
Hillerød Hospital
Dyrehavevej 29
DK 3400 Hillerød
cple@hih.regionh.dk
Fax: +45 4829 3935

Anne Marie Kodal, RN
Department of Anaesthesiology
Hillerød Hospital
Dyrehavevej 29
DK 3400 Hillerød

anmak@hih.regionh.dk

Jacob Anhøj, MD, DIT
Danish Society for Patient Safety
c/o Hvidovre Hospital
P610 Kettegård Alle 30
2650 Hvidovre

jacob.anhoej@regh.dk

Contributions

CvP, JA and AMK designed the study and collected data. CvP drafted the manuscript, JA and AMK revised and approved. JA performed the statistical analyses and produced the graphs.

Copyright

Abstract

Objectives: To describe experiences with the implementation of Global Trigger Tool (GTT) reviews in five Danish hospitals and to suggest ways to improve the performance of GTT review teams

Design: Retrospective observational study

Setting: The measurement and monitoring of harms is crucial to campaigns to improve the safety of patients. Increasingly, teams use the GTT to review patient records and measure harms in English and non-English-speaking countries. Meanwhile, it is not clear how the method performs in so diverse settings.

Participants: Review teams from five Danish pilot hospitals of the national Danish Safer Hospital Program

Primary and secondary outcome measures: We collected harm rates, background and anecdotal information and reported patient safety incidents (PSIs) from five pilot hospitals currently participating in the Danish Safer Hospital Programme. Experienced reviewers categorised harms by type. We plotted harm rates as run-charts and applied rules for the detection of patterns of non-random variation.

Results: The hospitals differed in size but had similar patient populations and activity. PSIs varied between 3 and 12 per 1000 patient days. The average harm rate for all hospitals was 60 per 1000 patient days ranging from 34 to 84. The percentage of harmed patients was 25 and ranged from 18 to 33. Overall, 96% of harms were temporary. Infections, pressure ulcers procedure related and gastrointestinal problems were common. Teams reported differences in training and review procedures such as the role of the secondary reviewer.

Conclusions: We found substantial variation of harm rates. Differences in training, review procedures and documentation in patient records probably contributed to these variations. Training reviewers as teams, specifying the roles of the different reviewers, training records and a database for findings of reviews may improve the application of the GTT.

Background

Patients run a high risk of being harmed during hospital admissions. Adverse events occur in up to 10 % of hospitalisations and can cause death, permanent or temporary disability.[1] For patients and health care workers, these harms and the underlying flaws of their health care systems that permit them to happen are deeply upsetting and completely unacceptable.

To improve the safety of patients, national and regional campaigns have been carried through [2-5] or are ongoing.[6] Improvements have been achieved in some areas such as reductions of catheter-related blood stream infections.[7] However, system wide progress is slow [8] and improvements are often limited to particular medical conditions or institutions. Indeed, a recent study [9] from the state of North Carolina, an active participant in large scale patient safety initiatives, concluded that overall rates of harm during 2002-2007 were not reduced. Thus, the challenge to improve the safety of patients in hospitals remains and specific and sensitive measures of harms are needed to assess and monitor the effects of changes to make hospitals safer.

In Denmark, the Operation Life campaign during 2006-2008 focused on patient safety in intensive care and during surgery. An estimated 1654 fewer patients died in the Danish population of 5.5 million during the campaign.[10] In 2010, another campaign the Safer Hospital Program (www.sikkerpatient.dk/fagfolk/patientsikkert-sygehus.aspx) was launched at five pilot hospitals to reduce mortality by 15 % and harms by 30 % through the implementation of 12 care bundles. The hospitals are required to measure and report harms.

Meanwhile, a gold standard for the measurement of harms does not exist. Methods like voluntary reports only detect a small fraction of harms,[11] chart reviews have low interrater reliability [12] and are very time consuming and so are direct observations of health care processes.[13] Studies comparing different methods of harm detection have found very little overlap of the detected harms.[14] Therefore, complete estimates of the incidence of harms probably require the combination of different methods. Meanwhile, such an approach is time consuming and results are often delayed, which is unsuitable for patient safety campaigns in which frequent and regular measurements of

harms are needed to evaluate and monitor the effects of interventions and organisational changes.

The Global Trigger Tool (GTT) has been developed for the purpose of monitoring harms at low cost.[15] Harm in this context is defined as an "Unintended physical injury resulting from or contributed to by medical care that requires additional monitoring, treatment or hospitalisation, or that results in death."[16] Thus the tool measures factual harm to patients while errors not leading to harm, near errors and errors of omission, are not included. A GTT-review is a trigger based chart audit of closed patient charts. Two reviewers, usually nurses or pharmacists, each review a limited number of randomly chosen charts with a given set of 56 triggers or hints of errors. The finding of such a hint triggers an investigation into whether and, if so, how severely, a patient actually has been harmed. Review time is limited to 20 minutes per admission. Finally, the two reviewers compare their conclusions and a supervisor, usually a physician, qualifies the number and severity of harms and decides in cases of disagreement. The number of harms is then expressed as a rate, e.g. harms per 1000 bed days. It has been suggested that GTT teams need a limited amount of training and practice to achieve good levels of reliability to identify harms.[16-18] The feasibility of the method invites for rapid adoption in health care systems around the world where practical ways to measure harms are much in demand. Nevertheless, experiences with the GTT in non-English speaking countries are limited. Thus careful calibration of the instrument and the review team that uses it is warranted to avoid evaluating the safety performance of hospitals with imprecise measurements.

