

BMJ Open Monitoring adverse events in Norwegian hospitals from 2010 to 2013

Ellen Tveter Deilkås,^{1,2} Geir Bukholm,^{3,4} Jonas Christoffer Lindstrøm,¹ Marion Haugen⁵

To cite: Deilkås ET, Bukholm G, Lindstrøm JC, *et al.* Monitoring adverse events in Norwegian hospitals from 2010 to 2013. *BMJ Open* 2015;5:e008576. doi:10.1136/bmjopen-2015-008576

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-008576>).

Received 24 April 2015
Revised 4 November 2015
Accepted 9 November 2015



CrossMark

¹Akershus University Hospital, Health Services Research Center, Lørenskog, Norway

²Division of Environmental Medicine and Infectious Disease Control, Norwegian Directorate of Health, National Patient Safety Program, Oslo, Norway

³Norwegian Institute of Public Health, Oslo, Norway

⁴Department of Chemistry, Biotechnology and Food Sciences, Norwegian University of Life Sciences, Ås, Norway

⁵Norwegian Computing Center, Oslo, Norway

Correspondence to
Dr Ellen Tveter Deilkås;
elde@ahus.no

ABSTRACT

Objectives: To describe how adverse event (AE) rates were monitored and estimated nationally across all Norwegian hospitals from 2010 to 2013, and how they developed during the monitoring period. Monitoring was based on medical record review with Global Trigger Tool (GTT).

Setting: All publicly and privately owned hospitals were mandated to review randomly selected medical records to monitor AE rates. The initiative was part of the Norwegian patient safety campaign, launched by the Norwegian Ministry of Health and Care Services. It started in January 2011 and lasted until December 2013. 2010 was the baseline for the review. One of the main aims of the campaign was to reduce patient harm.

Method: To standardise the medical record reviews in all hospitals, GTT was chosen as a standard method. GTT teams from all hospitals reviewed 40 851 medical records randomly selected from 2 249 957 discharges from 2010 to 2013. Data were plotted in time series for local measurement and national AE rates were estimated, plotted and monitored.

Results: AE rates were estimated and published nationally from 2010 to 2013. Estimated AE rates in severity categories E-I decreased significantly from 16.1% in 2011 to 13.0% in 2013 (−3.1% (95% CI −5.2% to −1.1%)).

Conclusions: Monitoring estimated AE rates emerges as a potential element in national systems for patient safety. Estimated AE rates in the category of least severity decreased significantly during the first 2 years of the monitoring.

INTRODUCTION

Based on several studies,^{1–5} the WHO has estimated that about 10% of patients are harmed in relation to hospital care in Western countries.^{6–7} This has created momentum for large-scale interventions in many countries.^{8–9} To facilitate monitoring of adverse events (AEs) within reasonable use of resources, the Institute for Healthcare Improvement (IHI) developed Global Trigger Tool (GTT) as a standard tool.^{9–10} The GTT definition of an AE is: “unintended

Strengths and limitations of this study

- The study covers, with few exceptions, all publicly and privately owned hospitals in Norway.
- The study is based on reviews of 40 851 hospital discharges.
- The Global Trigger Tool (GTT) reviewer teams applied the same standards and had the same training.
- The reviews were done by many teams and not only by one, which increases variability in results related to discretion.
- The size of reviewed samples was the same independent of hospital size.

physical injury resulting from or contributed to by medical care that requires additional monitoring, treatment or hospitalisation, or results in death”. Studies that used GTT in major US hospitals found AE rates of 27.7% and 33.2%.^{11–12} The GTT definition is at variance with the WHO definition, which may explain some of the discrepancy from the WHO estimate.

A research scan from the Health Foundation has considered GTT to compare well to other approaches, to be relatively sensitive and to identify significantly more AEs compared with self-reporting or other chart audit approaches. Most of the studies describing its use are based on large samples from multiple hospitals. Although the GTT is considered relevant for measuring harm at national level, few studies describe the process of doing so.^{13–14}

The Norwegian Ministry of Health and Care Services mandated all publicly and privately owned hospitals to review randomly selected medical records from 2010. GTT was chosen as a standard method in order to be able to aggregate results. The review started on medical records from March 2010 to make a baseline. Results have been published in annual reports since 2011 (for 2010–2013).¹⁵ Hospitals plotted data in diagrams on a web-based tool, facilitated by IHI, called Extranet, together with an additional

p-control chart tool. This has enabled them to follow and analyse their own results with statistical process control. The initiative from the health authorities was part of the planned national patient safety campaign, called 'In Safe Hands', which was launched in January 2011. The campaign aimed to reduce harmful events to patients, to establish competence and hospital routines for patient safety and to improve patient safety culture.

