

Accuracy of surgical complication rate estimation using ICD-10 codes

A. Storesund^{1,5} , A. S. Haugen¹, M. Hjortås⁷, M. W. Nortvedt^{3,4}, H. Flaatten^{1,5}, G. E. Eide^{2,6}, M. A. Boormeester⁸, N. Sevdalis⁹ and E. Søfteland^{1,5}

¹Department of Anaesthesia and Intensive Care and ²Centre for Clinical Research, Haukeland University Hospital, ³Centre for Evidence-Based Practice, Western Norway University of Applied Sciences, ⁴Department of Public Health and Services, City of Bergen, and Departments of ⁵Clinical Medicine and ⁶Global Public Health and Primary Care, University of Bergen, Bergen, and ⁷Department of Surgery, Førde Central Hospital, Førde, Norway, ⁸Department of Surgery, Academic Medical Centre Amsterdam, Amsterdam, the Netherlands, and ⁹Centre for Implementation Science, Health Service and Population Research Department, King's College London, London, UK

Correspondence to: Mrs A. Storesund, Department of Anaesthesia and Intensive Care, Haukeland University Hospital, Jonas Liesvei 65, N-5021 Bergen, Norway (e-mail: anette.storesund@helse-bergen.no)

Background: The ICD-10 codes are used globally for comparison of diagnoses and complications, and are an important tool for the development of patient safety, healthcare policies and the health economy. The aim of this study was to investigate the accuracy of verified complication rates in surgical admissions identified by ICD-10 codes and to validate these estimates against complications identified using the established Global Trigger Tool (GTT) methodology.

Methods: This was a prospective observational study of a sample of surgical admissions in two Norwegian hospitals. Complications were identified and classified by two expert GTT teams who reviewed patients' medical records. Three trained reviewers verified ICD-10 codes indicating a complication present on admission or emerging in hospital.

Results: A total of 700 admissions were drawn randomly from 12 966 procedures. Some 519 possible complications were identified in 332 of 700 admissions (47.4 per cent) from ICD-10 codes. Verification of the ICD-10 codes against information from patients' medical records confirmed 298 as in-hospital complications in 141 of 700 admissions (20.1 per cent). Using GTT methodology, 331 complications were found in 212 of 700 admissions (30.3 per cent). Agreement between the two methods reached 83.3 per cent after verification of ICD-10 codes. The odds ratio for identifying complications using the GTT increased from 5.85 (95 per cent c.i. 4.06 to 8.44) to 25.38 (15.41 to 41.79) when ICD-10 complication codes were verified against patients' medical records.

Conclusion: Verified ICD-10 codes strengthen the accuracy of complication rates. Use of non-verified complication codes from administrative systems significantly overestimates in-hospital surgical complication rates.

Paper accepted 26 July 2018

Published online in Wiley Online Library (www.bjs.co.uk). DOI: 10.1002/bjs.10985

Introduction

The Institute of Medicine's seminal report¹ on medical errors initiated safety awareness and implementation of preventive patient safety strategies. Patient harm remains a challenge in healthcare and up to 35 per cent of patients are exposed to complications during their hospital stay². A majority of identified complications (over 65 per cent) are attributed to surgical care^{3–5}.

A number of methods have been used to detect adverse events, patient harm or complications. These include prospective observation of unfolding care processes⁶, the

Clavien–Dindo classification of complications⁷, incident reporting⁸, and retrospective review of patient records, such as the Harvard method⁹ and the Global Trigger Tool (GTT) developed by the Institute for Healthcare Improvement (IHI)¹⁰. Under-reporting of complications in incident reporting systems remains a challenge¹¹. Full record review is thought to identify most complications, with the GTT method revealing ten times more complications than other methods¹². The GTT involves searching for 'trigger' words that can indicate a complication (such as decubitus, intubation, naloxone), tracking changes over

time¹³, and studying the effect of new interventions to improve patient safety¹⁴. The GTT is labour-intensive, and therefore mostly recommended for internal use. A less resource-demanding alternative is to use electronically extracted disease and complication codes from hospital administrative data that have already been entered into hospital databases^{15,16}.

ICD-9 and ICD-10 have been used by more than 100 countries, and contributed to more than 20 000 scientific publications¹⁷. In Norway, it has been mandatory to use the ICD-10 system since 1999. Discharging physicians have to code diseases and complications that are detected in patient records and hospital administrative systems. The codes are frequently also used for reimbursement. Comparing data on complications across nations based on ICD-10 codes is common, but, owing to variation in coding practices and poor quality of registered data, caution in interpreting patterns and comparisons is advised¹⁸.