A team of Danish experts translated the GTT to Danish [19] from the English and a Swedish version.[20] The tool was tested in four hospitals in different health regions.[21] The harm rate in these hospitals was around 20 per 1000 bed days. A recent report of harms to Danish patients with cancer found a rate of 68 per 1000 bed days.[22] Notwithstanding these variable rates, policy makers advocate the widespread [23] implementation of GTT reviews in Danish hospitals. Meanwhile in our opinion, it is not sufficiently clear, how the tool performs in the hands of Danish review teams.

The aim of this study was to describe experiences with the GTT in five Danish hospitals and suggest ways to improve the performance of GTT review teams and thus contribute to the accurate measurements of harms.

Methods

In this retrospective observational study, we present harm rates as measured by the GTT at the five hospitals participating in the Safer Hospital Program in Denmark. The hospitals had used the GTT for 18 months from January 2010 until June 2011. The harm rates are registered in a common database. The project managers at the hospitals supplied tabulated data on the size, activity and patient populations of the five hospitals from the year 2010. All data were collected between August and December 2011.

CvP, the project manager at one of the five hospitals and a consultant in pulmonary medicine and AMK, a registered nurse and experienced GTT reviewer interviewed members of the GTT teams and project managers on the telephone. The interviews comprised questions on the training of the GTT teams, team composition, roles of the team members and review processes. Moreover, they asked open questions about unexpected observations and changes. The GTT teams supplied lists of the recorded harms from their reviews.

To give an impression of the safety culture at the participating hospitals, we present the number and severity of reported patient safety incidents (PSIs) in 2010. JA, a physician by training who works in the Danish Society for Patient Safety, gathered these data from the Danish Patient Safety Database (www.dpsd.dk) for voluntary reporting of PSIs. Risk managers at the five hospitals classify the severity of PSIs in the Danish Safety Database into mild, moderate, severe and fatal. In Denmark, reporting of PSIs is mandatory, confidential and sanction free for health care personnel.

GTT review

The Danish translation of the GTT toolkit was the reference for the review teams.[19] Teams were to review a random sample of 10 admissions twice a month. Closed admissions of patients of at least 18 years of age and of at least 24 hours duration were eligible. The date of discharge was the index date. The GTT teams should review all available information from the admissions, i.e. physician and nurse notes, medication orders and history as well as results of diagnostic tests. Each primary reviewer should

review each record. Afterwards the primary reviewers should meet to discuss harms and come to consensus. Finally the primary reviewers should meet with the supervising physician to present their findings for approval and severity classification. Triggers, harms and the severity of harms should be recorded on standardised work sheets, then transferred to spreadsheets and stored locally. The team should classify the severity of harms according to the National Council for Medication Error Reporting and Prevention Index. The five hospitals do not have a shared database for their GTT findings.

Data analysis

We calculated the monthly harm rate as the total number of harms divided by the total number of patient days multiplied by 1000. The harm rate was then plotted on a run chart and analysed using four run chart rules for the detection of patterns of non-random variation such as shifts or trends.[24] We collected and managed the data in Microsoft Excel v. 2003 and produced the statistical analysis and graphs in R Statistical Software v. 2.13.1.

For the purpose of this study, two nurses from the GTT team of one of the hospitals who had done over 400 GTT reviews each retrospectively categorised harms found at all five hospitals into categories also used by Classen et al.[25] They added gastrointestinal complications and pressure ulcers as categories because these were common types of harm (Figure 1). Each harm was assigned to only one category.

The regional ethical committee deemed an ethical review of the study unnecessary.

Results

Background data of the five pilot hospitals

The five hospitals, one in each Danish health region vary in size but all hospitals have departments of internal medicine, orthopaedic and general surgery as well as of obstetrics and gynaecology. All hospitals use electronic patient records, but to varying degree parts of the documentation, such as nursing notes, are on paper. The populations of patients at the five hospitals were similar with regard to age and gender distribution. There was a four-fold difference in reporting of PSIs among hospitals (Table 1). Meanwhile, the distributions of PSIs by consequence were almost identical (Figure 2). All teams started reviewing for the safety program in May 2010 and reviewed records retrospectively from January 2010.

Table 1: Background information on the five hospitals (2010)

	Hillerød	Horsens	Kolding	Næstved	Thy-Mors
Discharges	60098	30377	27526	28677	11836
Average patient age (years)	55	57	53	59	58
Percent females	62	59	61	55	55
Patient days	231978	108060	90710	113353	49711
Outpatient visits	262547	212899	124184	184374	65165
Employees	3163	1367	1507	1668	689
Hospital Standardised Mortality Rate	95	97	96	112	100

Reported safety incidents	2736	365	923	1182	223
Reported patient safety incidents per 1000 patient days	12	3	10	10	4

Experiences with the implementation of the GTT

Reviewing patient files with the GTT was new to three of the five teams. Prior to the Safer Hospitals programme, Næstved and Hillerød Hospital had been using the GTT for reviewing patient records for one and two years respectively.

In May 2010, the GTT team from each of the five hospitals participated in a seven hour training session with experts in the method from Denmark and the Institute of Health Care Improvement (IHI). The session included an introduction to the Danish GTT manual, frontal teaching, review of three training records per team and plenary discussions of the findings (Table 2). Only the review team at Hillerød had in 2008 received a similar training. All teams used nurses as primary reviewers and physicians as supervisors. All teams received on-site expert coaching with reviews of 10 or more records, up to three times. The expert coach was a physician who was trained by experts from the IHI. Furthermore, all teams participated in two full day network seminars during the study period. All teams started reviewing patient records for the measurement of their baseline in May 2010 and retrospectively reviewed records from January to May 2010.

