BMJ Open

Does the rate of adverse events identified by the Global Trigger Tool depend on the sample size? An observational study of retrospective record reviews of two different sample sizes

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2015-010700
Article Type:	Research
Date Submitted by the Author:	30-Nov-2015
Complete List of Authors:	Mevik, Kjersti; Nordland Hospital Trust, Regional patient safety resource center Hansen, Tonje; Nordland Hospital Trust, Medical director Deilkås, Ellen; Akershus University Hospital, Centre for Health Service Research Griffin, Frances; Fran Griffin and Associates, LLC Vonen, Barthold; The Artic University of Norway, Institute for community medicine; Nordland Hospital Trust, CMO
Primary Subject Heading:	Qualitative research
Secondary Subject Heading:	Health services research
Keywords:	QUALITATIVE RESEARCH, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Global Trigger Tool, Sample size, Adverse events < THERAPEUTICS

Does the rate of adverse events identified by the Global Trigger Tool depend on the sample size? An observational study of retrospective record reviews of two different sample sizes

Authors

Kjersti Mevik, Regional Patient Safety Resource Center, Nordland Hospital Trust, Bodø, Norway

Tonje E Hansen, Medical director, Nordland Hospital, Bodø, Norway

Ellen T Deilkås, Center for Health Service Research, Akershus University Hospital, Lørenskog, Norway

Frances A Griffin, Fran Griffin & Associates, LLC, Neptune, New Jersey, USA

Barthold Vonen, CMO, Nordland Hospital Trust, Bodø, Norway and Institute for community medicine, The Artic University of Norway, Tromsø, Norway

Corresponding author

Kjersti Mevik, Regional Patient Safety Resource Center, Nordland Hospital Trust, Post box 1480, N-8092 Bodø, Norway

Email: kjersti.mevik@nlsh.no

Phone number: +4797123977

Keywords: qualitative research, quality in health care, adverse events, global trigger tool, sample size

Word count: 3572

Abstract

Objectives: To investigate the impact of the sample size to the rate of adverse events by reviewing two different samples sizes of records (1680 and 240) from the same population by the Global Trigger Tool.

Design: Retrospective observational study.

Setting: A Norwegian 524-bed general hospital trust

Participants: 1920 medical records selected from January 1th to December 31th 2010.

Primary and secondary outcomes: Rate, type and severity of adverse events in two different samples sizes of records. Risk ratio of identifying adverse events in the large sample compared to the small sample.

Results: In the large sample 1.45 (95 % confidence interval: 1.07 to 1.97) times more adverse events per 1000 patient days (39.3 adverse events/1000 patient days) were identified than in the small sample (27.2 adverse events/1000 patient days). Hospital-acquired infections were the most common adverse events in both samples and the distributions of the other categories of adverse events did not differ significantly between the samples. The distribution of severity level of adverse events did not differ between the samples.

Conclusions: We identified a significantly higher rate of adverse events in the large sample compared to the small sample thus demonstrating that the rate of adverse

events may depend on the sample size. The recommended sample size is sufficient to reveal the distribution of categories and the severity of adverse events, although further studies are needed to determine if larger samples than the recommended sample size are necessary to detect a more accurate rate of adverse events.

Article summary:

Strength and limitations of this study:

- The samples were similar in terms of age, sex and length of stay.
- The large sample is seven times larger than the recommend sample size.
- Preventability of the adverse events was not assessed.
- Only one sample size was compared to the recommended sample size.
- Records in the small sample were reviewed independently by two primary reviewers while records in the large sample were each reviewed by one of three primary reviewers.

This work was supported by The Northern Norwegian Regional Health Authority.

Competing interest: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: KM had financial support from The Northern Norway Regional Health Authority as a PhD grant for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

INTRODUCTION

For more than a decade considerable efforts have been invested across healthcare to reduce adverse events, resulting in many efforts to identify reliable and valid tools to measure such events. The Institute for Healthcare Improvement´s (IHI) Global Trigger Tool is a widely used and considered an effective tool for measuring adverse events[1–3]. The method includes reviewing samples of ten patient records selected randomly bi-weekly from the hospital discharge lists. The primary reviewers search for pre-defined triggers that could indicate possible adverse events. The adverse events identified in the bi-weekly periods provide the data for Statistical Process Control (SPC) charts used to analyse adverse events rates over time. However, concerns have been raised[2,4–8], about the method's ability to detect accurate rates of adverse events and changes in rates accurately, due to the small recommended sample size.

In Norway all hospitals are required by the National Health Authority to use the Global Trigger Tool to review the minimum of ten records selected continuously and bi-weekly in order to monitor the rates of adverse events at each hospital and at a national level. We wanted to assess whether a larger sample size than the recommended sample of ten records bi-weekly could yield a different rate of adverse events per patient days. The recommended sample size for the Global Trigger Tool has not been validated to our knowledge thus demonstrating the need for this study.

Our aim was to obtain the rate, categories and severity of adverse events in two different sample sizes of records selected from the same population: one sample corresponding to seven times larger than the recommended sample size and one sample corresponding to the recommended sample size. We hypothesised that increasing the sample size would not yield a different rate of adverse events per 1000 patient days.

METHODS

Study design

The study is an observational cross-sectional study including retrospective record review of two samples of records, respectively 1680 and 240 (figure 1).

Setting

The study was performed in a 524-bed hospital trust at three geographical locations in Nordland County, North-Norway. Both samples were selected from the same population discharged from January 1th to December 31th 2010. However the large sample was first stratified according to discharges from the nine services in the trust and then ten records were selected from five services and five records from four services respectively bi-weekly to a total of 70 records. The small sample included ten records selected bi-weekly from the aggregated discharge lists of all the nine services. Following the IHI guidelines, records were excluded for patients aged 17 years or younger, patients admitted primarily for psychiatric or rehabilitation care, or patients with a length of stay less than 24 hours. The whole hospitalization was reviewed including patient days at all services not only at the index service.

The study was approved by the Data protection official in Nordland Hospital trust and by the Norwegian Regional Ethics Committee (ref 2012/1691).

Record review method

Training of the reviewers followed the IHI recommendations and included theory, practical review exercises, and debriefing sessions provided by experienced reviewers. The IHI definition of an adverse event was used, i.e.,[1]: *"Unintended physical injury resulting from or contributed to by medical care that requires additional monitoring, treatment or hospitalization, or that results in death".* Both adverse events associated with treatment given prior, during or after (within 30 days) to the index discharge (the discharge selected from the discharge lists of the services) were included to evaluate the total number of adverse events resulting from

medical care. Preventability of the identified adverse events was not evaluated. The primary reviewers reviewed the records for a maximum of 20 minutes searching for triggers and possible adverse events.

The identified adverse events were grouped into 23 categories derived from the Norwegian translation[9] of the IHI's Global Trigger Tool. These categories were further aggregated into eight main categories (i.e., hospital acquired infections, surgical complications, bleeding/thrombosis, patient fall/fracture, medication harm, obstetric harm, pressure ulcer and other). The severity of adverse events was categorized into five levels (E – I) using definitions adapted from those of the National Coordinating Council for Medication Error Reporting and Prevention index (NCC MERP)[10]:

Category E:	Temporary harm to the patient and required intervention
Category F:	Temporary harm to the patient and required initial or prolonged hospitalization
Category G:	Permanent patient harm
Category H:	Intervention required to sustain life
Category I:	Patient death

The review process in the small sample followed the IHI guidelines[1] where two (nurses) primary reviewers (reviewer A and reviewer B) each reviewed all records independently and then reached consensus on presence, category and severity of events; this was then authenticated by a physician (reviewer C). Reviewing of records from the large sample was slightly modified where three reviewers (reviewer A, reviewer C and reviewer D (physician)) reviewed different records independently as primary reviewers. Each record was only reviewed by one reviewer. Reviewer A reviewed 65 % of the records, reviewer C reviewed 29 % and reviewer D reviewed 6 % of the records. After the primary review a consensus among the reviewers was reached on the presence, category and severity of adverse events identified (figure

1). No further authentication of the adverse events was performed for records in the large sample. The modification with only one primary reviewer per record in the reviewing process in the large sample was done due to limited resources available.

Statistical analysis

Demographic variables of the records were obtained. Categorical variables were compared between the samples with Chi-square test while continuous variables were compared using the Independent t-test.

Statistical Process Control (SPC) charts are used to evaluate variations between data points over time which is a recommended approach for evaluating the rates of adverse events measured by the Global Trigger Tool[1,11]. We used QI Macros in Excel 2013 to present the calculated rate of adverse events per 1000 patient days in U-charts and the calculated percentage of records with adverse events in a P-chart of both samples[12]. Test 1-3 of special cause variation (SCV) were applied in order to evaluate the rates. The tests are positive if data points are outside the control limits, eight or more data points are one the same side of the median or/and if six data points are either ascending or descending. We hypothesised that different rates of adverse events in the two samples would yield different results in terms of the tests and control limits.

To compare the calculated rates, proportions of severities and categories of adverse events between the samples we used Poisson regression in generalized linear models. Poisson regression was chosen as it accounts for variations in the number of cases reviewed and variations in length of stay. The number of adverse events was set as the dependent variable and log patient days as the offset variable (in the analysis of adverse events per patient day). When analysing adverse events per records and percentages of records with an adverse event, zero was set as the fixed value. A p value of < 0.05 was defined as statistically significant. We also adjusted for services and variables associated with the index service. Associations between

adverse events and demographic variables were explored using Pearson´s correlation and logistic regression. We used SPSS (version 22.0; SPSS Chicago, IL) for statistical analyses.

RESULTS

Demographics characteristics

A total of 1920 records were reviewed in the study using the Global Trigger Tool. Demographic characteristics in both samples and the overall population from which the samples were drawn from are shown in table 1. 12 % of the overall population (14267 discharges) was reviewed in the large samples while 2 % was reviewed in the small sample. Length of stay, age and sex were derived for the whole hospitalization and these did not differ between the large and the small sample. Patients in the large sample were different to the overall population in terms of sex while patients in the small sample did not differ from the overall population. Type of admissions (acute or planned), case mix (discharges diagnose), services (functional units), case mix index, admission to surgery and numbers of transfers were derived from the index discharge (source of the random selection) and adjusted for.

Table 1: Demographic characteristics of the two samples and the overall population and by presence or absence of an adverse event

	Samples			p-value			
	Large sample	Small sample	Overall population	1 vs 2	1 vs 3	2 vs 3	
n=	**1680**	**240**	**14267**				
Length of stay (days)*	6.8 (7.5)	6.9 (11.1)	6.3 (6.9)	0.852	0.014	0.400	±
Average age (years)*	62 (21)	61 (21)	62 (21)	0.487	0.592	0.344	±
Sex (percent women)**	62	59	57	0.446	<0.001	0.410	§

*Values presented as mean with standard deviations. **Values presented as percent. n.s =non-significant = p value>0.05.

±T-test, §Chi-square test

Comparison of adverse events

In the large sample of 1680 records comprising 11367 patient days, we identified 447 adverse events in 347 discharges. This corresponds to a rate of 39.3 adverse events per 1000 patient days (95 % confidence interval (CI): 35.8 to 43.1, standard error (SE) = 1.86) or 26.6 adverse events per 100 discharges (95 % CI: 24.3 to 29.2, SE= 1.26). The percentage of patients with an adverse event was 20.5 % in the large sample. In the small sample of 240 records comprising 1657 patient days, we identified 45 adverse events in 30 discharges. This corresponds to a rate of 27.2 adverse events per 1000 patient days (95 % CI: 20.3 to 36.4, SE = 4.05) or 18.8 adverse events per 100 discharges (95 % CI: 14.0 to 25.1, SE= 2.80). The percentages of patients experiencing an adverse event was 12.5 %. When reporting percentages of patients experiencing an adverse event one has to account for some patients experienced more than one adverse event. Patients experiencing adverse events were older, had longer hospital stays and were more frequently discharged with case mix of injury/poisoning and diseases of the circulatory system than patients without experiencing adverse events.

The rate of adverse events per 1000 patient days was 45 % higher in the large sample than in the small sample (risk ratio (RR) =1.45, 95 % CI: 1.07 to 1.97; P= 0.02). Likewise the rate of adverse events per record in the large sample was 42% higher in the large sample than in the small sample (RR= 1.42, 95 % CI: 1.04 to 1.93, P= 0.03). The percentages of records including an adverse event was 65 % higher in the large sample than in the small sample (RR=1.65, 95 % CI: 1.14 to 2.34, P=0.008). In figure 2 the rates of adverse events per 1000 patient days in both samples are presented in control U-charts and percentages of records with adverse events in control P-charts over the 24 bi-weekly periods in 2010. In both charts the control limits are much wider in the small sample than in the large sample. Special cause variations (positivity of tests 1) were identified only for the small sample. This is marked with a black dot in the U-chart. None of the other tests were positive for either of the samples.