Surgical complications often have a significant personal, family, economic and thus wider societal impact. Reliable knowledge of codes indicating complications, and methods to apply them, are warranted. Concerns have been raised regarding the reliability and validity of different diagnostic codes, such as those for venous thromboembolism¹⁹, stroke^{20,21}, sepsis²², infections²³ and myocardial infarction²⁴.

Consistent knowledge of surgical complications may inform and could influence healthcare policies and facilitate future safety targets. The aim of the present study was to investigate the accuracy of using ICD-10-coded surgical complications compared with the GTT as a reference standard, by conducting a concurrent validation study of ICD-10-coded complications. The ICD-10 classification system and the GTT method were chosen as they are well established nationally and globally. The hypothesis was that ICD-10 codes identifying complications, as currently used, overestimate actual procedure-related complications, especially as those present on admission are not distinguished from complications that arise during the hospital stay.

Methods

This observational study with prospective data collection investigated perioperative complications in two Norwegian hospitals: one tertiary teaching hospital (referral for 1·1 million inhabitants) and one community hospital (referral for 110 000 inhabitants). A sample of surgical admissions was drawn randomly from a larger group comprising various surgical procedures. Adult surgical patients (aged at least 18 years) admitted for hospital care (lasting at least 24 h) between November 2012 and March 2015 were included from the two hospitals. Exclusion criteria were:

rehabilitation admissions, ambulatory patients, donor surgery and patients who declined to participate in the study. The study was approved by the Western Norway Regional Ethical Research Committee (2012/560/REK West) and the data privacy unit at the central community hospital (Ref: 2012/3060). The study protocol was registered in ClinicalTrials.gov (NCT01872195).

Global Trigger Tool

The GTT was used to identify complications in patients' medical records. GTT-identified complications are covered by the IHI's definition of an adverse event: 'an unintended physical injury resulting from or contributed to by medical care that requires additional monitoring, treatment or hospitalization, or has a fatal outcome'¹³. The GTT method involves a two-stage review process performed by nurses and physicians. Reviewers searched for 'trigger' words that may or may not indicate patient harm. The Norwegian GTT protocol based on the IHI guidelines was followed¹³. Two GTT teams investigated patient records to identify any word from 55 predefined trigger words that could indicate patient harm. A positive trigger word led the two teams to classify the occurrence of complications from a list of 23 categories. Both teams consisted of registered nurses with clinical experience ranging from 7 to 35 years, and experience with use of the GTT ranging from beginner to 5 years. One team included a senior anaesthetist and the other a surgeon. The members of the two teams received a joint 2-h educational session delivered by two doctors experienced in use of the GTT. According to the GTT protocol, the teams reviewed medical summaries, medication logs, laboratory results, prescriptions, surgical procedural records, anaesthesia records, nursing registrations, discharge records, ICD-10 codes and other relevant documentation.

Severity of complications identified by the GTT was classified according to the international GTT template that is used routinely by Norwegian hospitals (not only as part of the present study): E, temporary harm – additional monitoring or treatment needed; F, temporary harm – initial or extended hospital stay; G, permanent harm; H, life-supporting treatment needed; and I, death²⁵. In admissions with several GTT-identified complications describing the same injury, the complication contributing to the injury was allocated a severity level. An example is postoperative bleeding resulting in reoperation: this was analysed as one complication (bleeding) with one severity level (F).

ICD-10 complication codes

Primary outcomes were complications during in-hospital care. A complication was defined as an adverse outcome:

‘an unintended and undesired occurrence in the healthcare process, which causes harm to the patient’²⁶. The ICD-10 codes indicating complications were identified by using complications as classified by the American College of Surgeons’ National Surgical Quality Improvement Program²⁷ and studies investigating surgical complications^{28–30}. Based on previous research publications on checklists and surgical complications, 154 ICD-10 complication codes were included in this study (*Table S1*, supporting information).

The codes investigated were extracted electronically from patient medical records using the hospital administrative data systems for routinely collected data. All patient records with any identified ICD-10 complication code were reviewed to verify whether the ICD-10 complication code was already linked to the patient’s condition at the time of admission or arose during the hospital stay. A complication resulting from a previous admission rather than the present one was not included as a complication in the admission analysed in the present study. Three clinical researchers (an intensive care nurse, a nurse anaesthetist and a senior intensivist), different from the GTT teams, independently reviewed the patient’s medical records and verified the codes as indicative of a complication already being present on admission, or one that emerged during the hospital stay and/or at discharge. Admissions with one or two complications were classified by a single reviewer. All admissions with three or more complications were discussed between all three reviewers, and consensus was obtained to ensure agreement in number and types of complications. The ICD-10 complication code reviewers and the GTT record review teams were blinded to each other’s reviews.