The compositions of teams changed between zero (Hillerød and Næstved) and three times (Thy) during the study period. Review intervals at the hospitals varied between two weeks, one monthly and irregular. Complete teams reviewed together and compared findings at two, later three hospitals. The role of the physician in the review team varied from only judging cases where the primary reviewers were in doubt or

disagreed (Hillerød) to identifying harms based on triggers found by primary reviewers (Horsens). Table 2 shows differences in review procedures at the hospitals.

Table 2: Characteristics and review procedures of GTT teams at five hospitals

	Hillerød	Horsens	Kolding	Næstved	Thy-Mors
Team characteristics					
Number of physicians	2	1	1	2	1
Number of nurses	3	3	4 later 3	4	3
Number of changes in team	0	1 (physician)	1	0	2 (nurses)
Review intervals	Twice per month	Monthly	Monthly (two half days)	Monthly	Variable
Training					
Hours of training	14[#]	7	7	7	7
Site visits by Danish expert (days)	1	3	1	4	2
Complete team present during site visit	-	-	+	-	-f
Number of records reviewed together with expert	<10	>10	>10	>10	>10
Review procedures					
Whole team meets for reviews	+	+	Since Jan. 2011	-	-
Physician acts as judge (J) in cases of disagreement or reviewer (R) based on triggers	J	R	J	J	J

Records entirely electronic	-	+	+	-	+
Dedicated person responsible to find records	+	-	+	+	-
Secretary plots triggers and harms	-	-	-	-	+

Team also trained by national expert in 2008

Anecdotal information about GTT reviews

At Næstved, the team sampled 24 records (in case some were incomplete) each month and sorted them in the order of the date of the admission and reviewed the first 20 records. Thus the sample became biased towards admissions in the earlier part of the month. Moreover, the team initially reviewed only admissions to the last department of a hospital admission. Thus they did not find harms that, for example, occurred during an admission to the intensive care unit earlier during the hospital stay. These errors were accidentally discovered during a site visit and the team did a new review for the period. The team at Kolding discovered after three months that their sampling procedure excluded admissions that had an appointment for ambulatory follow up after surgery, and they decided to discard the first three months from their baseline.

GTT findings

In total, 687 adverse events were identified in 11487 patient days, i.e. the overall average harm rate was 60 per 1000 patient days. The monthly harm rate ranged from 34 to 84 harms per 1000 patient days (Figure 3). The harm rates at all five hospitals showed only random variation.[23] The overall numbers of harms per 100 admissions were at Thy Mors 45, Næstved 24, Kolding 30, Horsens 43 and Hillerød 54. The percentage of harmed patients was 25, ranging from 18 (Horsens) to 33 (Hillerød) (Figure 4).

Overall, 96% of harms were temporary (grades E and F). However, the severity distribution varied between hospitals (Figure 5). Notably, the hospital with the highest harm rate (Hillerød) also had the highest proportion of grade E harms.

553 harms (80%) were recorded with sufficient detail in the hospital datasheets to categorise them by type. The proportion of harms without description varied: 5% (Kolding and Hillerød), 12% (Thy-Mors and Horsens) and 45% (Næstved). Common types of harm were infections, procedure related, pressure ulcers and gastrointestinal problems (Figure 1).

Discussion

We observed marked differences in the harm rates and types identified by GTT review teams in five Danish public hospitals. The GTT is not designed to compare hospitals but we were surprised by the magnitude of the variations. Therefore we designed this study to identify factors that contributed to the differences. The hospitals, their patient populations, structures and activity levels were similar but we found differences in the training, the review procedures and the experience of the review teams.

Other studies have also found variation of harms across hospitals. Naessens et al.[18] in a study of 1138 admissions to three academic health centres in three states of the US found a variation of harms by hospital between 23,1 % and 37,9 % of admissions. In a study of surgical harms the variation was between 5 % and 35 %,[26] Sharek et al.[27] observed harm rates between 0,18 and 1,28 per patient in 749 admissions to 15 newborn intensive care units in the US and Canada and Resar et al.[28] in 62 intensive care units in the US measured between 3,2 and 27,36 harms per 100 days. Thus significant variations of GTT findings seem to be common.

Several factors could explain the variation in rates of harm. First, there ~~they~~ can be ~~caused by real~~ differences in the safety of the clinical processes at the hospitals. However, it seems unlikely that such differences should cause as much as a 2,5 fold variation of harm rates given the similar patient populations at the five hospitals and the homogeneity of the Danish health care system in general. ~~Second, differences in case mix at the hospitals could cause the variations. However, the compositions of the patient populations in the five hospitals were similar.~~ We even found that the hospital with the highest mean age and the highest hospital standardised mortality rate had the lowest rate of harms. Second, the documentation of triggers and harms probably varies across hospitals. ~~Interestingly, the hospital with the highest PSI rate also observed a significant increase~~ High PSI reporting rates are generally considered a sign of a mature safety culture rather than of poor safety and one could speculate that staff is more likely to document harms in a hospital with such a culture. Interestingly, the hospital with the highest PSI rate also observed a significant increase of harm rates. Also different types of patient records (electronic and paper) and differences in layout and presentation could influence the results of the reviews. We do not have data to explore these questions but we certainly cannot exclude an influence on the different harm rates across hospitals. Third, differences in the training and the experience of the

review teams influence the subjective process of judging harms in any record review.[12] The team that found the most harms was the most experienced review team and had attended two training seminars. This team was also stable, reviewed regularly with the whole team present twice a month. ~~Interestingly, the harm rate at their hospital demonstrated a significant shift in the months after the second training seminar.~~ Finally, the teams conducted the review processes in slightly different ways. Most importantly, the roles of nurses and physicians varied. The role of the physician in the review team that found the highest harm rate was to judge in cases of disagreement while physicians in the other teams themselves identified harms. We assume that nurses are more prone to register harms of lower severity, while physicians might consider them insignificant. This interpretation is supported by our finding that the variation of harm rates was greatest in the least severe category "E".