To adjust for the stratification made before selection of records to the large sample we adjusted for the variables that were associated from the index discharge. The primary results did not alter as the risk ratio was 1.83 (95 % CI: 1.32 to 2.54, P<0.001) of identifying an adverse event per 1000 patient days in the large sample compared to the small sample when adjusting for these variables.

There were no difference of identified adverse events among the reviewers of the large sample in terms of adverse events per 1000 patient days, adverse events per records or percentages of patients with an adverse events.

Length of stay correlated moderately with number of adverse events detected in records in the large sample (r^2=0.21, P<0.001) while in records in the small sample length of stay correlated closely to number of adverse events (r^2=0.46, P<0.001). Age and number of adverse events correlated fairly in the large sample (r^2=0.03, P<0.001) while age correlated negatively with number of adverse events in the small sample (r^2= -0.003, P= 0.54).

Hospital acquired infections were the most frequent category of identified adverse events in both samples. There were no significant differences between the estimated proportions of identified adverse events between the samples for the six main categories of adverse events; hospital acquired infections (RR=1.52, 95 % CI: 0.94 to 2.47, P=0.09), surgical complications (RR=1.28, 95 % CI: 0.67 to 2.47, P=0.46), bleeding/thrombosis (RR=1.44, 95 % CI: 0.70 to 2.98, P=0.33), medication harm (RR=1.68, 95 % CI: 0.60 to 4.66, P=0.32), patient fall (RR=0.83, 95 % CI: 0.24 to 2.82, P=0.76) and pressure ulcers (RR=0.73, 95 % CI: 0.16 to 3.33, P=0.68) (supplementary file 1). For the categories obstetric harm and other, no adverse events were identified in the small sample and a comparison was not performed.

The least severe adverse events (category E) accounted for more than half of the adverse events identified in both samples (figure 3). No significant differences were

found between the rate of adverse events per 1000 patient days between the samples, when adverse events were analysed separately according to severity of the adverse events: E (RR=1.50, 95 %CI: 1.00 to 2.26, P=0.05) and F (RR=1.68, 95 % CI: 0.99 to 2.85, P=0.05) and F, G, H and I (RR=0.47, 95 %CI: 0.17 to 1.27, P=0.14) and G, H and I (RR=1.38, 95 % CI: 0.87 to 2.18, P=0.17).

DISCUSSION

We found that the rate of adverse events measured with the Global Trigger Tool for a one year time period was significantly higher when reviewing a larger sample size of records than the recommend sample size. The rate of adverse events was 1.45 higher in the large sample than in the small sample. When comparing adverse events per 1000 patient days the 95 % CI did fairly overlap and the SE was lower in the large sample (SE=1.86) compared to the small sample (SE=4.05). Our findings indicate that the sample size may influence the rate of identified adverse events. The differences regarding CI and SE indicates that increasing the sample size increase the reliability and validity of the results.

While evaluations of the Global Trigger Tool have reported both high sensitivity[3] and acceptable reliability[13,14] the impact of the sample size in determining the level of adverse events has hardly been discussed. We believe this is the first attempt to assess the impact of the sample size to the rate of adverse events identified with the Global Trigger Tool. Kennerly et al adjusted the sample size to the hospital sizes[15] without further comparisons between different sample sizes selected in the same time period. In the study of Landrigan et al[6] Global Trigger Tool was used to evaluate changes in adverse event rates over time. Their study did not apply SPC charts but used Poisson regression to conclude that change in the rate of adverse events over time had not occurred. However the samples was only ten records per quarter per hospital and the issue if the sample size was adequate to detect changes over time was not discussed. Classen et al identified 82 adverse events per 1000 patient days when reviewing a sample of 795 records selected from three hospitals and a period of one month[3] which is higher than our estimates of 39

adverse events per 1000 patient days in samples of 140 for one month. In regard to this we therefore determined it legitimate, necessary and original to assess whether using the Global Trigger Tool with different sample sizes would produce different results.

While our findings may challenge the sensitivity of the recommended small sample size in order to identify an accurate rate of adverse events, they also underline the ability of that sample size to reflect distribution of severities and categories of adverse events accurately. Our results in terms of this corresponded well with other studies[15,16]. In our small sample none adverse events of category I were identified, which is in accordance to that the Global Trigger Tool is not designed to measure this cases (category I) over time. Due to infrequent occurrence other methods should be used to monitor these specific types of events, for example investigating all hospital deaths[17,18]. Because of this we compared the rate of adverse events in category I along with the rate of adverse events in other categories (category F, G and H).

Several factors could explain the differences in the rate of adverse events identified in the two samples. First, the Simpson paradox defined as statistical results from aggregated data, could give a different result to that of a group-level analysis do[19]. A skewness regarding the variables associated to the index discharges could be present in our study as the large sample was stratified according to the services before sampling and the small sample was not. However, the primary results did not differ when adjusting for these variables. Neither did the demographic characteristic as sex, age and length of stay differ between the large and the small sample. Second, the study was undertaken for only one year's discharges comprising 240 records in the small sample. To increase the power of the study records from a longer time period could be reviewed as we did not assess whether a greater number of bi-weekly periods with a larger sample size would result in greater sensitivity. A meta-analysis of sample sizes showed that the variation of adverse event rates decreases as the sample size increases[4] thus underlining the importance of having a large enough sample size in order to obtain valid results. Third, the records in the

large sample were reviewed by three different primary reviewers where two of them were physicians. However each record in the large sample was reviewed only by one of the primary reviewers while each record in the small sample were reviewed by two primary reviewers. As there were no significantly differences in the rate of adverse events identified by the reviewers in the large sample we do not assume that this could explain the different rates identified in the samples. If number of primary reviewers should correlate to number of adverse events identified one should expect a higher rate of adverse events in the small sample which was not the case in our study.

CONCLUSION

We believe the findings in this study challenges the appropriateness of the sampling methods commonly used as the rate of adverse events increased when the number of records reviewed bi-weekly was increased, though limitations of the study have to be accounted for. Further studies are needed for this but number of discharges could be a guide to select an appropriate sample size.

Contributors KM and BV designed the study. TEH and KM reviewed the records. KM conducted the data analysis and wrote the first draft of the manuscript and revised drafts of the manuscript after all co-authors had reviewed. KM performed the statistical analyses and produced the graphs.

Competing interest None

Funding KM received a grant from the Northern Norway Regional Health Authority

.

Ethics approval Data protection official in Nordland Hospital trust. Norwegian Regional Ethics Committee (ref 2012/1691).

Acknowledgements The authors would like to thank Birger Hveding and Inger Lise Øvre who conducted the reviews together with the co-authors TEH and KM. Frank Federico and Carol Haraden for guidance on the development of the study protocol. Tom Wilsgaard for advice on statistical analysis and Alexander Ringdal and Marina Mineeva for help with data processing.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1 Griffin F, Resar R. IHI Global Trigger Tool for measuring adverse events. *IHI Innov Ser white Pap* 2007;:1–44.http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:IHI+Global+Trigger+Tool+for+Measuring+Adverse+Events#0 (accessed 26 Nov2014).

2 Commission S, Zealand N. The Global Trigger Tool : A Review of the Evidence. Wellington: 2013. www.hqsc.govt.nz

3 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. doi:10.1377/hlthaff.2011.0190

4 Lessing C, Schmitz A, Albers B, *et al.* Impact of sample size on variation of adverse events and preventable adverse events: systematic review on epidemiology and contributing factors. *Qual Saf Health Care* 2010;**19**:e24. doi:10.1136/qshc.2008.031435

5 Mattsson TO, Knudsen JL, Lauritsen J, *et al.* Assessment of the global trigger tool to measure, monitor and evaluate patient safety in cancer patients: reliability concerns are raised. *BMJ Qual Saf* 2013;**22**:571–9. doi:10.1136/bmjqs-2012-001219

6 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. doi:10.1056/NEJMsa1004404

7 James J. A new, evidence-based estimate of patient harms associated with hospital care. *J Patient Saf* Published Online First: 2013.http://journals.lww.com/journalpatientsafety/Abstract/2013/09000/A_New,_Evidence_based_Estimate_of_Patient_Harms.2.aspx (accessed 8 Jan2015).

8 Shojania KG, Thomas EJ. Trends in adverse events over time: why are we not improving? *BMJ Qual Saf* 2013;**22**:273–7. doi:10.1136/bmjqs-2013-001935

9 Strukturert journalundersøkelse, ved bruk av Global Trigger Tool for å identifisere og måle forekomst av skader i helsetjenesten. Oslo: 2010.

10 Hartwig SC, Denger SD, Schneider PJ. Severity-indexed, incident report-based medication error-reporting program. *Am J Hosp Pharm* 1991;**48**:2611–6.http://www.nccmerp.org/types-medication-errors

11 Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. *Qual Saf Health Care* 2003;**12**:458–64.http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1758030&tool=pmcentrez&rendertype=abstract

12 Provost LP, Murray S. *The Health Care Data Guide: Learning from Data for Improvement*. 2011. http://books.google.com/books?hl=no&lr=&id=pRLcaOkswQsC&pgis=1 (accessed 7 Jan2015).

13 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. doi:10.1111/j.1475-6773.2010.01156.x

14 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. doi:10.1093/intqhc/mzq026

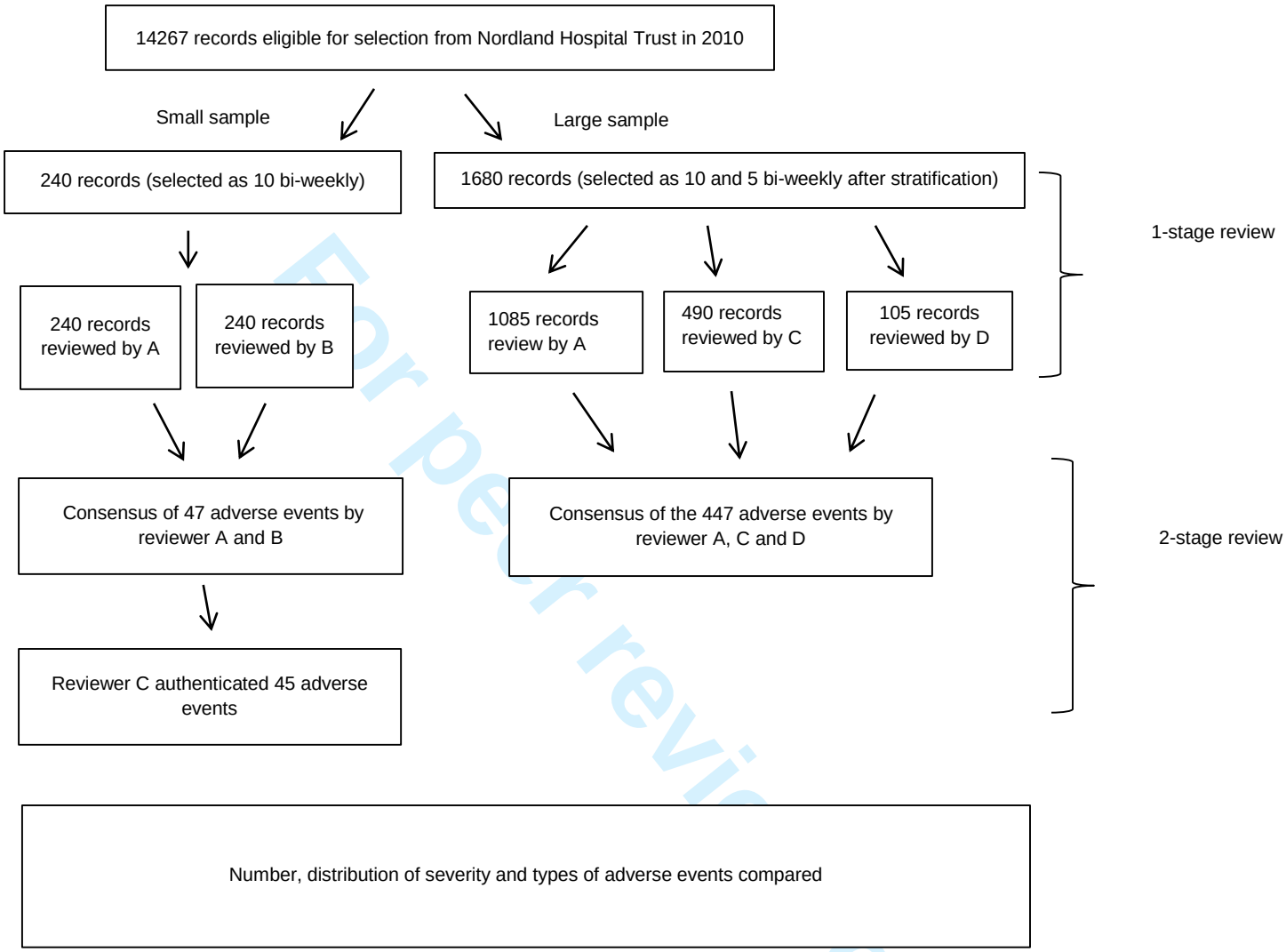
15 Kennerly D a, Saldaña M, Kudyakov R, *et al.* Description and evaluation of adaptations to the global trigger tool to enhance value to adverse event reduction efforts. *J Patient Saf* 2013;**9**:87–95. doi:10.1097/PTS.0b013e31827cdc3b