The aim of this study is to describe how the GTT method was applied at national level for monitoring AEs, and how the estimates developed during the monitoring period.

METHODS

Setting and inclusion criteria

All 19 publicly owned hospitals in Norway participated in the 2011 and 2012 reviews. Some of these hospitals consist of several hospital locations, organised together under the same board and Chief Executive Officer (CEO). Eighteen participated in 2010 and 2013. All five privately owned but publicly funded hospitals participated in 2010, 2012 and 2013, four in 2011. According to the GTT standard methodology, discharges from paediatric, psychiatric or rehabilitation units were excluded. Patients had to be admitted to the hospitals for more than 24 h, to be included.

TRANSLATION AND VALIDATION OF THE INSTRUMENT

The original GTT white paper was translated to Norwegian by a professional translator. The professional translation was validated by two clinical consultants (attending) who read it through and gave feedback. It was not back translated since it was considered a tool and not a questionnaire. The translated version was slightly modified by the campaign secretariat, regarding these three triggers on the trigger list:

- ▶ A medication commonly used against allergic reactions;
- ▶ The denomination used for blood glucose was changed;
- ▶ A blood test commonly used to monitor anticoagulation status.

Additional input for minor improvement of the translation came from record reviewers during the first months of the review in 2011. A protocol for how to perform the GTT review in the Norwegian campaign context was provided.¹⁶ It included a frequently asked questions (FAQ) list corresponding to the Norwegian context, in addition to the standard list.

Training and standardisation

GTT teams were trained according to the white paper standard and this is further described in the online supplementary appendix. The teams:

- ▶ Read the white paper and FAQ list before participating in the standardised 1-day course.

- ▶ Reviewed another 40 medical records before conducting regular reviews.
- ▶ Gathered at annual national meetings, to audit how the tool was applied.

Standardisation nationally was promoted through the following rules:

- ▶ Hospitals were required to establish at least one GTT team consisting of a doctor and two nurses.
- ▶ Each GTT team reviewed 240 medical records per year, strictly according to the IHI method.
- ▶ All eligible hospital discharges had to be included in the review.
- ▶ If there were more than one GTT team in a hospital, each team would make a sample of 240 discharges from different parts of the total population of eligible discharges that could not overlap.
- ▶ Identified AEs were categorised according to the severity categories E to I (description given in table 1), which are standard according to the GTT method, as well as type. GTT teams could not categorise AEs according to more than one severity.

Data reporting

The numbers of discharges with AEs, and AE severities and types per specified discharge were reported in an Excel template, which was annually sent to the campaign secretariat. In addition, the teams reported the total number of discharges that investigated records had been randomly selected from. This number was used for weighting the team results, when making national estimates. National estimated AE rates were presented in a

Table 1 Estimated mean difference in adverse event (AE) rates between 2013 and 2011 according to severity (with description of severity categories for the AE)

Severity category	Mean difference (95% CI)
E: AE contributed to temporary harm to the patient which required intervention	−2.6 (−3.4 to −0.4)*
F: AE contributed to temporary harm to the patient which required initial or prolonged hospitalisation	−0.9 (−1.7 to 1.0)
G: AE contributed to permanent patient harm	−0.2 (−0.6 to 0.2)
H: intervention was required to sustain life	−0.2 (−0.3 to −0.04)*
I: AE contributed to patient death	0.03 (−0.1 to 0.2)
E-I	−3.4 (−4.0 to −0.5)*
F-I	−1.3 (−2.0 to 0.6)

Mean difference in rates of AEs between 2013 and 2011, computed by the mean of the pairwise differences in each hospital. The lower and upper bounds of the 95% CI for the mean of the hospital pairwise differences are given below the estimate and are computed using 10 000 bootstrap simulations.

*Indicates a significant difference between the national AE rates in 2013 and 2011 at 5% significance level.

national report as per cent discharges with at least one AE in severity categories E to I compounded to E-I, and F to I compounded to F-I, together with a 95% CI.¹⁵ Discharges with multiple AEs in different severity categories were reported in each of the different categories. The compounding of the E-I and F-I categories took this into account by not counting these discharges more than once.