Thus in our opinion, the experience of the GTT teams and the way they perform the reviews strongly contributed to the differences in harm rates in the five hospitals. ~~Notably, four of the five teams found rates between~~ Moreover, differences in the documentation of harms in the patient records seem to influence the number of harms the GTT team can find. We did not expect these factors to be so important because the GTT was implemented according to current recommendations [15, 17] and was guided by some of their authors. Moreover, all the teams had attended a GTT network meeting with national and international experts and received site visits by a national expert. However these precautions, it seems, were not sufficient to secure at standardised reviewing process at the hospitals. Thus users of the GTT and its results, health care personnel, administrators, payers or the public, should be aware of the challenges of the implementation of the method and allow for sufficient training and evaluation of the results. Our findings also stress that GTT findings should guide hospitals in their efforts to improve patient safety but the results should not be used to compare hospitals.

Strengths, limitations and further research

The strength of this study is its relevance for the implementation of the GTT that increasingly is being used to monitor the safety in hospitals. Our contextual data are detailed and thus practical. The limitations are inherent to the observational nature of the study that prevents conclusions on causal links between the variations of harm rates

and the observed differences in team training and review processes. For the same reason, we cannot quantify the contribution of the different factors. Nevertheless, the findings are plausible and fit with the recommendations for the use of the method.[15]

Further research should address how teams' reviewing experience and training influence team performance and how team training can be optimised. Moreover, studies should investigate the influence of changes of documentation and presentation of information in patient charts and the use of the types of harm for improving patient safety.

In conclusion, differences in training, review processes, and documentation probably have contributed to variations in rates of harm as measured by the GTT. Thus health care staff and policy makers should be aware of the need for systematic training of the review teams and standardisation of the review process when implementing the GTT in new settings. These factors are related to the implementation of the GTT reviews and are not inherent to the method as such. Our study has implications for the implementation of the method in other settings and we suggest considering the following interventions to improve the implementation of the GTT in new settings:

- Secure that the review team is trained as a team
- Specify of the roles of the reviewers during the reviews to avoid over-/underestimation of especially harms of lesser severity depending on professional background
- Test review teams' abilities to find harms with a set of training charts to estimate their "sensitivity" before routine monitoring is instituted
- Define a minimum number of patient charts that the team should have reviewed before monitoring harms routinely
- Perform reviews with all team members present
- Ensure a structured review process, i.e. a space where the team can work without interruptions, regular time intervals between reviews to keep team "in shape"

- Implement a common database with individual patient data to allow for re-examination of reviewed charts to avoid problems such as sampling errors

Competing interests

CvP leads the current Safer Hospital campaign at one of the participating hospitals. AMK is a member of the GTT review team at the same hospital. JA is an advisor to the patient safety campaign at the national level. None of us has any financial interests related to this study.

Acknowledgements

GTT teams at the five hospitals: Hillerød: Lisbeth Dammegaard, Anne Marie Kodal, Minna Nielsen, Anne Marie Schlüter, Mette Østergaard. Horsens: Inge-Lise Johansen, Tina Kjær Pedersen, Tom Krintel Petersen (anæstesiologisk afdeling), Kiss Ruben Larsen. Kolding: Helle Guldborg Nielsen, Lynge Kirkegaard, Pernille Kølholt Langkilde, Pernille Lennert, Anita Pedersen, Pia Schrøder Petersen. Næstved: Arne Cyron, Helle Fugl, Kim Garde, Annette Knakkergaard, Nina Pedersen, Trine Zachariassen. Thy Mors: Connie Elbeck van der Koij, Mona Kyndi, Mie Landbo, Bente Ringgaard, Hans-Jörg Selter, Lisbeth Thomsen

Project managers at the five hospitals: Anne Balle Larsen, Søren Brogaard, Marianne Frandsen, Mona Kyndi Pedersen, Kirsten Løth Lysdahl, Søren Schousboe Laursen, Maria Staun, Inge Ulriksen

Funding

CvP received a grant from the Danish Society for Patient Safety to write this article.

Legends to figures

Figure 1: Harms by Type

The dot plots show the relative distribution of harms by type. VTE = venous thromboembolism.

Figure 2: Patient Safety Incidents by Consequence

The dot plots show the relative distribution of patient safety incidents by consequence as reported to the Danish national database. Categories *minor* and *moderate* represent no and temporary harms, *major* permanent harms. Overall 96% of the incidents were temporary.

Figure 3: Rates of Harm

The run charts show monthly rates of harms measured with the Global Trigger Tool. The curve shows the harm rate expressed as the number of adverse events per 1000 patient days. The horizontal line is the median harm rate.

Figure 4: Harmed Patients

The run charts show the percentages of harmed patients measured with the Global Trigger Tool. The horizontal line is the median percentage of harms.

Figure 5: Harms by Severity

The dot plots show the relative distribution of severity of harms in categories *E* – *I*, where *E* and *F* are temporary, *G* – *H* permanent harms and *I* death. Overall, 96% of harms were temporary.

References

1 de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. Qual Saf Health Care 2008 Jun;**17**(3):216-223.

2 Berwick DM, Calkins DR, McCannon CJ, Hackbarth AD. The 100,000 lives campaign: setting a goal and a deadline for improving health care quality. JAMA 2006 Jan 18;**295**(3):324-327.