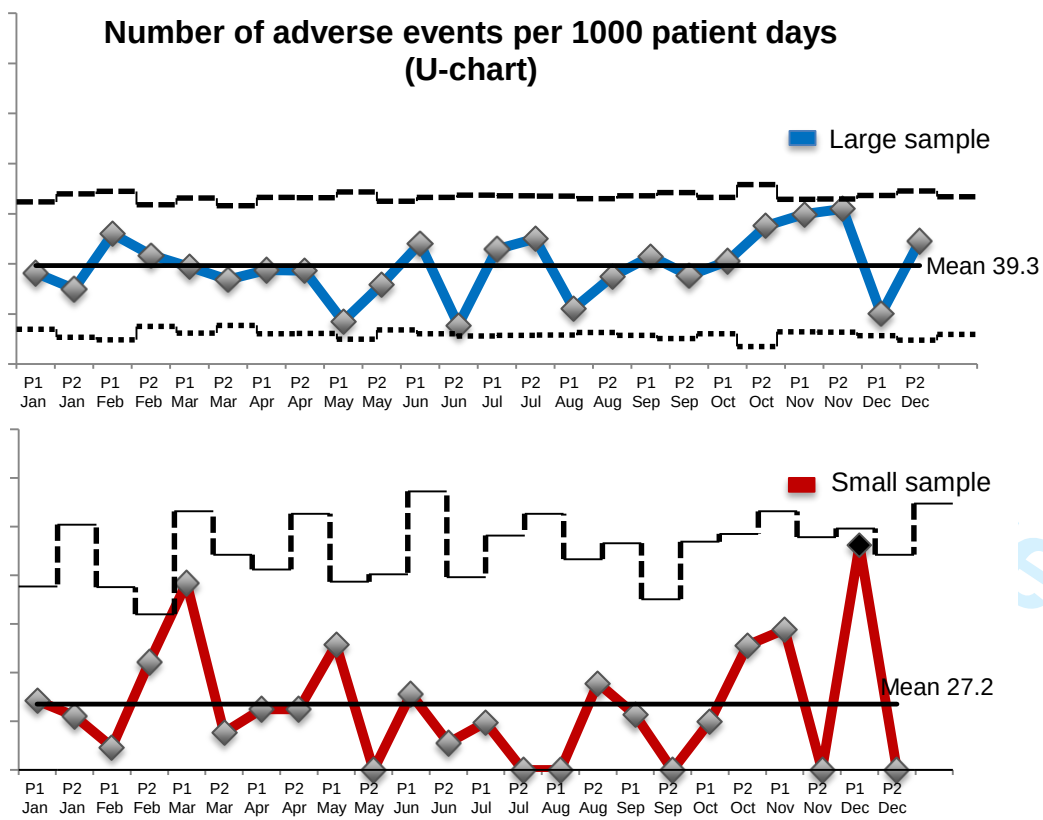
16 Rutberg H, Borgstedt Risberg M, Sjödahl R, *et al.* Characterisations of adverse events detected in a university hospital: a 4-year study using the Global Trigger Tool method. *BMJ Open* 2014;**4**:e004879. doi:10.1136/bmjopen-2014-004879

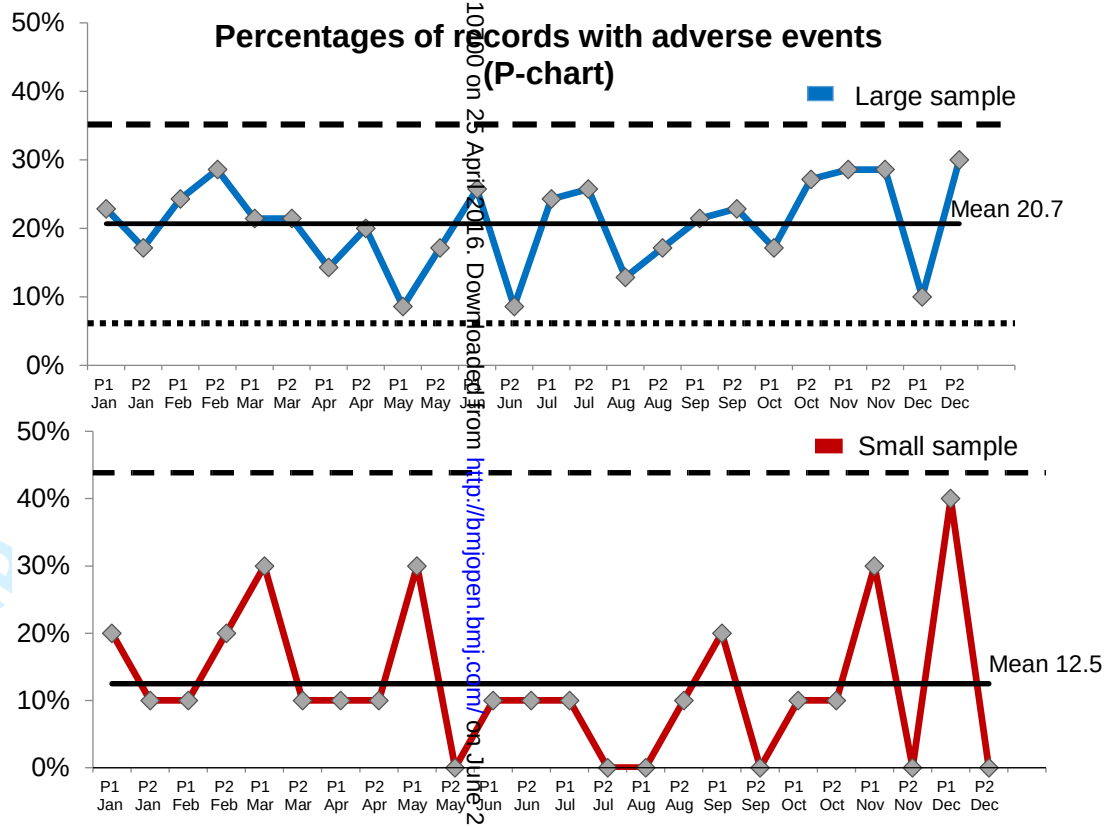
17 Lau H, Litman KC. Saving lives by studying deaths: Using standardized mortality reviews to improve inpatient safety. Jt. Comm. J. Qual. Patient Saf. 2011;**37**:400–8.

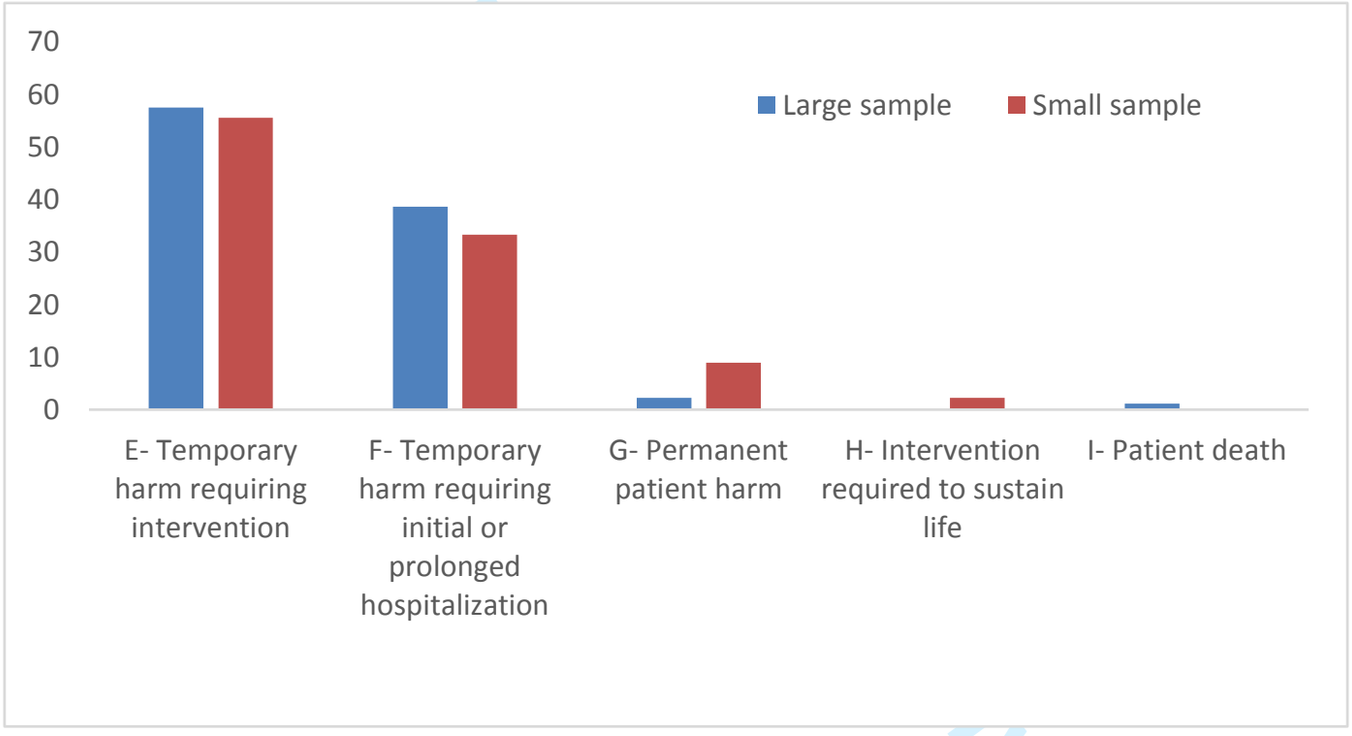
18 Move your dot: Measuring, Evaluating, and Reducing Hospital Mortality Rates (part 1) IHI Innovation Series white paper. Boston: 2003. http://scholar.google.no/scholar?q=Move+Your+Dot%E2%84%A2%3A&btnG=&hl=no&as_sdt=0%2C5#1 (accessed 5 Jun2015).

19 Blyth CR. On Simpson's Paradox and the Sure-Thing Principle. *J Am Stat Assoc* 1972;**67**:364–6. doi:10.1080/01621459.1972.10482387









Supplementary file 1: Percentages of types and severity level of the adverse events identified in the large and small sample.

	Large sample size						Small sample size						
	E	F	G	H	I	Total	E	F	G	H	I	Total	Risk ratio (95 %CI)
Hospital acquired infections	26 %	15 %	0 %	0 %	1 %	42 %	31 %	9 %	0 %	0 %	0 %	40 %	1.50 (0.94-2.47)
Urinary tract infection	14 %	2 %	0 %	0 %	0 %	16 %	11 %	0 %	0 %	0 %	0 %	11 %	
CVC infection	0 %	0 %	0 %	0 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	2 %	
Ventilator associated pneumonia	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	
Other infection	8 %	12 %	0 %	0 %	1 %	21 %	9 %	2 %	0 %	0 %	0 %	11 %	
Lower respiratory infection	4 %	1 %	0 %	0 %	0 %	5 %	11 %	4 %	0 %	0 %	0 %	16 %	
Surgical complications	6 %	12 %	1 %	0 %	0 %	20 %	7 %	7 %	7 %	2 %	0 %	22 %	1.28 (0.67-2.47)
Infection after surgery	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	2 %	0 %	2 %	
Respiratory complications after surgery	1 %	1 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	0 %	0 %	0 %	
Return to surgery	0 %	3 %	0 %	0 %	0 %	4 %	0 %	4 %	0 %	0 %	0 %	4 %	
Injury, repair or removal of organ	1 %	1 %	0 %	0 %	0 %	1 %	4 %	0 %	7 %	0 %	0 %	11 %	
Occurrence of any operative complication	2 %	2 %	0 %	0 %	0 %	4 %	2 %	2 %	0 %	0 %	0 %	4 %	
Switch in surgery	0 %	0 %	0 %	0 %	0 %	1 %	0 %	0 %	0 %	0 %	0 %	0 %	
Other	3 %	5 %	0 %	0 %	0 %	8 %	0 %	0 %	0 %	0 %	0 %	0 %	
Bleeding/thrombosis	14 %	3 %	0 %	0 %	0 %	18 %	16 %	2 %	0 %	0 %	0 %	18 %	1.44 (0.70-2.98)
Thrombosis/Embolism	1 %	1 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	0 %	0 %	0 %	
Bleeding	9 %	1 %	0 %	0 %	0 %	10 %	4 %	2 %	0 %	0 %	0 %	7 %	
Bleeding after surgery	4 %	1 %	0 %	0 %	0 %	5 %	11 %	0 %	0 %	0 %	0 %	11 %	
Patient fall /fracture	1 %	2 %	0 %	0 %	0 %	4 %	0 %	7 %	0 %	0 %	0 %	7 %	0.83 (0.24-2.82)
Patient fall	1 %	1 %	0 %	0 %	0 %	2 %	0 %	2 %	0 %	0 %	0 %	2 %	
Fracture	0 %	1 %	0 %	0 %	0 %	1 %	0 %	4 %	0 %	0 %	0 %	4 %	
Other	1 %	1 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	0 %	0 %	0 %	NA
Allergy	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	