Reliability and validity

Reliability was assessed for both teams classifying complications using the GTT method in the same 20 random medical records. After classification, agreement on the presence of a complication, numbers of complications and levels of severity was tested. In addition, three clinical researchers, with no involvement in the GTT classification, reviewed the same discharge ICD-10 codes in 30 new random medical records. The agreement on patients having a complication or not during the hospital stay and number of complications was tested.

In the second phase, concurrent validity³¹ was studied, comparing complications using the two different methods: GTT (reference standard) and ICD-10 complication codes. Validation here refers to agreement in identifying complications in the same admissions using the two different methods³².

Statistical analysis

Sample size calculations were based on the assumption that 14 per cent of the study population would acquire a complication in hospital according to ICD-10 codes, based on available evidence^{28,30}. Because patient record review is expected to reveal more complications¹², it was further assumed that, if an ICD-10 complication code were attributed to an admission, the risk of identifying a complication according to the GTT (patient harm of category E, F, G, H, I) would be twice the risk had no such code been present. Based on these assumptions, to obtain 90 per cent power and a significance level of 5 per cent, inclusion of at least 636 patient admissions was required.

A Venn diagram was used to illustrate associations between surgical complications identified by ICD-10 codes and GTT reviews. Cohen’s κ and weighted κ statistics were used to test reliability, with assessment of the strength of agreement among the ICD-10 code reviewers and between the GTT teams by means of inter-rater reliability tests³³. Standard classification of κ coefficient values was used: less than 0.20, poor agreement; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, good; and 0.81–1.00, very good³³.

Logistic regression was used to analyse the relationship between complications identified using a verified ICD-10 code compared with complications identified by the GTT review of patients’ records; the results are reported as odds ratios (ORs) with 95 per cent confidence intervals. $P \leq 0.050$ was considered statistically significant. Data were analysed using SPSS® version 24 for Windows® (IBM, Armonk, New York, USA). Weighted κ analysis was performed using Stata® version 14.0, and Venn diagrams were drawn using the Stata procedure *pvenn* (StataCorp, College Station, Texas, USA).

Results

A study sample of 700 surgical admissions in 695 patients was drawn randomly from a larger group of 12 966 surgical procedures. Some 87.4 per cent were from the tertiary hospital and 12.6 per cent from the community hospital. Surgical procedures in the community hospital included gastrointestinal surgery (such as appendectomy and colonic resection) and urology (for example prostatectomy and ureteric stent). Those in the tertiary hospital included neurosurgery (such as disc herniation surgery, excision of intracranial lesion, evacuation of haematoma, external drainage), gynaecology (hysterectomy, oophorectomy, vaginal fistula repair, perineorrhaphy), orthopaedics (osteosynthesis or reposition of fractured limbs, hip or knee replacements, external fixation, malleolus surgery) and thoracic surgery (ascending aorta vascular prosthesis,

Table 1 Characteristics of 700 surgical patient admissions in two hospitals in western Norway from November 2012 to March 2015

	No. of patients (n = 700)
Age (years)	
18–64	417 (59.6)
≥ 65	283 (40.4)
Sex	
M	309 (44.1)
F	391 (55.9)
Duration of hospital stay (days)	
1	72 (10.3)
2–7	350 (50.0)
8–14	199 (28.4)
≥ 15	79 (11.3)
Incision time (min)	
≤ 30	83 (11.9)
31–60	125 (17.9)
61–180	392 (56.0)
≥ 181	100 (14.3)
ASA fitness grade	
I	115 (16.4)
II	305 (43.6)
III	249 (35.6)
IV	30 (4.3)
V	1 (0.1)
Urgency of surgery	
Elective	395 (56.4)
Emergency	305 (43.6)
Surgical specialty	
Neurosurgery	129 (18.4)
Orthopaedics	223 (31.9)
Gynaecology	111 (15.9)
Thoracic	149 (21.3)
General	88 (12.6)
Hospital type	
Tertiary	612 (87.4)
Central	88 (12.6)

Values in parentheses are percentages.

cardiopulmonary bypass, aortic valve replacement, circulatory anastomosis). Patient characteristics are shown in *Table 1*. Mean(s.d.) age was 58.3(18.1) (range 18–99) years. In total, the data set represented 5350 days of admission, with a median of 5.8 (i.q.r. 3.1–8.8) and mean(s.d.) of 7.6(8.3) days per stay.