3 Benning A, Dixon-Woods M, Nwulu U, Ghaleb M, Dawson J, Barber N, et al. Multiple component patient safety intervention in English hospitals: controlled evaluation of second phase. BMJ 2011 Feb 3;**342**:d199.

4 Benning A, Ghaleb M, Suokas A, Dixon-Woods M, Dawson J, Barber N, et al. Large scale organisational intervention to improve patient safety in four UK hospitals: mixed method evaluation. BMJ 2011 Feb 3;**342**:d195.

5 McCannon CJ, Schall MW, Calkins DR, Nazem AG. Saving 100,000 lives in US hospitals. BMJ 2006 Jun 3;**332**(7553):1328-1330.

6 NHS Scotland. Scottish Patient Safety Programme. Available at: http://www.patientsafetyalliance.scot.nhs.uk/programme. Accessed 11/03, 2011.

7 Pronovost P, Needham D, Berenholtz S, Sinopoli D, Chu H, Cosgrove S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. N Engl J Med 2006 Dec 28;**355**(26):2725-2732.

8 Vincent C, Aylin P, Franklin BD, Holmes A, Iskander S, Jacklin A, et al. Is health care getting safer? BMJ 2008 Nov 13;**337**:a2426.

9 Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal trends in rates of patient harm resulting from medical care. N Engl J Med 2010 Nov 25;**363**(22):2124-2134.

10 Dansk Selskab for Patientsikkerhed. The Danish Safer Hospital Programme. 2011; Available at: http://www.sikkerpatient.dk/about-sikker-patient/the-danish-safer-hospital-programme.aspx. Accessed 11/03, 2011.

11 Sari AB, Sheldon TA, Cracknell A, Turnbull A. Sensitivity of routine system for reporting patient safety incidents in an NHS hospital: retrospective patient case note review. BMJ 2007 Jan 13;**334**(7584):79.

12 Thomas EJ, Lipsitz SR, Studdert DM, Brennan TA. The reliability of medical record review for estimating adverse event rates. Ann Intern Med 2002 Jun 4;**136**(11):812-816.

13 Buckley MS, Erstad BL, Kopp BJ, Theodorou AA, Priestley G. Direct observation approach for detecting medication errors and adverse drug events in a pediatric intensive care unit. Pediatr Crit Care Med 2007 Mar;**8**(2):145-152.

14 Naessens JM, Campbell CR, Huddleston JM, Berg BP, Lefante JJ, Williams AR, et al. A comparison of hospital adverse events identified by three widely used detection methods. Int J Qual Health Care 2009 Aug;**21**(4):301-307.

15 Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events. IHI Innovation Series White paper. 2009.

16 Classen DC., Lloyd RC., Provost L., Griffin FA., Resar R. Development and Evaluation of the Institute for Healthcare Improvement Global Trigger Tool. Journal of Patient Safety 2008;**4**(3):169-177.

17 Adler L, Denham CR, McKeever M, Purinton R, Guilloteau F, Moorhead JD, et al. Global Trigger Tool: Implementation Basics. J Patient Saf 2008;**4**:245-249.

18 Naessens JM, O'Byrne TJ, Johnson MG, Vansuch MB, McGlone CM, Huddleston JM. Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool. Int J Qual Health Care 2010 Aug;**22**(4):266-274.

19 Anonymous. Global Trigger Tool. En oversættelse af IHI's værktøj tilpasset danske forhold. Center for Kvalitet. Region Syddanmark 2008.

20 Anonymous. Strukturerat journalgranskning för att identifiera och mätta förekomst av skador i vården enligt metoden Global Trigger Tool. Institute for Health Care Improvement Innovation series 2007, Svensk översätning och anpassning. Stockholm. Kommentus Förlag 2008.

21 Rabøl LI, Jølving LR, Bjørn B and Viskum B. Systematisk journalgennemgang med IHI's Global Trigger Tool - Erfaringer fra afprøvning i danske sygehuse. Copenhagen. Dansk Selskap for Patientsikkerhed 2008.

22 Anonymous. Patientsikkerhed i kræftbehandlingen - journalgennemgang med Global Trigger Tool. Rapport nr 3/2010. Copenhagen. Kræftens Bekæmpelse 2010.

23 Auerbach AD, Landefeld CS, Shojania KG. The tension between needing to improve care and knowing how to do it. N Engl J Med 2007 Aug 9**;357**(6):608-613.

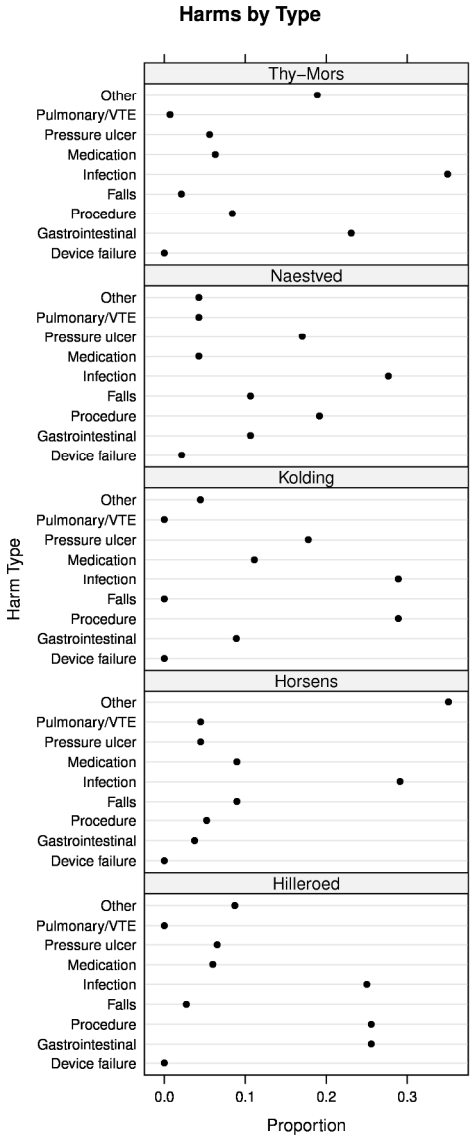
24 Perla RJ, Provost LP, Murray SK. The run chart: a simple analytical tool for learning from variation in healthcare processes. BMJ Qual Saf 2011 Jan**;20**(1):46-51.