Medical technical harm	1 %	1 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	0 %	0 %	0 %	
Deterioration and cronic illness	0 %	0 %	0 %	0 %	0 %	1 %	0 %	0 %	0 %	0 %	0 %	0 %	
Medication harm	5 %	4 %	0 %	0 %	0 %	10 %	0 %	7 %	2 %	0 %	0 %	9 %	1.68 (0.60-4.66)
Obstetric harm	1 %	0 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	0 %	0 %	0 %	NA
Pressure ulcer	2 %	0 %	0 %	0 %	0 %	2 %	2 %	2 %	0 %	0 %	0 %	4 %	0.73 (0.16-3.33)
Total	57 %	39 %	2 %	0 %	1 %	100 %	56 %	33 %	9 %	2 %	0 %	100 %	

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*

Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1 and 2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any pre-specified hypotheses	3
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4, 5 and 6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*5,6 and 7*
Bias	9	Describe any efforts to address potential sources of bias	10 and 14
Study size	10	Explain how the study size was arrived at	3 and 4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	4, 5 and 6
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6 and 7
		(*b*) Describe any methods used to examine subgroups and interactions	7,8 and 9
		(*c*) Explain how missing data were addressed	Na
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	4
		(*e*) Describe any sensitivity analyses	
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Figure 1
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	6-12
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	10 and 11
Discussion			
Key results	18	Summarise key results with reference to study objectives	10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	11 and 12
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	2

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Does increasing the size of bi-weekly samples of records influence results when using the Global Trigger Tool? An observational study of retrospective record reviews of two different sample sizes

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2015-010700.R1
Article Type:	Research
Date Submitted by the Author:	03-Feb-2016
Complete List of Authors:	Mevik, Kjersti; Nordland Hospital Trust, Regional patient safety resource center Griffin, Frances; Fran Griffin and Associates, LLC Hansen, Tonje; Nordland Hospital Trust, Medical director Deilkås, Ellen; Akershus University Hospital, Centre for Health Service Research Vonen, Barthold; The Artic University of Norway, Institute for community medicine; Nordland Hospital Trust, CMO
Primary Subject Heading:	Qualitative research
Secondary Subject Heading:	Health services research
Keywords:	QUALITATIVE RESEARCH, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Global Trigger Tool, Sample size, Adverse events < THERAPEUTICS

Does increasing the size of bi-weekly samples of records influence results when using the Global Trigger Tool? An observational study of retrospective record reviews of two different sample sizes

Authors

Kjersti Mevik, Regional Patient Safety Resource Center, Nordland Hospital Trust, Bodø, Norway

Frances A Griffin, Fran Griffin & Associates, LLC, Neptune, New Jersey, USA

Tonje E Hansen, Medical director, Nordland Hospital Trust, Bodø, Norway

Ellen T Deilkås, Center for Health Service Research, Akershus University Hospital, Lørenskog, Norway

Barthold Vonen, CMO, Nordland Hospital Trust, Bodø, Norway and Institute for community medicine, The Artic University of Norway, Tromsø, Norway

Corresponding author

Kjersti Mevik, Regional Patient Safety Resource Center, Nordland Hospital Trust, Post box 1480, N-8092 Bodø, Norway

Email: kjersti.mevik@nlsh.no

Phone number: +4797123977

Keywords: qualitative research, quality in health care, adverse events, global trigger tool, sample size

Word count: 3085

Abstract

Objectives: To investigate the impact of increasing sample of records reviewed bi-weekly with the Global Trigger Tool method to identify adverse events in hospitalized patients.

Design: Retrospective observational study.

Setting: A Norwegian 524-bed general hospital trust

Participants: 1920 medical records selected from January 1th to December 31th 2010.

Primary outcomes: Rate, type and severity of adverse events identified in two different samples sizes of records selected as 10 and 70 records bi-weekly.

Results: In the large sample 1.45 (95 % confidence interval: 1.07 to 1.97) times more adverse events per 1000 patient days (39.3 adverse events/1000 patient days) were identified than in the small sample (27.2 adverse events/1000 patient days). Hospital-acquired infections were the most common category of adverse events in both samples and the distributions of the other categories of adverse events did not differ significantly between the samples. The distribution of severity level of adverse events did not differ between the samples.

Conclusions: The findings suggest that while the distribution of categories and severity are not dependent on the sample size, the rate of adverse events is. Further studies are needed to conclude if the optimal sample size may need to be adjusted based on the hospital size in order to detect a more accurate rate of adverse events.

Article summary:

Strength and limitations of this study:

- The samples were similar in terms of age, sex and length of stay.
- Preventability of the adverse events was not assessed.
- Only two sample sizes were compared.
- Method for authentication of events differed slightly for each set of samples however high inter-rater reliability between the review teams indicates consistency and thus did not likely affect results.

This work was supported by The Northern Norwegian Regional Health Authority.

Competing interest: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: KM had financial support from The Northern Norway Regional Health Authority as a PhD grant for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

INTRODUCTION

For more than a decade considerable efforts have been invested across healthcare to reduce adverse events, resulting in many efforts to identify reliable and valid tools to measure such events. The Institute for Healthcare Improvement (IHI) Global Trigger Tool is a widely used and considered an effective tool for measuring adverse events[1–3]. The method includes reviewing bi-weekly samples of ten patient records selected randomly from the hospital discharge lists. Two non-physician reviewers search independently for pre-defined triggers that could indicate possible adverse events. A physician authenticates their consensus on presence of adverse events and severity. The adverse events identified in the bi-weekly periods provide data for Statistical Process Control (SPC) charts used to analyse adverse events rates over time. However, concerns have been raised[2,4–8], about the method's ability to accurately detect rates of adverse events and changes in rates, due to the small sample size of ten records bi-weekly recommended in the IHI method.

In Norway all hospital trusts are required by the National Health Authority to use a translated version of the Global Trigger Tool to review a minimum of ten records selected continuously and bi-weekly in order to monitor the rates of adverse events in each hospital trust and at a national level. Good et al[9] suggest that sample size should be adjusted to hospital size and based on this we increased the sample size at our trust to seven times greater than that required by the Health Authority as we believed this would detect a more accurate rate of adverse events. Our rates of adverse events have been higher than other comparable trusts that are reviewing bi-weekly samples of ten records thus we sought to assess whether our higher rates were due to the larger sample size. The impact of sample size on adverse event rates has not been validated to our knowledge thus demonstrating the need for this study.

Our aim was to obtain the rate, category and severity of identified adverse events in two different sample sizes of records selected from the same population bi-weekly: one sample corresponding to the IHI recommendation and one sample seven times

larger. We hypothesised that increasing the sample size would not yield a different rate of adverse events per 1000 patient days.

METHODS

Study design

The study is an observational cross-sectional study including retrospective record review of two samples of records, respectively 1680 and 240 (figure 1).

Setting

The study was performed in a 524-bed hospital trust at three geographical locations in Nordland County, North-Norway. Both samples were selected from the same population discharged from January 1 to December 31 2010. However the large sample was first stratified according to discharges from the nine services in the trust and then ten records were selected from five services and five records from four services respectively for a total of 70 records bi-weekly. The small sample included ten records selected bi-weekly from the aggregated discharge lists of all the nine services. Following the IHI guidelines, records were excluded in both samples for patients aged 17 years or younger, patients admitted primarily for psychiatric or rehabilitation care, or patients with a length of stay less than 24 hours. The whole hospitalization was reviewed including patient days at all services not only at the index service.

The study was approved by the Data protection official in Nordland Hospital trust and by the Norwegian Regional Ethics Committee (ref 2012/1691).

Record review method

Training of the reviewers followed the IHI recommendations and included theory, practical review exercises, and debriefing sessions provided by experienced

reviewers. The IHI definition of an adverse event was used, i.e.,[1]: *"Unintended physical injury resulting from or contributed to by medical care that requires additional monitoring, treatment or hospitalization, or that results in death".* Both adverse events associated with treatment given prior, during or after (within 30 days) to the index discharge (the discharge selected from the discharge lists of the services) were included to evaluate the total number of adverse events resulting from medical care. Preventability of the identified adverse events was not evaluated.

The identified adverse events were grouped into 23 categories derived from the Norwegian translation[10] of the IHI Global Trigger Tool. These categories were further aggregated into eight main categories (i.e., hospital-acquired infections, surgical complications, bleeding/thrombosis, patient fall/fracture, medication harm, obstetric harm, pressure ulcer and other). The severity of adverse events was categorized into five levels (E – I) using definitions adapted from those of the National Coordinating Council for Medication Error Reporting and Prevention index (NCC MERP)[11]:

Category E:	Temporary harm to the patient and required intervention
Category F:	Temporary harm to the patient and required initial or prolonged hospitalization
Category G:	Permanent patient harm
Category H:	Intervention required to sustain life
Category I:	Patient death

The review process for both sets of samples followed the IHI method[1] were reviewers checked each record for the presence of triggers from a standard list of triggers in the Norwegian translation of the Global Trigger Tool. When a trigger was identified they checked for documentation indicating that an adverse event had occurred; for any adverse event detected, whether by a trigger or not, one of the above eight categories and a severity level was assigned. The process for

authentication of adverse events differed slightly between the two sets of samples. For the small samples, two nurses (reviewer A and reviewer B) each reviewed all records independently and then together reached consensus on presence, category and severity of adverse events. A physician (reviewer C) then authenticated their findings. The reviewing process of authentication with records from the large samples was slightly different in that each record was reviewed by one reviewer – either a nurse (reviewer A) or one of two physicians (reviewers C and D). The three reviewers discussed their findings and reached consensus of presence, category and severity of adverse events identified (figure 1). The modification with only one reviewer per record in the reviewing process for the large samples was due to limited resources available.

Statistical analysis

Demographic variables of the records were obtained. Categorical variables were compared between the samples with Chi-square test while continuous variables were compared using the Independent t-test.

Statistical Process Control (SPC) charts are used to evaluate variations between data points over time which is a recommended approach for evaluating the rates of adverse events measured by the Global Trigger Tool[1,12]. We used QI Macros in Excel 2013 to present the calculated rate of adverse events per 1000 patient days in U-charts and the calculated percentage of records with adverse events in a P-chart of both samples[13]. Test 1-3 of special cause variation (SCV) were applied in order to evaluate the rates. The tests are positive if data points are outside the control limits, eight or more data points are one the same side of the median or/and if six data points are either ascending or descending. We hypothesised that different rates of adverse events in the two samples would yield different results in terms of the tests and control limits.

To compare the calculated rates, proportions of severities and categories of adverse events between the samples we used Poisson regression in generalized linear models to calculate the relative risk of adverse events between the samples as the risk ratio. Poisson regression was chosen as it accounts for variations in the number of cases reviewed and variations in length of stay. The number of adverse events was set as the dependent variable and log patient days as the offset variable (in the analysis of adverse events per patient day). When analysing adverse events per records and percentages of records with an adverse event, zero was set as the fixed value. A p value of < 0.05 was defined as statistically significant. We also adjusted for services and variables associated with the index service. Associations between adverse events and demographic variables were explored using Pearson´s correlation and logistic regression. To assess the inter-rater reliability between the review teams of the two samples we used kappa and weighted kappa statistics. The following interpretations from Landis and Koch was used for the Cohen Kappa coefficient: poor (<0.0), slight (0.00-0.20), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61-0.80) and almost perfect (0.81-1.00)[14]. We used SPSS (version 22.0; SPSS Chicago, IL) for statistical analyses.

RESULTS

Demographics characteristics

A total of 1920 records were reviewed in the study using the Global Trigger Tool. Demographic characteristics in both samples and the overall population from which the samples were drawn from are shown in table 1. 12 % of the overall population (14267 discharges) was reviewed in the large samples while 2 % was reviewed in the small sample. Length of stay, age and sex were derived for the whole hospitalization and these did not differ between the large and the small sample. Patients in the large sample were different to the overall population in terms of sex and length of stay while patients in the small sample did not differ from the overall population. Type of admissions (acute or planned), case mix (discharge diagnose), services (functional units), case mix index, admission to surgery and numbers of transfers were derived from the index discharge (source of the random selection) and adjusted for.