Complications detected by the Global Trigger Tool method

Using the GTT method, a total of 331 (range 1–7) complications were identified in 212 of 700 admissions (30.3 per cent). Seventy-seven admissions were identified with more than one complication describing an injury. The distribution of the GTT complications is shown in *Table 2*. A majority were classified as temporary: E in 111 of 331 (33.5 per cent) and F in 200 (60.4 per cent). Thirteen (4.0 per cent) were regarded as representing permanent harm and classified as G. None were classified as H (life-supporting treatment needed) and complications in seven patients (2.1

Table 2 Complications classified according to the Global Trigger Tool in 23 categories for the 212 of 700 patient admissions with patient harm in two hospitals in western Norway from November 2012 to March 2015

	One or more GTT complications*
Other surgical complications†	86 (26.0)
Surgical-site infection	35 (10.6)
Urinary tract infection	34 (10.3)
Low respiratory infection	30 (9.1)
Other infection	26 (7.9)
Postoperative	24 (7.3)
bleeding/haematoma	
Postoperative respiratory complication	23 (6.9)
Reoperation	20 (6.0)
Ventilator-associated pneumonia	10 (3.0)
Organ failure	10 (3.0)
Medication-related (including blood and fluid therapy)	9 (2.7)
Deteriorating chronic condition	6 (1.8)
Bleeding	5 (1.5)
Thrombosis/emboli	3 (0.9)
Decubitus	2 (0.6)
Other	2 (0.6)
Allergy	1 (0.3)
Fracture	1 (0.3)
Central venous line infection	1 (0.3)
Medical technical equipment failure	1 (0.3)
Postpartum/obstetric complication	1 (0.3)
Wrong surgical site	1 (0.3)
Fall	0 (0)
Total no. of complications	331 (100)

Values in parentheses are percentage of total number of complications.

*Among 212 patient admissions. †Drop foot, rupture of dura, pleural fluid, necrosis, vision disturbances, infarction, atrial fibrillation, other. GTT, Global Trigger Tool.

per cent) were classified as I (death). Infection-related complications constituted 41.1 per cent and 26.0 per cent were classified as other surgical complications.

ICD-10 complication code classification

Electronic extraction of ICD-10 codes identified 519 complication codes in 332 patient records of the 700 admissions (complication rate 47.4 per cent). After excluding codes representing complications already present on admission, 141 of 700 admissions (20.1 per cent) with a total of 298 complications were found to occur in hospital. The number of complications per hospital stay ranged from one to six. The distribution of the ICD-10 complication codes is summarized in *Table 3*. After verifying the complications, the order of frequency of complication types changed from cardiac, fall, respiratory and infections to cardiac, respiratory, infections and other. Of note, all 96 codes for patient falls

Table 3 Distribution of complications in 332 surgical admissions identified using ICD-10 complication codes, and distribution of verified complications in 141 surgical admissions from patients' records in two western Norwegian hospitals from November 2012 to March 2015

	Extracted ICD-10 codes (n = 332 admissions)	Verified ICD-10 codes (n = 141 admissions)
Respiratory	79 (15.2)	55 (18.5)
Pneumonia	21 (4.0)	20 (6.7)
Respiratory, other	58 (11.2)	35 (11.7)
Cardiac	151 (29.1)	95 (31.9)
Cardiac arrhythmia	65 (12.5)	49 (16.4)
Congestive heart failure	17 (3.3)	11 (3.7)
Cardiac, other	69 (13.3)	35 (11.7)
Infections	65 (12.5)	47 (15.8)
Sepsis	13 (2.5)	9 (3.0)
Surgical site	20 (3.9)	13 (4.4)
Urinary tract	24 (4.6)	20 (6.7)
Infections, other	8 (1.5)	5 (1.7)
Surgical wound rupture	5 (1.0)	4 (1.3)
Nervous system	13 (2.5)	11 (3.7)
Delirium, somnolence, other	3 (0.6)	2 (0.7)
Cerebral infarction	10 (1.9)	9 (3.0)
Bleeding	17 (3.3)	15 (5.0)
Embolism	5 (1.0)	2 (0.7)
Nutrition	28 (5.4)	23 (7.7)
Malnutrition, other nutritional deficiencies	12 (2.3)	11 (3.7)
Other disorders of fluid, electrolyte and acid-base balance	16 (3.1)	12 (4.0)
Anaesthesia	3 (0.6)	3 (1.0)
Mechanical implantation	16 (3.1)	7 (2.3)
Fall	96 (18.5)	0 (0)
Other complications	41 (7.9)	36 (12.1)
Total no. of complications	519 (100)	298 (100)