25 Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al. 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured. Health Aff (Millwood) 2011 Apr**;30**(4):581-589.

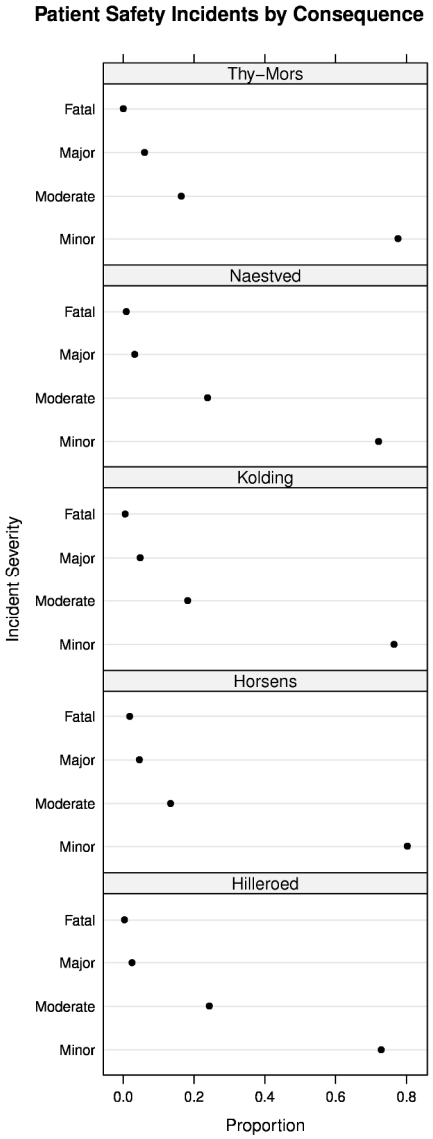
26 Griffin FA, Classen DC. Detection of adverse events in surgical patients using the Trigger Tool approach. Qual Saf Health Care 2008 Aug**;17**(4):253-258.

27 Sharek PJ, Horbar JD, Mason W, Bisarya H, Thurm CW, Suresh G, et al. Adverse events in the neonatal intensive care unit: development, testing, and findings of an NICU-focused trigger tool to identify harm in North American NICUs. Pediatrics 2006 Oct**;118**(4):1332-1340.

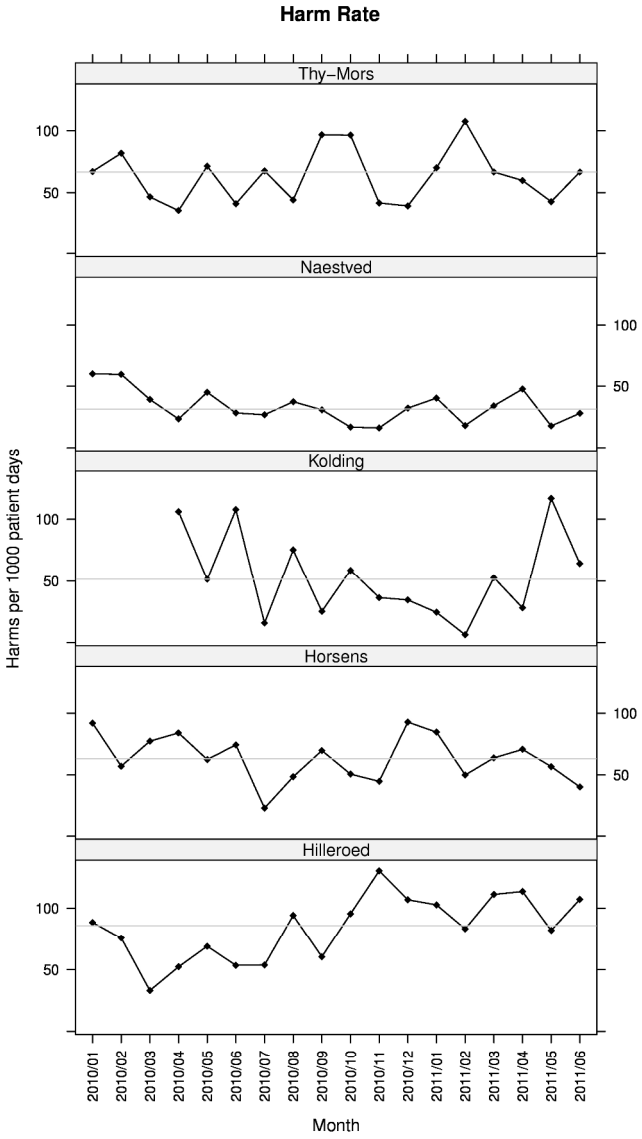
28 Resar RK, Rozich JD, Simmonds T, Haraden CR. A trigger tool to identify adverse events in the intensive care unit. Jt Comm J Qual Patient Saf 2006 Oct**;32**(10):585-590.