Table 1: Demographic characteristics of the two samples and the overall population

	Samples			p-value			
	Large sample	**Small sample**	**Overall population**	Large versus small sample	Large versus overall population	Small versus overall population	
n=	**1680**	**240**	**14267**				
Length of stay (days)*	6.8 (7.5)	6.9 (11.1)	6.3 (6.9)	0.852	0.014	0.400	±
Average age (years)*	62 (21)	61 (21)	62 (21)	0.487	0.592	0.344	±
Sex (percent women)**	62	59	57	0.446	<0.001	0.410	§

*Values presented as mean with standard deviations. **Values presented as percent. n.s =non-significant = p value>0.05.

±T-test, §Chi-square test

Comparison of adverse events

In the large sample of 1680 records comprising 11367 patient days, we identified 447 adverse events in 347 discharges. This corresponds to a rate of 39.3 adverse events per 1000 patient days (95 % confidence interval (CI): 35.8 to 43.1, standard error (SE) = 1.86) or 26.6 adverse events per 100 discharges (95 % CI: 24.3 to 29.2, SE= 1.26). The percentage of patients with an adverse event was 20.5 % in the large sample. In the small sample of 240 records comprising 1657 patient days, we identified 45 adverse events in 30 discharges. This corresponds to a rate of 27.2 adverse events per 1000 patient days (95 % CI: 20.3 to 36.4, SE = 4.05) or 18.8 adverse events per 100 discharges (95 % CI: 14.0 to 25.1, SE= 2.80). The percentages of patients experiencing an adverse event was 12.5 %. Some patients experienced more than one adverse event. Patients experiencing adverse events had longer hospital stays (large sample r^2=0.21, P<0.001 and small sample r^2=0.46, P<0.001) than patients without experiencing adverse events. In the large sample age

correlated (r^2=0.03, P<0.001) with number of adverse events while in the small sample age did not correlate with number of adverse events (r^2= -0.003, P= 0.54).

The rate of adverse events per 1000 patient days was 45 % higher in the large sample than in the small sample (risk ratio (RR) =1.45, 95 % CI: 1.07 to 1.97; P= 0.02). Likewise the rate of adverse events per record in the large sample was 42% higher in the large sample than in the small sample (RR= 1.42, 95 % CI: 1.04 to 1.93, P= 0.03). The percentages of records including an adverse event was 65 % higher in the large sample than in the small sample (RR=1.65, 95 % CI: 1.14 to 2.34, P=0.008). In figure 2 the rates of adverse events per 1000 patient days in both samples are presented in control U-charts and percentages of records with adverse events in control P-charts over the 24 bi-weekly periods in 2010. In both charts the control limits are much wider in the small sample than in the large sample. Special cause variations (positivity of tests 1) were identified only for the small sample. This is marked with a black dot in the U-chart. None of the other tests were positive for either of the samples.

To adjust for the stratification made before selection of records to the large sample we adjusted for the variables that were associated from the index discharge. The primary results did not alter as the risk ratio was 1.83 (95 % CI: 1.32 to 2.54, P<0.001) of identifying an adverse event per 1000 patient days in the large sample compared to the small sample when adjusting for these variables.

The inter-rater reliability of the two teams that reviewed the different sets of samples was obtained to assess for possible impact from the different authentication processes. The two review teams reviewed a set of 50 patient records and agreement regarding presence of adverse events (kappa (κ)=0.75), number of adverse events (κ=0.68) and severity level (κ=0.69) was substantial.

Hospital-acquired infections were the most frequent category of identified adverse events in both samples. There were no significant differences between the estimated proportions of identified adverse events between the samples for the six main categories of adverse events; hospital-acquired infections (RR=1.52, 95 % CI: 0.94 to 2.47, P=0.09), surgical complications (RR=1.28, 95 % CI: 0.67 to 2.47, P=0.46), bleeding/thrombosis (RR=1.44, 95 % CI: 0.70 to 2.98, P=0.33), medication harm (RR=1.68, 95 % CI: 0.60 to 4.66, P=0.32), patient fall (RR=0.83, 95 % CI: 0.24 to 2.82, P=0.76) and pressure ulcers (RR=0.73, 95 % CI: 0.16 to 3.33, P=0.68) (supplementary file 1). For the categories obstetric harm and other, no adverse events were identified in the small sample and a comparison was not performed.

The least severe adverse events (category E) accounted for more than half of the adverse events identified in both samples. Severity level including prolonged stay accounted for the same amount (30-40 %) in both samples. No significant differences were found between the rate of adverse events per 1000 patient days between the samples, when adverse events were analysed separately according to severity of the adverse events: E (RR=1.50, 95 %CI: 1.00 to 2.26, P=0.05) and F (RR=1.68, 95 % CI: 0.99 to 2.85, P=0.05) and F, G, H and I (RR=0.47, 95 %CI: 0.17 to 1.27, P=0.14) and G, H and I (RR=1.38, 95 % CI: 0.87 to 2.18, P=0.17).

DISCUSSION

The rate of adverse events was 1.45 higher in the large sample than in the small sample. Our findings indicate that the sample size may influence the rate of identified adverse events. The differences regarding CI and SE indicates that increasing the sample size decreases the variation, as expected.

While evaluations of the Global Trigger Tool have reported both high sensitivity[3] and acceptable reliability[15,16] the impact of the sample size in determining the level of adverse events has hardly been discussed. We believe this is the first attempt to assess the impact of the sample size to the rate of adverse events

identified with the Global Trigger Tool. Good et al adjusted the sample size to the hospital sizes without further comparisons between different sample sizes selected in the same time period[9]. We wanted to evaluate whether a larger sample of records reviewed bi-weekly could yield higher rates of adverse events than a sample of ten records reviewed bi-weekly. Our trust had increased our bi-weekly samples to correspond to 10 % of the total number of discharges and found higher rates of adverse events than comparable Norwegian trusts that reviewed samples of ten records bi-weekly. Thus we determined it legitimate, necessary and original to assess whether using the Global Trigger Tool with different sample sizes would produce different results.

While our findings may challenge the sensitivity of the recommended small sample size in order to identify an accurate rate of adverse events, they also underline the ability of that sample size to reflect distribution of severities and categories of adverse events accurately. Our results in terms of this corresponds well with other studies[17,18]. In the small sample none adverse events of category I were identified. This is most likely due to the fact that the Global Trigger Tool is not designed to identify all such cases (category I). Due to their infrequent occurrence, other methods should be used to monitor these specific types of events, for example investigating all hospital deaths[19,20]. Thus we compared the rate of adverse events in category I along with the rate of adverse events in other categories (category F, G and H).

Several factors could explain the differences in the rate of adverse events identified in the two samples. First, the authentication processes differed slightly for the two samples. To assess for possible bias we evaluated the inter-rater reliability of the two teams that reviewed the different samples. We found substantial agreement between the two review teams regarding presences, number and severity level of adverse events, thus conclude that the difference in adverse event rates between samples is not due to bias from the different authentication processes. These findings are supported in the work of Zegers et al[21]. Second, the Simpson paradox defined as statistical results from aggregated data, could give a different result to that of a group-level analysis do[22]. A skewness regarding the variables associated to the

index discharges could be present in our study as the large sample was stratified according to the services before sampling and the small sample was not. However, the primary results did not differ when adjusting for these variables. Neither did the demographic characteristic as sex, age and length of stay did not differ between the large and the small sample. Third, the study was undertaken for only one year of discharges comprising 240 records in the small sample. A meta-analysis of different sample sizes showed that the variation of adverse event rates decreases as the sample size increases[4] thus underlining the importance of having a large enough sample size in order to obtain valid results.

CONCLUSION

We believe the findings in this study could challenge the appropriateness of the sampling methods commonly used as the rate of adverse events increased when the number of records reviewed bi-weekly was increased, though limitations of the study must be considered. The distributions of adverse event categories and severity level did not differ between the samples and only the rate of adverse events appeared to be influenced by the sample size. Further studies are needed to determine whether there is an optimal sample size and whether it should be based on hospital size. Reviewing 10 % of the total discharges may be considered as optimal upon further studies.

Contributors KM and BV designed the study. TEH and KM reviewed the records. KM conducted the data analysis and wrote the first draft of the manuscript and rewrote drafts of the manuscript after all co-authors had reviewed and revised. KM performed the statistical analyses and produced the graphs.

Competing interest None

Funding KM received a grant from the Northern Norway Regional Health Authority

.

Ethics approval Data protection official in Nordland Hospital trust. Norwegian Regional Ethics Committee (ref 2012/1691).

Acknowledgements The authors would like to thank Birger Hveding and Inger Lise Øvre who conducted the reviews together with the co-authors TEH and KM. Frank Federico and Carol Haraden for guidance on the development of the study protocol. Tom Wilsgaard for advice on statistical analysis and Alexander Ringdal and Marina Mineeva for help with data processing.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1 Griffin F, Resar R. IHI Global Trigger Tool for measuring adverse events. *IHI Innov Ser white Pap* 2007;:1–44.http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:IHI+Global+Trigger+Tool+for+Measuring+Adverse+Events#0 (accessed 26 Nov2014).

2 Commission S, Zealand N. The Global Trigger Tool : A Review of the Evidence. Wellington: 2013. www.hqsc.govt.nz

3 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. doi:10.1377/hlthaff.2011.0190

4 Lessing C, Schmitz A, Albers B, *et al.* Impact of sample size on variation of adverse events and preventable adverse events: systematic review on epidemiology and contributing factors. *Qual Saf Health Care* 2010;**19**:e24. doi:10.1136/qshc.2008.031435

5 Mattsson TO, Knudsen JL, Lauritsen J, *et al.* Assessment of the global trigger tool to measure, monitor and evaluate patient safety in cancer patients:

reliability concerns are raised. *BMJ Qual Saf* 2013;**22**:571–9. doi:10.1136/bmjqs-2012-001219

6 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. doi:10.1056/NEJMsa1004404

7 James J. A new, evidence-based estimate of patient harms associated with hospital care. *J Patient Saf* Published Online First: 2013.http://journals.lww.com/journalpatientsafety/Abstract/2013/09000/A_New,_Evidence_based_Estimate_of_Patient_Harms.2.aspx (accessed 8 Jan2015).

8 Shojania KG, Thomas EJ. Trends in adverse events over time: why are we not improving? *BMJ Qual Saf* 2013;**22**:273–7. doi:10.1136/bmjqs-2013-001935

9 Good VS, Saldaña M, Gilder R, *et al.* Large-scale deployment of the Global Trigger Tool across a large hospital system: refinements for the characterisation of adverse events to support patient safety learning opportunities. *BMJ Qual Saf* 2011;**20**:25–30. doi:10.1136/bmjqs.2008.029181

10 Strukturert journalundersøkelse, ved bruk av Global Trigger Tool for å identifisere og måle forekomst av skader i helsetjenesten. Oslo: 2010.

11 Hartwig SC, Denger SD, Schneider PJ. Severity-indexed, incident report-based medication error-reporting program. *Am J Hosp Pharm* 1991;**48**:2611–6.http://www.nccmerp.org/types-medication-errors

12 Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. *Qual Saf Health Care* 2003;**12**:458–64.http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1758030&tool=pmcentrez&rendertype=abstract

13 Provost LP, Murray S. *The Health Care Data Guide: Learning from Data for Improvement*. 2011. http://books.google.com/books?hl=no&lr=&id=pRLcaOkswQsC&pgis=1 (accessed 7 Jan2015).

14 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.http://www.ncbi.nlm.nih.gov/pubmed/843571 (accessed 20 Jul2014).

15 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. doi:10.1111/j.1475-6773.2010.01156.x

16 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. doi:10.1093/intqhc/mzq026

17 Kennerly D a, Saldaña M, Kudyakov R, *et al.* Description and evaluation of adaptations to the global trigger tool to enhance value to adverse event reduction efforts. *J Patient Saf* 2013;**9**:87–95. doi:10.1097/PTS.0b013e31827cdc3b

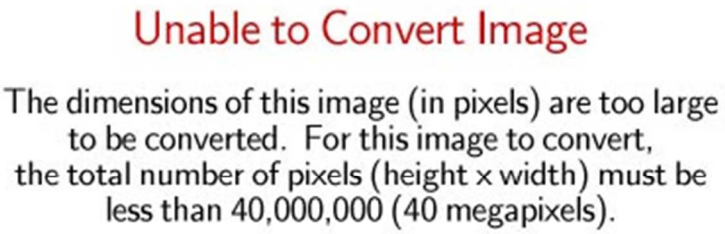
18 Rutberg H, Borgstedt Risberg M, Sjödahl R, *et al.* Characterisations of adverse events detected in a university hospital: a 4-year study using the Global Trigger Tool method. *BMJ Open* 2014;**4**:e004879. doi:10.1136/bmjopen-2014-004879

19 Lau H, Litman KC. Saving lives by studying deaths: Using standardized mortality reviews to improve inpatient safety. Jt. Comm. J. Qual. Patient Saf. 2011;**37**:400–8.