Values in parentheses are percentage of total number of complications. Detailed list of included ICD-10 complication codes can be found in *Table S1* (supporting information).

were found to represent falls occurring before, and not during, the hospital stay.

Reliability analysis

Analysis of agreement in classifying complications in 20 random medical records using the GTT method revealed that the two teams reached 85 per cent agreement in terms of the presence of a complication, 65 per cent regarding numbers of complications and 75 per cent on the levels of severity. The κ values for inter-rating agreement between the teams were 0.700, 0.504 (weighted) and 0.688 (weighted) respectively. Three clinical researchers reviewed the same discharge ICD-10 codes in 30 random medical records. Agreement was 91 per cent in terms of patients having a complication or not during the hospital stay, and 77 per cent for agreement on actual number of complications. Accordingly, the κ values for inter-rater reliability were 0.816 and 0.731 respectively.

Validating complications by ICD-10 versus Global Trigger Tool

To investigate concurrent validity, it was determined whether admissions with ICD-10 complications were the same admissions as those identified as having one or more complications by the GTT methodology. The similarity between the two classification methods increased from 68.3 per cent before clinical verification of the ICD-10 complication codes to 83.3 per cent after excluding ICD-10 codes representing complications already present on admission (*Fig. 1*).

Logistic regression was used to quantify the importance of clinically verifying ICD-10 complication codes rather

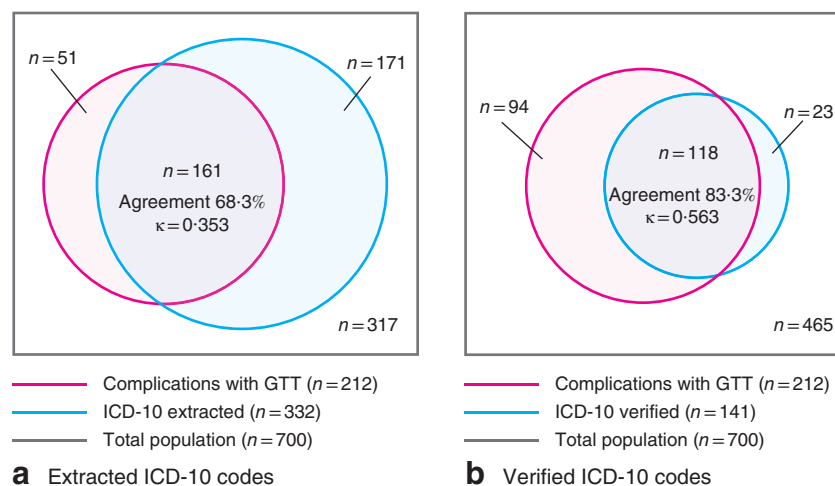


Fig. 1 Agreement between methods of identifying admissions with complications versus no complications: **a** using ICD-10 codes extracted from administrative data and **b** using ICD-10 codes verified from patients' records. GTT, Global Trigger Tool

than using them without verification. Admissions with unverified ICD-10 codes (332) were at increased odds of also having a GTT-identified complication (OR 5.85, 95 per cent confidence interval 4.06 to 8.44), whereas admissions with verified ICD-10 codes (141) increased the odds substantially (OR 25.38, 15.41 to 41.79). Ninety-four admissions with complications according to GTT methodology did not have an ICD-10 code reflecting a complication (*Fig. 1*).

Discussion

This study found that complications during the hospital stay were overestimated when crude ICD-10 codes were used in surgical admissions. By excluding codes representing conditions already present on admission, the complication rate decreased from 47.4 to 20.1 per cent. This provides quantifiable evidence of the detrimental impact of coding practices on the ability of ICD-10 codes to indicate a true complication in patient care. Based on the present findings, it does not appear feasible to detect and disclose all complications and level of severity using a single method. A substantial decrease in complications was found with accurate ICD-10-verified complication codes compared with ICD-10 codes present on admission. These findings support the hypothesis of the study. The GTT method is designed to inform about local complications and patient safety initiatives over longer periods of time¹³, whereas the ICD-10 (if used accurately) may be used both locally and in large epidemiological studies to inform on larger patient safety interventions.