Ethics

Information that could be used to identify individual patients, for example, age, gender and date of discharge, were not reported, according to regulations for personal protection.

Statistical analysis

Cross-sectional analysis was performed on data annually reported in Excel templates from the GTT teams to calculate national AE rates with associated 95% CIs, divided into types and severities (table 1). Types of AEs were not estimated in this study, but are reported elsewhere.¹⁵ The national AE rates were calculated as a weighted average of individual means for the GTT teams from all hospitals. Since AEs according to type and severity were not symmetrically distributed across the GTT teams, non-parametric CIs for the national AE rates were calculated using 10 000 bootstrap simulations.¹⁷ The bootstrap was performed by sampling with replacement from the individual means for all GTT teams. An observation was drawn from the data with probability equal to the weight of the GTT team. For all analyses, $p < 0.05$ was regarded as statistically significant. Microsoft Excel V.2010 was used for collecting and managing the data, cross-sectional analysis and for producing the graph. Bootstrap CIs, based on 10 000 simulations, were calculated with the R statistical software.¹⁸

At local level, AE data were measured, registered and plotted in run charts twice a month in Extranet by the GTT teams. Data could also be exported to make p-control diagrams providing analysis over the development of AEs over time, both at local and national level. Local p-control diagrams are not shown here. The national p-control diagram is given in figure 1. The chart was made based on bimonthly data registered in Extranet and the national AE rate is calculated as a weighted average of individual means for all GTT teams. The number of GTT teams contributing to each data point in the chart varied between 32 and 36, meaning that not all of the GTT teams have registered data twice a month in Extranet (39–47 teams did the annual reviews, see table 2). Numbers of discharges reviewed at each point varied between 7 and 20 for each team, and between 320 and 381 at national level. The bimonthly data from Extranet was only used to develop the national p-control diagram; the rest of the analysis was based on data annually reported from the GTT teams in the Excel templates.

To explore the relation between size of hospitals and AE rates for the severity categories E-I, a mixed-effect model with the percentage of discharges with AEs as response was fitted. Year and the number of discharges (used as a proxy for hospital size, reflecting hospital's workload) were included as fixed effects, and hospital was included as a random effect. The data were weighted according to the proportion of medical records that were sampled from each hospital. A 95% bootstrap CI based on 10 000 simulations was used to make inferences of the parameter of interest, the effect of number of discharges. Model fit was assessed using the marginal R^2 for mixed-effect models.¹⁹

To explore how the national AE rates and severity develop over time, we looked at the differences between