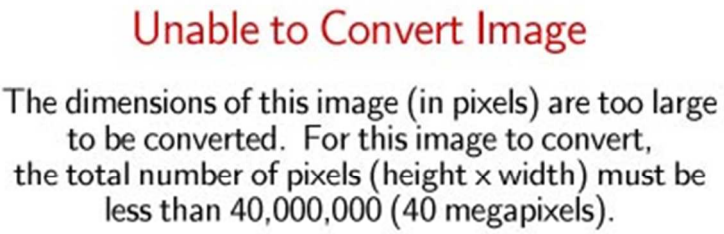
20 Move your dot: Measuring, Evaluating, and Reducing Hospital Mortality Rates (part 1) IHI Innovation Series white paper. Boston: 2003. http://scholar.google.no/scholar?q=Move+Your+Dot%E2%84%A2%3A&btnG=&hl=no&as_sdt=0%2C5#1 (accessed 5 Jun2015).

21 Zegers M, de Bruijne MC, Wagner C, *et al.* The inter-rater agreement of retrospective assessments of adverse events does not improve with two reviewers per patient record. *J Clin Epidemiol* 2010;**63**:94–102. doi:10.1016/j.jclinepi.2009.03.004

22 Blyth CR. On Simpson's Paradox and the Sure-Thing Principle. *J Am Stat Assoc* 1972;**67**:364–6. doi:10.1080/01621459.1972.10482387



Overview of the study design. Non-physician reviewers; reviewer A and B, physician reviewers; reviewer C and D



Comparison of statistical process control charts (U-chart) and (P-chart) between large and small sample.
−−−−= Upper control limits, - - - =Lower control limits

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*

Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1, 2 and 3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any pre-specified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	5
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6,7 and 8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*6,7 and 8*
Bias	9	Describe any efforts to address potential sources of bias	10 and 12-13
Study size	10	Explain how the study size was arrived at	4 and 5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7,8 and 10
		(*b*) Describe any methods used to examine subgroups and interactions	9, 10 and 11
		(*c*) Explain how missing data were addressed	Na
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	5
		(*e*) Describe any sensitivity analyses	
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Figure 1
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8 and 9
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	9-10
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8-11
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	10 and 11
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	12 and 13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	12-13
Generalisability	21	Discuss the generalisability (external validity) of the study results	12-13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	3

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Supplementary file 1: Percentages of types and severity level of the adverse events identified in the large and small sample.

	Large sample size						Small sample size						
	E	F	G	H	I	Total	E	F	G	H	I	Total	Risk ratio (95 %CI)
Hospital acquired infections	26 %	15 %	0 %	0 %	1 %	42 %	31 %	9 %	0 %	0 %	0 %	40 %	1.50 (0.94-2.47)
Urinary tract infection	14 %	2 %	0 %	0 %	0 %	16 %	11 %	0 %	0 %	0 %	0 %	11 %	
CVC infection	0 %	0 %	0 %	0 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	2 %	
Ventilator associated pneumonia	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	
Other infection	8 %	12 %	0 %	0 %	1 %	21 %	9 %	2 %	0 %	0 %	0 %	11 %	
Lower respiratory infection	4 %	1 %	0 %	0 %	0 %	5 %	11 %	4 %	0 %	0 %	0 %	16 %	
Surgical complications	6 %	12 %	1 %	0 %	0 %	20 %	7 %	7 %	7 %	2 %	0 %	22 %	1.28 (0.67-2.47)
Infection after surgery	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	2 %	0 %	2 %	
Respiratory complications after surgery	1 %	1 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	0 %	0 %	0 %	
Return to surgery	0 %	3 %	0 %	0 %	0 %	4 %	0 %	4 %	0 %	0 %	0 %	4 %	
Injury, repair or removal of organ	1 %	1 %	0 %	0 %	0 %	1 %	4 %	0 %	7 %	0 %	0 %	11 %	
Occurrence of any operative complication	2 %	2 %	0 %	0 %	0 %	4 %	2 %	2 %	0 %	0 %	0 %	4 %	
Switch in surgery	0 %	0 %	0 %	0 %	0 %	1 %	0 %	0 %	0 %	0 %	0 %	0 %	
Other	3 %	5 %	0 %	0 %	0 %	8 %	0 %	0 %	0 %	0 %	0 %	0 %	
Bleeding/thrombosis	14 %	3 %	0 %	0 %	0 %	18 %	16 %	2 %	0 %	0 %	0 %	18 %	1.44 (0.70-2.98)
Thrombosis/Embolism	1 %	1 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	0 %	0 %	0 %	
Bleeding	9 %	1 %	0 %	0 %	0 %	10 %	4 %	2 %	0 %	0 %	0 %	7 %	
Bleeding after surgery	4 %	1 %	0 %	0 %	0 %	5 %	11 %	0 %	0 %	0 %	0 %	11 %	
Patient fall /fracture	1 %	2 %	0 %	0 %	0 %	4 %	0 %	7 %	0 %	0 %	0 %	7 %	0.83 (0.24-2.82)
Patient fall	1 %	1 %	0 %	0 %	0 %	2 %	0 %	2 %	0 %	0 %	0 %	2 %	
Fracture	0 %	1 %	0 %	0 %	0 %	1 %	0 %	4 %	0 %	0 %	0 %	4 %	
Other	1 %	1 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	0 %	0 %	0 %	NA
Allergy	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	

Medical technical harm	1 %	1 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	0 %	0 %	0 %	
Deterioration and cronic illness	0 %	0 %	0 %	0 %	0 %	1 %	0 %	0 %	0 %	0 %	0 %	0 %	
Medication harm	5 %	4 %	0 %	0 %	0 %	10 %	0 %	7 %	2 %	0 %	0 %	9 %	1.68 (0.60-4.66)
Obstetric harm	1 %	0 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	0 %	0 %	0 %	NA
Pressure ulcer	2 %	0 %	0 %	0 %	0 %	2 %	2 %	2 %	0 %	0 %	0 %	4 %	0.73 (0.16-3.33)
Total	57 %	39 %	2 %	0 %	1 %	100 %	56 %	33 %	9 %	2 %	0 %	100 %	

BMJ Open

Does increasing the size of bi-weekly samples of records influence results when using the Global Trigger Tool? An observational study of retrospective record reviews of two different sample sizes

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2015-010700.R2
Article Type:	Research
Date Submitted by the Author:	04-Mar-2016
Complete List of Authors:	Mevik, Kjersti; Nordland Hospital Trust, Regional patient safety resource center Griffin, Frances; Fran Griffin and Associates, LLC Hansen, Tonje; Nordland Hospital Trust, Medical director Deilkås, Ellen; Akershus University Hospital, Centre for Health Service Research Vonen, Barthold; The Artic University of Norway, Institute for community medicine; Nordland Hospital Trust, CMO
Primary Subject Heading:	Qualitative research
Secondary Subject Heading:	Health services research
Keywords:	QUALITATIVE RESEARCH, Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Global Trigger Tool, Sample size, Adverse events < THERAPEUTICS

Does increasing the size of bi-weekly samples of records influence results when using the Global Trigger Tool? An observational study of retrospective record reviews of two different sample sizes

Authors

Kjersti Mevik, Regional Patient Safety Resource Center, Nordland Hospital Trust, Bodø, Norway

Frances A Griffin, Fran Griffin & Associates, LLC, Neptune, New Jersey, USA

Tonje E Hansen, Medical director, Nordland Hospital Trust, Bodø, Norway

Ellen T Deilkås, Center for Health Service Research, Akershus University Hospital, Lørenskog, Norway

Barthold Vonen, CMO, Nordland Hospital Trust, Bodø, Norway and Professor, Institute for community medicine, The Artic University of Norway, Tromsø, Norway

Corresponding author

Kjersti Mevik, Regional Patient Safety Resource Center, Nordland Hospital Trust, Post box 1480, N-8092 Bodø, Norway

Email: kjersti.mevik@nlsh.no

Phone number: +4797123977

Keywords: qualitative research, quality in health care, adverse events, global trigger tool, sample size

Word count: 3070

Abstract

Objectives: To investigate the impact of increasing sample of records reviewed bi-weekly with the Global Trigger Tool method to identify adverse events in hospitalized patients.

Design: Retrospective observational study.

Setting: A Norwegian 524-bed general hospital trust

Participants: 1920 medical records selected from January 1th to December 31th 2010.

Primary outcomes: Rate, type and severity of adverse events identified in two different samples sizes of records selected as 10 and 70 records bi-weekly.

Results: In the large sample 1.45 (95 % confidence interval: 1.07 to 1.97) times more adverse events per 1000 patient days (39.3 adverse events/1000 patient days) were identified than in the small sample (27.2 adverse events/1000 patient days). Hospital-acquired infections were the most common category of adverse events in both samples and the distributions of the other categories of adverse events did not differ significantly between the samples. The distribution of severity level of adverse events did not differ between the samples.

Conclusions: The findings suggest that while the distribution of categories and severity are not dependent on the sample size, the rate of adverse events is. Further studies are needed to conclude if the optimal sample size may need to be adjusted based on the hospital size in order to detect a more accurate rate of adverse events.

Article summary:

Strength and limitations of this study:

- The samples were similar in terms of age, sex and length of stay.
- Preventability of the adverse events was not assessed.
- Only two sample sizes were compared.
- Method for authentication of events differed slightly for each set of samples however high inter-rater reliability between the review teams indicates consistency and thus did not likely affect results.

This work was supported by The Northern Norwegian Regional Health Authority.

INTRODUCTION

For more than a decade considerable efforts have been invested across healthcare to reduce adverse events, resulting in many efforts to identify reliable and valid tools to measure such events. The Institute for Healthcare Improvement (IHI) Global Trigger Tool is a widely used and considered an effective tool for measuring adverse events[1–3]. The method includes reviewing bi-weekly samples of ten patient records selected randomly from the hospital discharge lists. Two non-physician reviewers search independently for pre-defined triggers that could indicate possible adverse events. A physician authenticates their consensus on presence of adverse events and severity. The adverse events identified in the bi-weekly periods provide data for Statistical Process Control (SPC) charts used to analyse adverse events rates over time. However, concerns have been raised[2,4–8], about the method's ability to accurately detect rates of adverse events and changes in rates, due to the small sample size of ten records bi-weekly recommended in the IHI method.

In Norway all hospital trusts are required by the National Health Authority to use a translated version of the Global Trigger Tool to review a minimum of ten records selected continuously and bi-weekly in order to monitor the rates of adverse events in each hospital trust and at a national level[9]. Good et al[10] suggest that sample size should be adjusted to hospital size and based on this we increased the sample size at our trust to seven times greater than that required by the Health Authority as we believed this would detect a more accurate rate of adverse events. Our rates of adverse events have been higher than other comparable trusts that are reviewing bi-weekly samples of ten records thus we sought to assess whether our higher rates were due to the larger sample size. The impact of sample size on adverse event rates has not been validated to our knowledge thus demonstrating the need for this study.

Our aim was to obtain the rate, category and severity of identified adverse events in two different sample sizes of records selected from the same population bi-weekly: one sample corresponding to the IHI recommendation and one sample seven times

larger. We hypothesised that increasing the sample size would not yield a different rate of adverse events per 1000 patient days.

METHODS

Study design

The study is an observational cross-sectional study including retrospective record review of two samples of records, respectively 1680 and 240 (figure 1).

Setting

The study was performed in a 524-bed hospital trust at three geographical locations in Nordland County, North-Norway. Both samples were selected from the same population discharged from January 1 to December 31 2010. However the large sample was first stratified according to discharges from the nine services in the trust and then ten records were selected from five services and five records from four services respectively for a total of 70 records bi-weekly. The small sample included ten records selected bi-weekly from the aggregated discharge lists of all the nine services. Following the IHI guidelines, records were excluded in both samples for patients aged 17 years or younger, patients admitted primarily for psychiatric or rehabilitation care, or patients with a length of stay less than 24 hours. The whole hospitalization was reviewed including patient days at all services not only at the index service.

The study was approved by the Data protection official in Nordland Hospital trust and by the Norwegian Regional Ethics Committee (ref 2012/1691).

Record review method

Training of the reviewers followed the IHI recommendations and included theory, practical review exercises, and debriefing sessions provided by experienced

reviewers. The IHI definition of an adverse event was used, i.e.,[1]: *"Unintended physical injury resulting from or contributed to by medical care that requires additional monitoring, treatment or hospitalization, or that results in death".* Both adverse events associated with treatment given prior, during or after (within 30 days) to the index discharge (the discharge selected from the discharge lists of the services) were included to evaluate the total number of adverse events resulting from medical care. Preventability of the identified adverse events was not evaluated.