The complication rate obtained using the GTT in the present study was 30.3 per cent of all admissions. This is at the upper end of the range reported in studies included in a recent systematic review². That review, however, included studies across both medical and surgical specialties. Focusing solely on surgical patient populations, as in the present study, would be expected to result in higher rates than in mixed patient populations⁵. Regarding level of severity, the majority of complications identified by the GTT (93.9 per cent) were found to be associated with temporary harm. Similar findings regarding severity have been documented elsewhere^{34,35}.

In the present study, the agreement between the ICD-10 and GTT methods increased from 68.3 to 83.3 per cent following clinical researchers' verification of the ICD coding. Other studies^{7,15,36} have investigated complications using different detection methods. The high rates of agreement here might be explained by avoidance of use of complications reported voluntarily by healthcare personnel as a comparator. There is evidence for under-reporting

of complications in voluntary reporting systems¹², which would likely lead to lower agreement between methods. The present analysis included a large number of complication codes (154 in total), which might have increased the number of complications identified, thus offering a broader perspective on surgical complication analyses. Moreover, a large number of clinically reviewed patient records were included, which is likely to have increased the number of complications found and analysed by this methodology compared with smaller studies³⁵.

A total of 94 admissions with GTT-identified complications were not identified by ICD-10 codes. There may be several reasons for this discrepancy. In a busy clinical practice, physicians may fail to use correct ICD-10 codes owing to lack of training in the use of such codes and/or time constraints, as pointed out in a national report³⁷. The finding also demonstrates differences in methodology between the two systems for identifying complications. The GTT method may include complications before admission if they are linked to medical treatment¹³, whereas the ICD-10 codes should consider only complications that emerge in hospital to be 'true' complications. The present findings have significant practical implications. If hospitals are to work on preventing or addressing patient safety risks, reliable knowledge of risk factors will be needed. Deriving such knowledge and developing patient safety programmes based solely on administratively collected complication data does not represent an effective strategy, based on the present findings. More accurate evidence concerning in-hospital complications is needed to tailor surgical patient safety interventions. Examples from this study suggest that a focus on respiratory and cardiac complications, infections and nutrition is needed. It was also shown here that all patient falls occurred before admission. These findings are important as ICD-10 coding is widely used to report on complications, carry out research, and to inform healthcare policies and hospital funding¹⁷. Yet few studies have reported similar procedures for clinical verification of ICD-10-coded patient-level data³⁰. Such studies are urgently required to inform decision-making and funding. On a practical level, an electronic 'flag' built into ICD-10 classification systems can be recommended, so that the coder can identify a 'complication' already present on admission. Such a flagging option is available in the USA, Canada and Australia³⁸. This improves coding accuracy without the requirement for significant financial investment or training, thereby enhancing the value of inexpensive complication reports based on routinely collected data.

Prospective recording of complications on a severity scale, using a validated system such as the Clavien–Dindo classification⁷, would be ideal. This would probably lead to

the availability of more accurate and clinician-reported data in prospective databases of postoperative morbidity, which could offer a better picture of surgical care quality. However, this would have training and resource implications if introduced as standard practice, and this is not currently done routinely in Norwegian hospitals.

The present study has limitations. Only surgical patients were included, so the results cannot be extrapolated directly to the larger cohort of medical admissions. Second, a standard Norwegian version of the GTT protocol was used and not a trigger protocol especially designed for surgical patients, known as the Surgical Trigger Toolkit. This was because the expert GTT teams had already been trained to use the standard version; in addition, there is no validated Norwegian version of the Surgical Trigger Toolkit available. However, the GTT actually covers all but two of the trigger words available in the Surgical Trigger Toolkit and hence the coverage is very similar. Third, the preventability of the identified complications was not investigated. Classifying preventability is not included as part of national GTT team training in Norway, nor is it recommended as a part of the GTT protocol¹³. Further research should analyse preventability in a similarly structured manner^{2,39}. Furthermore, when studying in-hospital complications, those related to previous admissions had to be excluded. This may have led to under-reporting of complications, mainly owing to coding practices being related to each hospital admission and not to each patient throughout the healthcare pathway. Finally, as a result of natural differences between the ICD-10 and GTT systems, it may be questioned whether admissions identified by both methods actually had the same (type of) complications. Simply put, although an admission might have been identified as complicated by both tools, the type of complication identified by one of the two systems may have differed from that identified by the other. This would not affect overall complication rates, but could affect the types of complication found and consequently the hospital's targets for improvement.