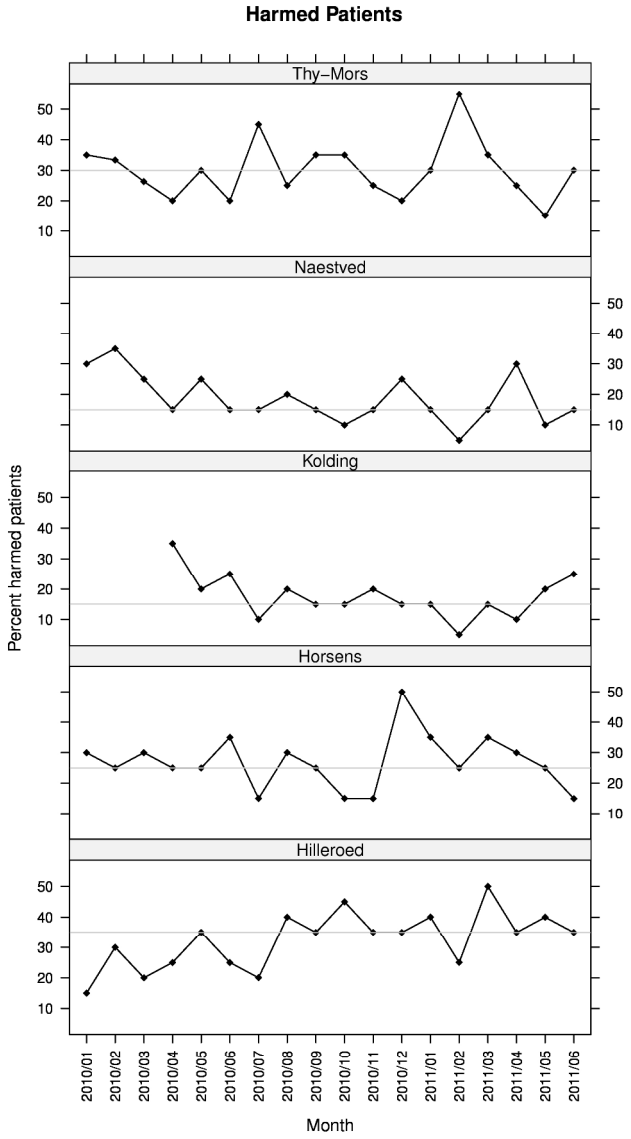
209x296mm (300 x 300 DPI)



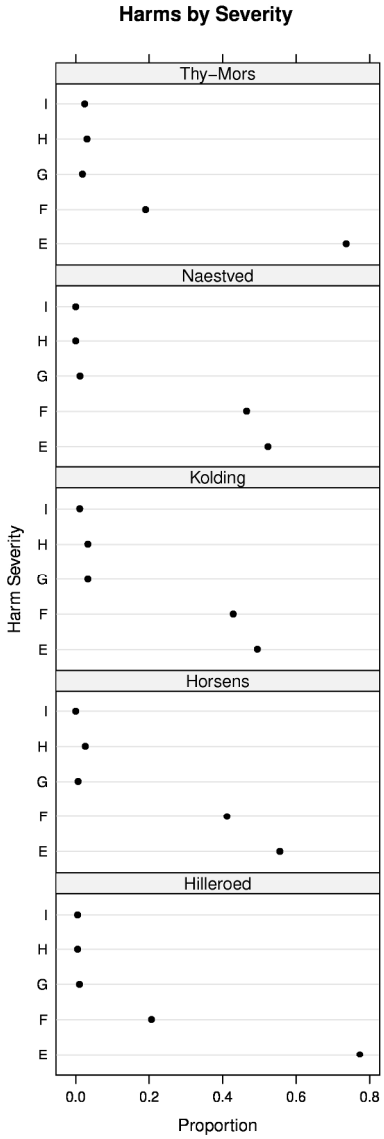
209x296mm (300 x 300 DPI)



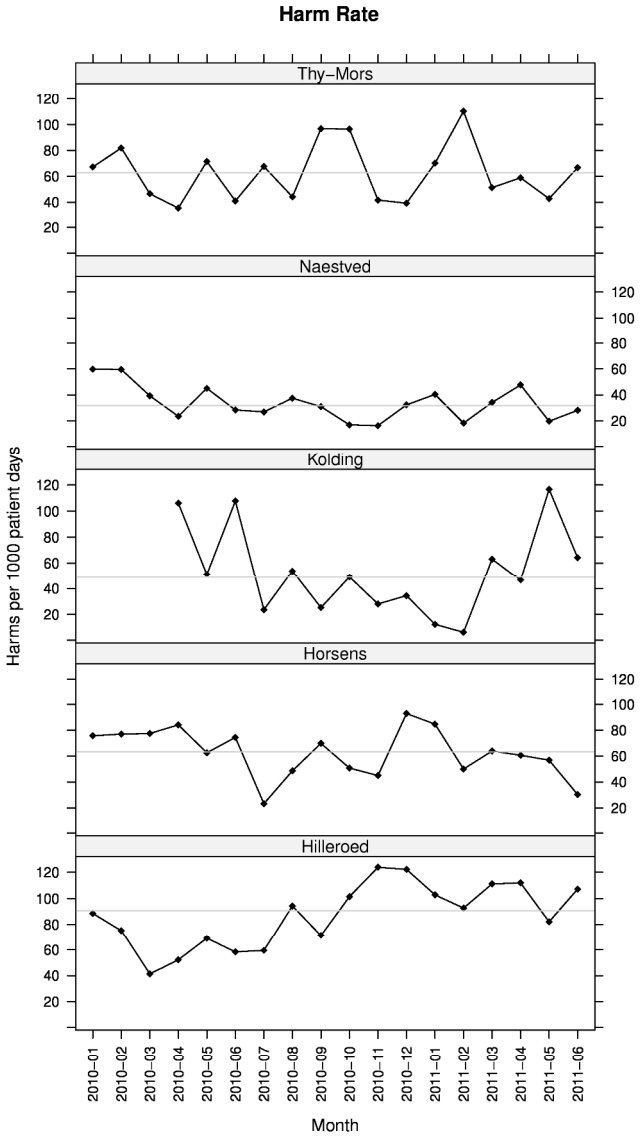
209x296mm (300 x 300 DPI)