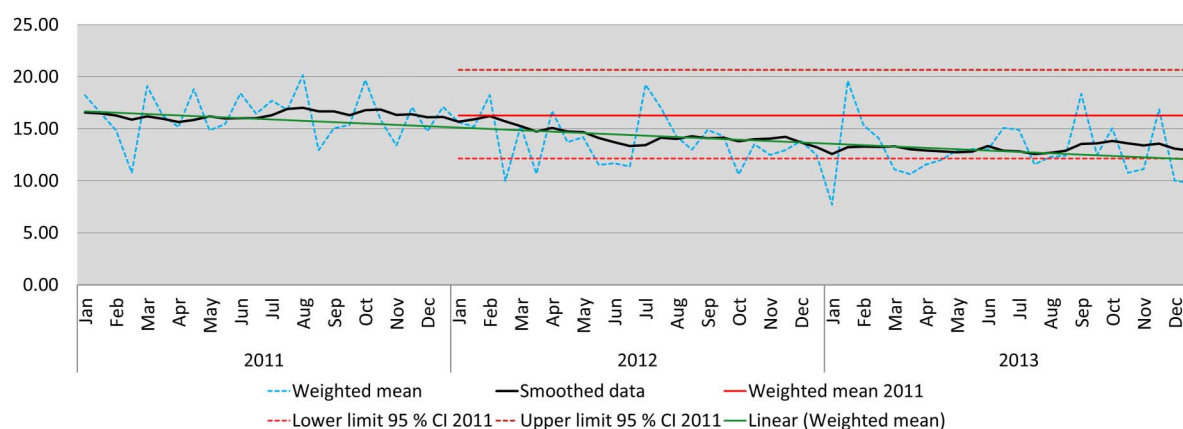


Figure 1 Estimated per cent discharges with at least one adverse event (AE) in severity categories E-I across all hospitals, 2011–2013. The light blue dashed line shows AE rates over time, calculated as the weighted average of individual means for all Global Trigger Tool (GTT) teams. The black smoothed line shows means of 10 consecutive dots. A red line marks the mean level in 2011 (set as baseline from 2012 and onwards since several teams did not have complete data in 2010) with 95% CIs marked with dashed red lines. The green line shows the linear trend. A regression of the AE rates against time shows a decline in AEs for the whole period ($p < 0.001$).

Table 2 Estimated discharges with at least one AE according to severity

	2010	2011	2012	2013
Per cent discharges with at least one AE in severity categories E-I (95% CI)	15.9 (13.8 to 17.8)	16.1 (14.6 to 17.5)	13.7 (12.5 to 15.1)	13.0 (11.7 to 14.2)
Per cent discharges with at least one AE in severity categories F-I (95% CI)	8.9 (7.3 to 10.5)	8.8 (7.9 to 9.8)	7.7 (6.8 to 8.6)	7.6 (6.6 to 8.7)
Number of reviewed discharges	7849	10 288	11 728	10 986
Number of discharges the records were randomly selected from	501 549*	585 648	593 046	569 714
Number of GTT teams	39	47	47	45

National AE rates from 2010 to 2013 with associated 95% CIs, divided into the severity categories E-I and F-I. From 2010 to 2013, 178 teams reviewed in total 40 851 medical records randomly selected from 2 249 957 discharges.

*The review in 2010 only included the months from March until December.

AE, adverse event; GTT, Global Trigger Tool.

the AE rates in 2013 and 2011. To control for the dependence between hospitals at the two time points, the pairwise differences in each hospital were used. Ninety-five per cent CIs for the mean of the hospital pairwise differences were calculated using 10 000 bootstrap simulations (table 1). The analysis of the individual severity categories were considered supplemental to the analysis of the E-I and F-I categories. This is elaborated in the discussion. Since the categories overlapped, usual multiple comparison procedures, such as the Bonferroni correction, were inappropriate.

RESULTS

In 2010, 85% of the country's eligible discharges from March to December were included in the hospital teams' sample frames. From 2011 and onwards, 98% or more of eligible discharges were included in the study.

Table 2 shows annual estimates and 95% CIs of discharges with at least one AE in severity categories E-I and F-I from 2010 to 2013. These are the parameters that were made public in annual reports.¹⁵ The table also gives annual numbers of reviewed discharges, numbers of discharges the records were randomly selected from and number of reviewing GTT teams.

The estimates of discharges with at least one AE in severity categories E-I show a significant decrease from 2011 to 2012 (−2.4% (−4.5% to −0.3%)) and from 2011 to 2013 (−3.1% (−5.2% to −1.1%)). The decrease from 2012 to 2013 is not statistically significant (−0.7% (−2.7% to 1.2%)). The estimates of discharges with at least one AE in severity categories F-I do not show any statistically significant decrease.

The model used to estimate the relation between size of hospitals and AE rates had an acceptable fit, with a marginal R^2 of 0.31. Estimated AE rates in severity categories E-I were reduced by 0.09 percentage points (95% CI −0.15 to −0.04) for each 1000 additional discharges, which we used as a proxy for hospital size.

As seen from table 1, estimated decline in national AE rates from 2013 to 2011 was significant for the severity categories E, H and E-I when looking at the pairwise differences in each hospital.