The identified adverse events were grouped into 23 categories derived from the Norwegian translation[11] of the IHI Global Trigger Tool. These categories were further aggregated into eight main categories (i.e., hospital-acquired infections, surgical complications, bleeding/thrombosis, patient fall/fracture, medication harm, obstetric harm, pressure ulcer and other). The severity of adverse events was categorized into five levels (E – I) using definitions adapted from those of the National Coordinating Council for Medication Error Reporting and Prevention index (NCC MERP)[12]:

Category E:	Temporary harm to the patient and required intervention
Category F:	Temporary harm to the patient and required initial or prolonged hospitalization
Category G:	Permanent patient harm
Category H:	Intervention required to sustain life
Category I:	Patient death

The review process for both sets of samples followed the IHI method[1] were reviewers checked each record for the presence of triggers from a standard list of triggers in the Norwegian translation of the Global Trigger Tool. When a trigger was identified they checked for documentation indicating that an adverse event had occurred; for any adverse event detected, whether by a trigger or not, one of the above eight categories and a severity level was assigned. The process for

authentication of adverse events differed slightly between the two sets of samples. For the small samples, two nurses (reviewer A and reviewer B) each reviewed all records independently and then together reached consensus on presence, category and severity of adverse events. A physician (reviewer C) then authenticated their findings. The reviewing process of authentication with records from the large samples was slightly different in that each record was reviewed by one reviewer – either a nurse (reviewer A) or one of two physicians (reviewers C and D). The three reviewers discussed their findings and reached consensus of presence, category and severity of adverse events identified (figure 1). The modification with only one reviewer per record in the reviewing process for the large samples was due to limited resources available.

Statistical analysis

Demographic variables of the records were obtained. Categorical variables were compared between the samples with Chi-square test while continuous variables were compared using the Independent t-test.

Statistical Process Control (SPC) charts are used to evaluate variations between data points over time which is a recommended approach for evaluating the rates of adverse events measured by the Global Trigger Tool[1,13]. We used QI Macros in Excel 2013 to present the calculated rate of adverse events per 1000 patient days in U-charts and the calculated percentage of records with adverse events in a P-chart of both samples[14]. Test 1-3 of special cause variation (SCV) were applied in order to evaluate the rates. The tests are positive if data points are outside the control limits, eight or more data points are on the same side of the median or/and if six data points are either ascending or descending. We hypothesised that different rates of adverse events in the two samples would yield different results in terms of the tests and control limits.

To compare the calculated rates, proportions of severities and categories of adverse events between the samples we used Poisson regression in generalized linear models to calculate the relative risk of adverse events between the samples as the risk ratio. Poisson regression was chosen as it accounts for variations in the number of cases reviewed and variations in length of stay. The number of adverse events was set as the dependent variable and log patient days as the offset variable (in the analysis of adverse events per patient day). When analysing adverse events per records and percentages of records with an adverse event, zero was set as the fixed value. A p value of < 0.05 was defined as statistically significant. We also adjusted for services and variables associated with the index service. Associations between adverse events and demographic variables were explored using Pearson´s correlation and logistic regression. To assess the inter-rater reliability between the review teams of the two samples we used kappa and weighted kappa statistics. The following interpretations from Landis and Koch was used for the Cohen Kappa coefficient: poor (<0.0), slight (0.00-0.20), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61-0.80) and almost perfect (0.81-1.00)[15]. We used SPSS (version 22.0; SPSS Chicago, IL) for statistical analyses.

RESULTS

Demographics characteristics

A total of 1920 records were reviewed in the study using the Global Trigger Tool. Demographic characteristics in both samples and the overall population from which the samples were drawn from are shown in table 1. 12 % of the overall population (14267 discharges) was reviewed in the large samples while 2 % was reviewed in the small sample. Length of stay, age and sex were derived for the whole hospitalization and these did not differ between the large and the small sample. Patients in the large sample were different to the overall population in terms of sex and length of stay while patients in the small sample did not differ from the overall population. Type of admissions (acute or planned), case mix (discharge diagnose), services (functional units), case mix index, admission to surgery and numbers of transfers were derived from the index discharge (source of the random selection) and adjusted for.

Table 1: Demographic characteristics of the two samples and the overall population

	Samples			p-value			
	Large sample	Small sample	Overall population	Large versus small sample	Large versus overall population	Small versus overall population	
n=	**1680**	**240**	**14267**				
Length of stay (days)*	6.8 (7.5)	6.9 (11.1)	6.3 (6.9)	0.852	0.014	0.400	±
Average age (years)*	62 (21)	61 (21)	62 (21)	0.487	0.592	0.344	±
Sex (percent women)**	62	59	57	0.446	<0.001	0.410	§

*Values presented as mean with standard deviations. **Values presented as percent. n.s =non-significant = p value>0.05.

±T-test, §Chi-square test

Comparison of adverse events

In the large sample of 1680 records comprising 11367 patient days, we identified 447 adverse events in 347 discharges. This corresponds to a rate of 39.3 adverse events per 1000 patient days (95 % confidence interval (CI): 35.8 to 43.1, standard error (SE) = 1.86) or 26.6 adverse events per 100 discharges (95 % CI: 24.3 to 29.2, SE= 1.26). The percentage of patients with an adverse event was 20.5 % in the large sample. In the small sample of 240 records comprising 1657 patient days, we identified 45 adverse events in 30 discharges. This corresponds to a rate of 27.2 adverse events per 1000 patient days (95 % CI: 20.3 to 36.4, SE = 4.05) or 18.8 adverse events per 100 discharges (95 % CI: 14.0 to 25.1, SE= 2.80). The percentages of patients experiencing an adverse event was 12.5 %. Some patients experienced more than one adverse event. Patients experiencing adverse events had longer hospital stays (large sample r^2=0.21, P<0.001 and small sample r^2=0.46, P<0.001) than patients without experiencing adverse events. In the large sample age

correlated (r^2=0.03, P<0.001) with number of adverse events while in the small sample age did not correlate with number of adverse events (r^2= -0.003, P= 0.54).

The rate of adverse events per 1000 patient days was 45 % higher in the large sample than in the small sample (risk ratio (RR) =1.45, 95 % CI: 1.07 to 1.97; P= 0.02). Likewise the rate of adverse events per record in the large sample was 42% higher in the large sample than in the small sample (RR= 1.42, 95 % CI: 1.04 to 1.93, P= 0.03). The percentages of records including an adverse event was 65 % higher in the large sample than in the small sample (RR=1.65, 95 % CI: 1.14 to 2.34, P=0.008). In figure 2 the rates of adverse events per 1000 patient days in both samples are presented in control U-charts and percentages of records with adverse events in control P-charts over the 24 bi-weekly periods in 2010. In both charts the control limits are much wider in the small sample than in the large sample. Special cause variations (positivity of tests 1) were identified only for the small sample. This is marked with a black dot in the U-chart. None of the other tests were positive for either of the samples.

To adjust for the stratification made before selection of records to the large sample we adjusted for the variables that were associated from the index discharge. The primary results did not alter as the risk ratio was 1.83 (95 % CI: 1.32 to 2.54, P<0.001) of identifying an adverse event per 1000 patient days in the large sample compared to the small sample when adjusting for these variables.

The inter-rater reliability of the two teams that reviewed the different sets of samples was obtained to assess for possible impact from the different authentication processes. The two review teams reviewed a set of 50 patient records and agreement regarding presence of adverse events (kappa (κ) =0.75), number of adverse events (κ=0.68) and severity level (κ=0.69) was substantial.

Hospital-acquired infections were the most frequent category of identified adverse events in both samples. There were no significant differences between the estimated proportions of identified adverse events between the samples for the six main categories of adverse events; hospital-acquired infections (RR=1.52, 95 % CI: 0.94 to 2.47, P=0.09), surgical complications (RR=1.28, 95 % CI: 0.67 to 2.47, P=0.46), bleeding/thrombosis (RR=1.44, 95 % CI: 0.70 to 2.98, P=0.33), medication harm (RR=1.68, 95 % CI: 0.60 to 4.66, P=0.32), patient fall (RR=0.83, 95 % CI: 0.24 to 2.82, P=0.76) and pressure ulcers (RR=0.73, 95 % CI: 0.16 to 3.33, P=0.68) (supplementary file 1). For the categories obstetric harm and other, no adverse events were identified in the small sample and a comparison was not performed.

The least severe adverse events (category E) accounted for more than half of the adverse events identified in both samples. Severity level including prolonged stay accounted for the same amount (30-40 %) in both samples. No significant differences were found between the rate of adverse events per 1000 patient days between the samples, when adverse events were analysed separately according to severity of the adverse events: E (RR=1.50, 95 %CI: 1.00 to 2.26, P=0.05) and F (RR=1.68, 95 % CI: 0.99 to 2.85, P=0.05) and F, G, H and I (RR=0.47, 95 %CI: 0.17 to 1.27, P=0.14) and G, H and I (RR=1.38, 95 % CI: 0.87 to 2.18, P=0.17).

DISCUSSION

The rate of adverse events was 1.45 higher in the large sample than in the small sample. Our findings indicate that the sample size may influence the rate of identified adverse events. The differences in CI and SE indicates that increasing the sample size decreases the variation, as expected. We believe that the higher rate of adverse events detected was due to the use of a larger sample and may be more reflective of the total population given the size of the hospital. Since the distribution of severity level and types of adverse events were the same in both sample sizes, we suggest that these distributions are unaffected by sample size.

While evaluations of the Global Trigger Tool have reported both high sensitivity[3] and acceptable reliability[16,17] the impact of the sample size in determining the

level of adverse events has hardly been discussed. We believe this is the first attempt to assess the impact of the sample size to the rate of adverse events identified with the Global Trigger Tool. Good et al adjusted the sample size to the hospital sizes without further comparisons between different sample sizes selected in the same time period[10]. We wanted to evaluate whether a larger sample of records reviewed bi-weekly could yield higher rates of adverse events than a sample of ten records reviewed bi-weekly. Our trust had increased our bi-weekly samples to correspond to 12 % of the total number of discharges and found higher rates of adverse events than comparable Norwegian trusts that reviewed samples of ten records bi-weekly. Thus we determined it legitimate, necessary and original to assess whether using the Global Trigger Tool with different sample sizes would produce different results.

While our findings may challenge the sensitivity of the recommended small sample size in order to identify an accurate rate of adverse events, they also underline the ability of that sample size to reflect distribution of severities and categories of adverse events accurately. Our results in terms of this corresponds well with other studies[18,19]. In the small sample no adverse events of category I were identified. This is most likely due to the fact that the Global Trigger Tool is not designed to identify all such cases (category I). Due to their infrequent occurrence, other methods should be used to monitor these specific types of events, for example investigating all hospital deaths[20,21]. Thus we compared the rate of adverse events in category I along with the rate of adverse events in other categories (category F, G and H).

Several factors could explain the differences in the rate of adverse events identified in the two samples. First, the authentication processes differed slightly for the two samples. To assess for possible bias we evaluated the inter-rater reliability of the two teams that reviewed the different samples. We found substantial agreement between the two review teams regarding presence, number and severity level of adverse events, thus conclude that the difference in adverse event rates between the samples are most likely not due to bias from the minor difference in authentication processes. These findings are supported by the work of Zegers et al[22]. Second, the

Simpson paradox, implying that statistical results from aggregated data could give a different result from a group-level analysis [23]. A skewness regarding the variables associated to the index discharges could be present in our study as the large sample was stratified according to the services before sampling and the small sample was not. However, the primary results did not differ when adjusting for these variables. Neither did the demographic characteristics sex, age and length of stay differ between the large and the small sample. Third, the study was undertaken for only one year of discharges comprising 240 records in the small sample. A meta-analysis of different sample sizes showed that the variation of adverse event rates decreases as the sample size increases[4] thus underlining the importance of having a large enough sample size in order to obtain valid results.

CONCLUSION

We believe the findings in this study could challenge the appropriateness of the sampling methods commonly used as the rate of adverse events increased when the number of records reviewed bi-weekly was increased, though limitations of the study must be considered. The distributions of adverse event categories and severity level did not differ between the samples and only the rate of adverse events appeared to be influenced by the sample size. Further studies are needed to determine whether there is an optimal sample size and if it should be based on hospital size, especially as reviewing larger sample sizes requires more resources. Until further studies, we suggest using a relative increase in sample size to 8-10 % of total number of discharges.

Contributors KM and BV designed the study. TEH and KM reviewed the records. KM conducted the data analysis and wrote the first draft of the manuscript and rewrote drafts of the manuscript after all co-authors had reviewed and revised. KM performed the statistical analyses and produced the graphs.

Competing interest None

Funding KM received a grant from the Northern Norway Regional Health Authority

Ethics approval Data protection official in Nordland Hospital trust. Norwegian Regional Ethics Committee (ref 2012/1691).