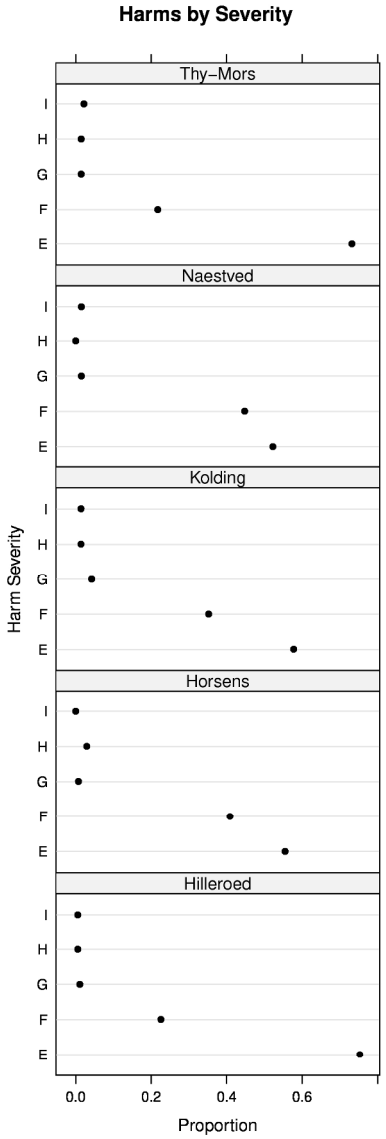
209x296mm (300 x 300 DPI)



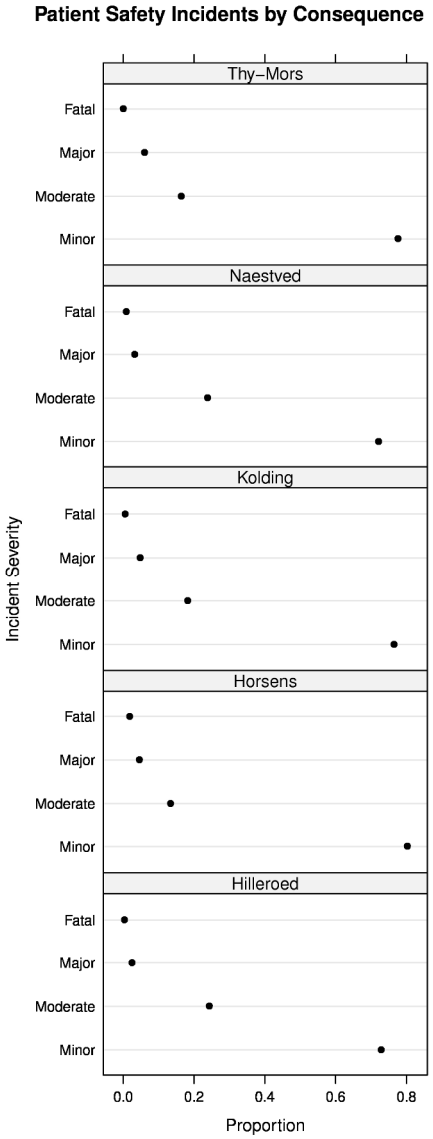
209x296mm (300 x 300 DPI)



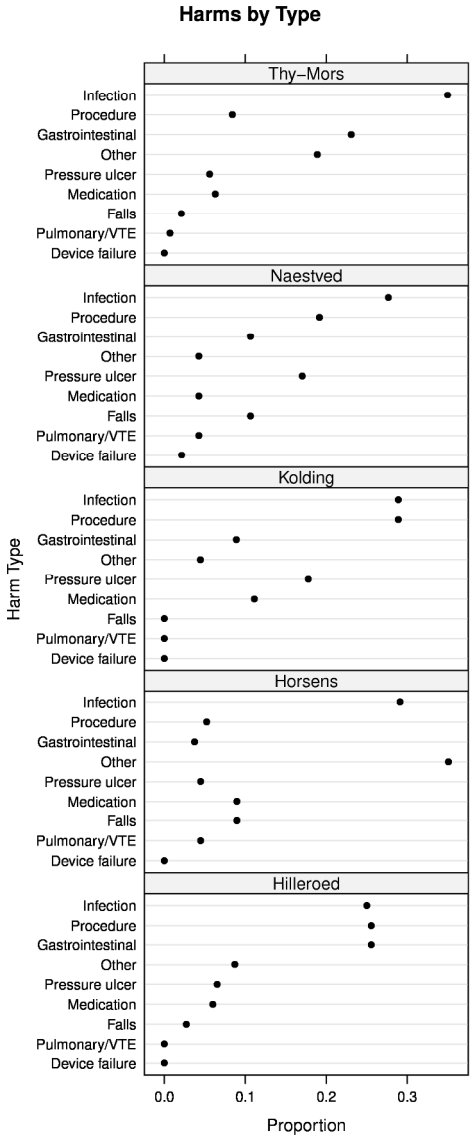
209x296mm (300 x 300 DPI)



209x296mm (300 x 300 DPI)



209x296mm (300 x 300 DPI)



209x296mm (300 x 300 DPI)