DISCUSSION

This study describes the variation in estimates of AEs in Norwegian hospitals during a 4 years period obtained as a result of national monitoring. To what extent could the results be biased by the way the review was organised? The reviewers' experiences from various specialties may have influenced their view on whether, for example, an AE caused intermediate or permanent harm. That could influence whether an AE was categorised as severity F or G. To reduce this potential for bias, we constructed the compounded severity category F-I. Although teams were trained in the same way (except one hospital that had already started), they could have varied in how carefully they examined the medical records. Teams that spend more time on reading nurse notes might find more AEs in the category E, than teams that read them more briefly. Such AEs could, for example, be mild decubitus ulcers that only demanded some cream put on, or falls that gave small bruises, only demanding adhesive plaster. This could have influenced hospital results. To reduce this potential for bias, we constructed the compounded severity category E-I, and counted discharges with multiple AEs only once. Team adjustment of interpretation and application of the GTT definitions and method over time could have influenced the AE estimates and contributed to the decline. Brian Bjørn in the Danish society for patient safety has communicated to the authors that in their experience AE rates of least severity, in the E category, can be exposed to variability in the teams application of the GTT definition over time. To counter we had annual national gatherings where Norwegian GTT teams audited cases and routines, to synchronise and solidify their practices. The meetings contributed to correction of occasional deviation in practice, but did not reveal widespread changes in how teams had applied the method and definitions. We can, however, not be certain about this without scientific verification. The validity of our AE estimates is to some extent confirmed by that they correlate with patients reported perceptions of patient safety surveyed in the same hospitals.²⁰

The results indicate a decline in AEs in the compounded category of E-I and in the category of least severity, the E category. One could easily conclude that the significant decrease in the E category fully explains the decrease in the E-I category. Non-significant decreases in the other severity categories as well, contradict that. They can also contribute to the observed effect in the E-I category. The E-I category is not just the sum of the individual subcategories. One discharge with two AEs in the E and G categories would only be counted once in the E-I category, but twice in the individual subcategories, one in the E category and one in the G category. The observed decline of AEs in the H category could mean that fewer patients experienced severe AEs requiring lifesaving intervention. We consider that the result should be interpreted with caution since the CI is near zero. When we explored the relation between number of discharges in the hospitals and AE rates, we found that hospitals with fewer discharges were more likely to have slightly higher estimates of AEs than those with many discharges. Causality between the decline in AEs and the interventions of the national patient safety campaign¹⁵ could not be inferred.

There was a public criticism in 2012 regarding why possible confounding effects of demographic characteristics, like, for example, gender and age, had not been analysed. Collection of demographic data had been prevented by regulations regarding personal protection. In addition, it had been considered that the risk for the sample to be skewed regarding demographic characteristics was minute since the random samples from each year in the study were very large and had been drawn from more or less all somatic hospital discharges in the country.

GTT may reveal risks that are not necessarily captured by diagnosis-based quality registers. This is because the GTT review is done regardless of the patients' somatic diagnosis. An additional value is that GTT supports plotting results in run charts and control charts for local monitoring, based on small data samples over time. Other opportunities lie in adapting the GTT method, so that its definitions of AEs are applied to more specific clinical contexts, and exploring how AE rates can be used for improvement purposes at hospital and clinical level.

The Norwegian patient safety campaign initiated medical record review in all hospitals to monitor AE rates. The idea was that the information could motivate hospital CEO's efforts to reduce risk for harming patients, and that the monitoring could inform the Norwegian Ministry of Health and Care Services in their governing efforts. Since the GTT tool has facilitated the Ministry to mandate 50% reduction in preventable AEs over the next 5 years, it can be said to have succeeded as a surveillance tool. But we have had little systematic information about how CEOs have used the information to monitor or improve patient safety at hospital level. It is reported that clinicians get frustrated by the data, because they mostly exist at hospital level, and therefore are not relevant to evaluate interventions at clinical

level. To further expand understanding on how structural conditions and organisational culture influence AE rates, AE estimates could be compared with surveys of workplace environment and patient safety culture. These are opportunities for future research.

CONCLUSIONS

The study describes a successful experience of monitoring national estimates of AEs based on medical record reviews in all hospitals. Estimated AE rates in the category of least severity decreased significantly during the first 2 years of the monitoring. Smaller hospitals had slightly higher AE rates than larger hospitals. Monitoring AE estimates emerges as a potential element in national systems for patient safety.

Acknowledgements The authors thank all GTT teams in the hospitals for their contributions to the study. They also thank the three referees for valuable comments, which have led to great improvements to the quality of the paper.

Contributors ETD was responsible for designing the study, training of reviewer teams, data collection, analysis and interpretation of data, and writing the manuscript. GB participated in designing the study and revising the manuscript critically. JCL participated in the statistical analysis and interpretation, in addition to revising the manuscript. MH participated in data collection, and was responsible for the statistical analysis and interpretation, in addition to revising the manuscript critically. All authors read and approved the final manuscript.