Data sharing No additional data available.

Acknowledgements The authors would like to thank Birger Hveding and Inger Lise Øvre who conducted the reviews together with the co-authors TEH and KM. Frank Federico and Carol Haraden for guidance on the development of the study protocol. Tom Wilsgaard for advice on statistical analysis and Alexander Ringdal and Marina Mineeva for help with data processing.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1 Griffin F, Resar R. IHI Global Trigger Tool for measuring adverse events. *IHI Innov Ser white Pap* 2007;:1–44.http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:IHI+Global+Trigger+Tool+for+Measuring+Adverse+Events#0 (accessed 26 Nov2014).

2 Commission S, Zealand N. The Global Trigger Tool : A Review of the Evidence. Wellington: 2013. www.hqsc.govt.nz

3 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. doi:10.1377/hlthaff.2011.0190

4 Lessing C, Schmitz A, Albers B, *et al.* Impact of sample size on variation of adverse events and preventable adverse events: systematic review on epidemiology and contributing factors. *Qual Saf Health Care* 2010;**19**:e24. doi:10.1136/qshc.2008.031435

5 Mattsson TO, Knudsen JL, Lauritsen J, *et al.* Assessment of the global trigger tool to measure, monitor and evaluate patient safety in cancer patients: reliability concerns are raised. *BMJ Qual Saf* 2013;**22**:571–9. doi:10.1136/bmjqs-2012-001219

6 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. doi:10.1056/NEJMsa1004404

7 James J. A new, evidence-based estimate of patient harms associated with

hospital care. *J Patient Saf* Published Online First: 2013.http://journals.lww.com/journalpatientsafety/Abstract/2013/09000/A_New,_Evidence_based_Estimate_of_Patient_Harms.2.aspx (accessed 8 Jan2015).

8 Shojania KG, Thomas EJ. Trends in adverse events over time: why are we not improving? *BMJ Qual Saf* 2013;**22**:273–7. doi:10.1136/bmjqs-2013-001935

9 Deilkås E, Bukholm G, Lindstrøm J, *et al.* Monitoring adverse events in Norwegian hospitals from 2010 to 2013. *BMJ Open* Published Online First: 2015.http://bmjopen.bmj.com/content/5/12/e008576.short (accessed 4 Mar2016).

10 Good VS, Saldaña M, Gilder R, *et al.* Large-scale deployment of the Global Trigger Tool across a large hospital system: refinements for the characterisation of adverse events to support patient safety learning opportunities. *BMJ Qual Saf* 2011;**20**:25–30. doi:10.1136/bmjqs.2008.029181

11 Strukturert journalundersøkelse, ved bruk av Global Trigger Tool for å identifisere og måle forekomst av skader i helsetjenesten. Oslo: 2010.

12 Hartwig SC, Denger SD, Schneider PJ. Severity-indexed, incident report-based medication error-reporting program. *Am J Hosp Pharm* 1991;**48**:2611–6.http://www.nccmerp.org/types-medication-errors

13 Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. *Qual Saf Health Care* 2003;**12**:458–64.http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1758030&tool=pmcentrez&rendertype=abstract

14 Provost LP, Murray S. *The Health Care Data Guide: Learning from Data for Improvement*. 2011. http://books.google.com/books?hl=no&lr=&id=pRLcaOkswQsC&pgis=1 (accessed 7 Jan2015).

15 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.http://www.ncbi.nlm.nih.gov/pubmed/843571 (accessed 20 Jul2014).

16 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. doi:10.1111/j.1475-6773.2010.01156.x

17 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. doi:10.1093/intqhc/mzq026

18 Kennerly D a, Saldaña M, Kudyakov R, *et al.* Description and evaluation of adaptations to the global trigger tool to enhance value to adverse event reduction efforts. *J Patient Saf* 2013;**9**:87–95. doi:10.1097/PTS.0b013e31827cdc3b

19 Rutberg H, Borgstedt Risberg M, Sjödahl R, *et al.* Characterisations of adverse events detected in a university hospital: a 4-year study using the Global Trigger Tool method. *BMJ Open* 2014;**4**:e004879. doi:10.1136/bmjopen-2014-004879

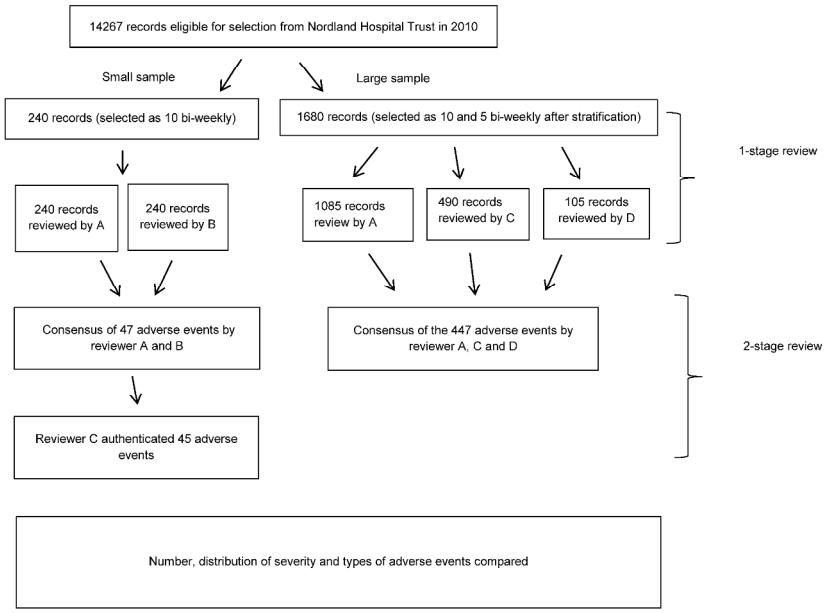
20 Lau H, Litman KC. Saving lives by studying deaths: Using standardized

mortality reviews to improve inpatient safety. Jt. Comm. J. Qual. Patient Saf. 2011;**37**:400–8.

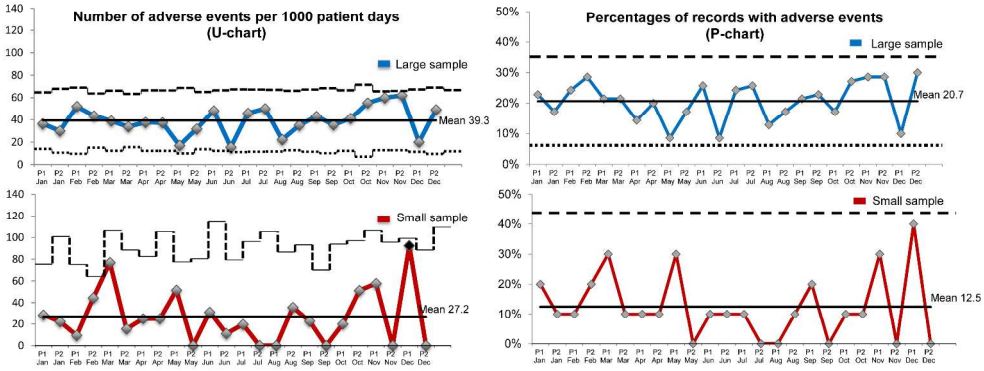
21 Move your dot: Measuring, Evaluating, and Reducing Hospital Mortality Rates (part 1) IHI Innovation Series white paper. Boston: 2003. http://scholar.google.no/scholar?q=Move+Your+Dot%E2%84%A2%3A&btnG=&hl=no&as_sdt=0%2C5#1 (accessed 5 Jun2015).

22 Zegers M, de Bruijne MC, Wagner C, *et al.* The inter-rater agreement of retrospective assessments of adverse events does not improve with two reviewers per patient record. *J Clin Epidemiol* 2010;**63**:94–102. doi:10.1016/j.jclinepi.2009.03.004

23 Blyth CR. On Simpson's Paradox and the Sure-Thing Principle. *J Am Stat Assoc* 1972;**67**:364–6. doi:10.1080/01621459.1972.10482387



Overview of the study design. Non-physician reviewers; reviewer A and B, physician reviewers; reviewer C and D
215x279mm (300 x 300 DPI)



Comparison of statistical process control charts (U-chart) and (P-chart) between large and small sample.
−−−−= Upper control limits, - - - =Lower control limits
297x210mm (300 x 300 DPI)

Supplementary file 1: Percentages of types and severity level of the adverse events identified in the large and small sample.

	Large sample size						Small sample size						
	E	F	G	H	I	Total	E	F	G	H	I	Total	Risk ratio (95 %CI)
Hospital acquired infections	26 %	15 %	0 %	0 %	1 %	42 %	31 %	9 %	0 %	0 %	0 %	40 %	1.50 (0.94-2.47)
Urinary tract infection	14 %	2 %	0 %	0 %	0 %	16 %	11 %	0 %	0 %	0 %	0 %	11 %	
CVC infection	0 %	0 %	0 %	0 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	2 %	
Ventilator associated pneumonia	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	
Other infection	8 %	12 %	0 %	0 %	1 %	21 %	9 %	2 %	0 %	0 %	0 %	11 %	
Lower respiratory infection	4 %	1 %	0 %	0 %	0 %	5 %	11 %	4 %	0 %	0 %	0 %	16 %	
Surgical complications	6 %	12 %	1 %	0 %	0 %	20 %	7 %	7 %	7 %	2 %	0 %	22 %	1.28 (0.67-2.47)
Infection after surgery	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	2 %	0 %	2 %	
Respiratory complications after surgery	1 %	1 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	0 %	0 %	0 %	
Return to surgery	0 %	3 %	0 %	0 %	0 %	4 %	0 %	4 %	0 %	0 %	0 %	4 %	
Injury, repair or removal of organ	1 %	1 %	0 %	0 %	0 %	1 %	4 %	0 %	7 %	0 %	0 %	11 %	
Occurrence of any operative complication	2 %	2 %	0 %	0 %	0 %	4 %	2 %	2 %	0 %	0 %	0 %	4 %	
Switch in surgery	0 %	0 %	0 %	0 %	0 %	1 %	0 %	0 %	0 %	0 %	0 %	0 %	
Other	3 %	5 %	0 %	0 %	0 %	8 %	0 %	0 %	0 %	0 %	0 %	0 %	
Bleeding/thrombosis	14 %	3 %	0 %	0 %	0 %	18 %	16 %	2 %	0 %	0 %	0 %	18 %	1.44 (0.70-2.98)
Thrombosis/Embolism	1 %	1 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	0 %	0 %	0 %	
Bleeding	9 %	1 %	0 %	0 %	0 %	10 %	4 %	2 %	0 %	0 %	0 %	7 %	
Bleeding after surgery	4 %	1 %	0 %	0 %	0 %	5 %	11 %	0 %	0 %	0 %	0 %	11 %	
Patient fall /fracture	1 %	2 %	0 %	0 %	0 %	4 %	0 %	7 %	0 %	0 %	0 %	7 %	0.83 (0.24-2.82)
Patient fall	1 %	1 %	0 %	0 %	0 %	2 %	0 %	2 %	0 %	0 %	0 %	2 %	
Fracture	0 %	1 %	0 %	0 %	0 %	1 %	0 %	4 %	0 %	0 %	0 %	4 %	
Other	1 %	1 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	0 %	0 %	0 %	NA
Allergy	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	

Medical technical harm	1 %	1 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	0 %	0 %	0 %	
Deterioration and cronic illness	0 %	0 %	0 %	0 %	0 %	1 %	0 %	0 %	0 %	0 %	0 %	0 %	
Medication harm	5 %	4 %	0 %	0 %	0 %	10 %	0 %	7 %	2 %	0 %	0 %	9 %	1.68 (0.60-4.66)
Obstetric harm	1 %	0 %	0 %	0 %	0 %	2 %	0 %	0 %	0 %	0 %	0 %	0 %	NA
Pressure ulcer	2 %	0 %	0 %	0 %	0 %	2 %	2 %	2 %	0 %	0 %	0 %	4 %	0.73 (0.16-3.33)
Total	57 %	39 %	2 %	0 %	1 %	100 %	56 %	33 %	9 %	2 %	0 %	100 %	

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*

Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1, 2 and 3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any pre-specified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	5
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6,7 and 8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*6,7 and 8*
Bias	9	Describe any efforts to address potential sources of bias	10 and 12-13
Study size	10	Explain how the study size was arrived at	4 and 5
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7,8 and 10
		(*b*) Describe any methods used to examine subgroups and interactions	9, 10 and 11
		(*c*) Explain how missing data were addressed	Na
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed	

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	5
		(*e*) Describe any sensitivity analyses	
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Figure 1
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	8 and 9
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	9-10
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8-11
		(*b*) Report category boundaries when continuous variables were categorized	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	10 and 11
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	12 and 13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	12-13
Generalisability	21	Discuss the generalisability (external validity) of the study results	12-13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	3

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.