The study also has strengths, including: bringing together two methods for assessing surgical safety; the overall high level of expertise among the reviewers; the inclusion of two separate hospitals; and the good reliability of the analyses. Regarding reliability, the inter-rater reliability analysis is a methodological strength. The GTT teams showed good agreement for detection and severity of complications, and moderate agreement regarding the number of complications present. The two GTT teams had expert members from both hospitals (with knowledge of local reporting practices). The inter-rater agreement among the ICD-10 reviewers was even stronger. This is a

prerequisite for studies reporting data that require clinical judgement and the seniority of the reviewers ensured this.

The accuracy of ICD-10 complication codes is improved when in-hospital complications are verified with record reviews. Crude data with unverified ICD-10 codes significantly overestimate surgical complications within hospitals because complications present on admission are included. This can represent a severe bias for national and international comparisons of quality and safety of surgical care.

Acknowledgements

The authors thank the hospitals' management and staff for making the data collection possible; the professional GTT teams at Haukeland University Hospital (J. Veim, J. K. Kleiva and G. Gran) and Førde Health Trust (K. Furevik and W. B. Sjøstad) for the classification work; S. Harthug for constructive feedback; H. Waldeland, T.-L. Thorsen and N. E. J. Widnes for electronic data collection; and R. Kufner Lein at the University of Bergen Library for support with systematic searching and reference management.

A.S. was supported by a research grant from the Western Norway Regional Health Authority Trust, the Patient Safety Programme (grant number HV1173) and the Norwegian Nurses Organization (grant number 15/0023). A.S.H. was supported by a postdoctoral fellow research grant from the Western Norway Regional Health Authority Trust, the Patient Safety Programme (grant number HV1172). The research carried out by N.S. is supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South London at King's College Hospital NHS Foundation Trust. N.S. is director of King's Improvement Science, which is part of the NIHR Collaboration for Leadership in Applied Health Research and Care South London, and comprises a specialist team of improvement scientists and senior researchers based at King's College London. Its work is funded by King's Health Partners (Guy's and St Thomas' NHS Foundation Trust, King's College Hospital NHS Foundation Trust, King's College London and South London and Maudsley NHS Foundation Trust), Guy's and St Thomas' Charity, the Maudsley Charity and the Health Foundation. No funding source had any role in the design and conduct of the study; collection, management, analysis or interpretation of the data; or preparation, review or approval of the manuscript. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR or the Department of Health. N.S. is the director of London Training & Safety Solutions, which delivers

team assessment and training to hospitals on a consultancy basis.

Discourse: The authors declare no other conflict of interest.