Funding Reviewer teams were trained by the National Patient Safety campaign, The Norwegian Directorate of Health. The teams work was funded by the hospitals. The research was funded by Akershus University Hospital's Health Services Research Unit and by Norwegian Computing Center.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

1. Wilson RM, Runciman WB, Gibberd RW, *et al.* The quality in Australian Health Care Study. *Med J Aust* 1995;163:458–71.
2. Brennan TA, Leape LL, Laird NM, *et al.* Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med* 1991;324:370–6.
3. Thomas EJ, Studdert DM, Burstin HR, *et al.* Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care* 2000;38:261–71.
4. Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. *BMJ* 2001;322:517–19.
5. Schioler T, Lipczak H, Pedersen BL, *et al.* Incidence of adverse events in hospitals. A retrospective study of medical records. *Ugeskr Laeger* 2001;163:5370–8.
6. Donaldson L. *World alliance for patient safety*. France: WHO, 2005.
7. World Health Organization, Fifty Fifth World Health Assembly. Quality of care: Patient safety – Report by the Secretariat. Geneva, 2002.
8. Botwinick L, Bisognano M, Haraden C. Leadership Guide to Patient Safety. IHI Innovation Series white paper. Cambridge, MA: Institute for Healthcare Improvement, 2006.
9. Griffin F, Resar R. *IHI Global Trigger Tool for measuring adverse events IHI innovation series*. 2nd edn. Cambridge, MA, 2009.

10. Resar RK, Rozich JD, Classen D. Methodology and rationale for the measurement of harm with trigger tools. *Qual Saf Health Care* 2003;12:ii39–45.
11. 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.
12. 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.
13. Health Foundation. *Evidence scan: Global Trigger Tools*. London, UK: The Health Foundation, 2010.
14. Sharek PJ, Parry G, Goldmann D, *et al.* Performance characteristics of a methodology to quantify adverse events over time in hospitalized patients. *Health Serv Res* 2011;46:654–78.
15. Deilkås ET. *Rapport for Nasjonal Journalundersøkelse med Global Trigger Tool 2013*. Oslo: Rapport fra Kunnskapssenteret, 2014.
16. The Norwegian patient safety campaign. *Gjennomføring av journalundersøkelse med Global Trigger Tool (GTT) i den norske pasientsikkerhetskampanjen*. Oslo: Nasjonal pasientsikkerhetskampanje "I trygge hender", 2013.
17. Efron B, Tibshirani RJ. *An introduction to the bootstrap*. London: Chapman & Hall, 1993.
18. R Core Team. R: A Language and Environment for Statistical Computing [computer program]. Vienna, Austria: R Foundation for Statistical Computing, 2015. <http://www.R-project.org/>
19. Nakagawa S, Schielzeth H. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol Evol* 2013;4:133–42.
20. Bjertnaes O, Deilkås ET, Skudal KE, *et al.* The association between patient-reported incidents in hospitals and estimated rates of patient harm. *Int J Qual Health Care* 2015;27: 26–30.

Appendix

Training and standardisation

Reviewer teams were trained according to the GTT standard described in the white paper. They read the white paper before going through a one day course which included reviewing five anonymised medical records. Patients had signed agreements to let us use them for training purposes. Further training included reviewing another 40 medical records before the teams could conduct regular reviews. The teams could call national facilitator and trainer for help. They were gathered at annual national meetings, to audit how the tool was applied. The purpose was to keep up a common understanding and practice according to the standard. In the meetings teams audited how the samples were randomised, how they did the review and how definitions were applied on specific examples. Results were exchanged, and teams described how they were used by executive leaders. Hospitals were required to establish at least one GTT team. All but one of the teams consisted of a doctor and two nurses. Some established reviewer teams at lower organisational levels than hospital level, like for example one team for the medical division and one team for the surgical division. If there were more than one GTT team in a hospital, each team would make a sample of 240 discharges from only a part of the total population of discharges. The populations of admissions which teams within the same hospital selected records from could not overlap. Each GTT team reviewed 240 medical records per year, strictly according to the IHI method. All hospital admissions relevant to the method had to be included in the review. Identified AEs were categorised according to severity (Table 1) as well as type. AEs could not be categorised according to more than one severity. There were 23 categories of types, for example, postoperative infection, bleeding and deep venous thrombosis. The categories of types were not mutually exclusive. If relevant, an AE could be categorised as more than one type, e.g. both a postoperative bleeding and a reoperation.