References

- Kohn LT, Corrigan JM, Donaldson MS (eds). *To Err is Human. Building a Safer Healthcare System*; 1999. http://www.pearsoneducation.nl/wps/media/objects/13902/14236351/H%2007_To%20Err%20Is%20Human.pdf [accessed 2 January 2018].
- Hibbert PD, Molloy CJ, Hooper TD, Wiles LK, Runciman WB, Lachman P et al. The application of the Global Trigger Tool: a systematic review. *Int J Qual Health Care* 2016; **28**: 640–649.
- Gawande AA, Thomas EJ, Zinner MJ, Brennan TA. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. *Surgery* 1999; **126**: 66–75.
- Zegers M, de Bruijne MC, de Keizer B, Merten H, Groenewegen PP, van der Wal G et al. The incidence, root-causes, and outcomes of adverse events in surgical units: implication for potential prevention strategies. *Patient Saf Surg* 2011; **5**: 13.
- de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boormeester MA. The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care* 2008; **17**: 216–223.
- Olsen S, Neale G, Schwab K, Psaila B, Patel T, Chapman EJ et al. Hospital staff should use more than one method to detect adverse events and potential adverse events: incident reporting, pharmacist surveillance and local real-time record review may all have a place. *Qual Saf Health Care* 2007; **16**: 40–44.
- Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 2004; **240**: 205–213.
- Donaldson LJ, Panesar SS, Darzi A. Patient-safety-related hospital deaths in England: thematic analysis of incidents reported to a national database, 2010–2012. *PLoS Med* 2014; **11**: e1001667.
- Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA et al. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med* 1991; **324**: 377–384.
- Classen DC, Lloyd RC, Provost L, Griffin FA, Resar R. Development and evaluation of the Institute for Healthcare Improvement Global Trigger Tool. *J Patient Saf* 2008; **4**: 169–177.
- Noble DJ, Panesar SS, Pronovost PJ. A public health approach to patient safety reporting systems is urgently needed. *J Patient Saf* 2011; **7**: 109–112.
- Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N et al. ‘Global trigger tool’ shows that adverse events in hospitals may be ten times greater than previously measured. *Health Aff (Millwood)* 2011; **30**: 581–589.
- Griffen F, Resar R. *Global Trigger Tool for Measuring Adverse Events (Second Edition)*; 2009. <http://www.ihl.org/resources/Pages/IHIWhitePapers/IHIGlobalTriggerToolWhitePaper.aspx> [accessed 2 January 2018].
- Sharek PJ, Parry G, Goldmann D, Bones K, Hackbarth A, Resar R et al. Performance characteristics of a methodology to quantify adverse events over time in hospitalized patients. *Health Serv Res* 2011; **46**: 654–678.
- Naessens JM, Campbell CR, Huddleston JM, Berg BP, Lefante JJ, Williams AR et al. A comparison of hospital adverse events identified by three widely used detection methods. *Int J Qual Health Care* 2009; **21**: 301–307.
- Raleigh VS, Cooper J, Bremner SA, Scobie S. Patient safety indicators for England from hospital administrative data: case-control analysis and comparison with US data. *BMJ* 2008; **337**: a1702.
- WHO. *History of the ICD*. <http://www.who.int/classifications/icd/en/> [accessed 2 January 2018].
- OECD. *Health at a Glance 2013: OECD Indicators*; 2013. http://www.oecd-ilibrary.org/social-issues-migration-health/health-at-a-glance-2013_health_glance-2013-en [accessed 2 January 2018].
- Severinsen MT, Kristensen SR, Overvad K, Dethlefsen C, Tjønneland A, Johnsen SP. Venous thromboembolism discharge diagnoses in the Danish National Patient Registry should be used with caution. *J Clin Epidemiol* 2010; **63**: 223–228.
- Sedova P, Brown RD Jr, Zvolosky M, Kadlecova P, Bryndziar T, Volny O et al. Validation of stroke diagnosis in the National Registry of Hospitalized Patients in the Czech Republic. *J Stroke Cerebrovasc Dis* 2015; **24**: 2032–2038.
- Hall R, Mondor L, Porter J, Fang J, Kapral MK. Accuracy of administrative data for the coding of acute stroke and TIAs. *Can J Neurol Sci* 2016; **43**: 765–773.
- Aardal S, Berge K, Breivik K, Flaatten HK. Medical records, DRG and intensive care patients. *Tidsskr Nor Laegeforen* 2005; **125**: 903–906.
- Barber C, Lacaille D, Fortin PR. Systematic review of validation studies of the use of administrative data to identify serious infections. *Arthritis Care Res (Hoboken)* 2013; **65**: 1343–1357.
- McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of myocardial infarction diagnoses in administrative databases: a systematic review. *PLoS One* 2014; **9**: e92286.
- Hartwig SC, Denger SD, Schneider PJ. Severity-indexed, incident report-based medication error-reporting program. *Am J Hosp Pharm* 1991; **48**: 2611–2616.
- Kristensen S, Mainz J, Bartels P. *A Patient Safety Vocabulary Safety Improvement for Patients in Europe: SimPatIE – Work Package 4*; 2007. http://www.zdravstvo-kvaliteta.org/attachments/article/18/Patient_safety_indicator_development [accessed 1 August 2018].
- American College of Surgeons. *American College of Surgeons’ National Surgical Quality Improvement Program*. <http://www.facs.org/cqi/outcomes.html> [accessed 3 January 2018].

- 28 de Vries EN, Prins HA, Crolla RM, den Outer AJ, van Andel G, van Helden SH *et al.*; SURPASS Collaborative Group. Effect of a comprehensive surgical safety system on patient outcomes. *N Engl J Med* 2010; **363**: 1928–1937.
- 29 Haynes AB, Weiser TG, Berry WR, Lipsitz SR, Breizat AH, Dellinger EP *et al.*; Safe Surgery Saves Lives Study Group. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med* 2009; **360**: 491–499.
- 30 Haugen AS, Sjøfteland E, Almeland SK, Sevdalis N, Vonen B, Eide GE *et al.* Effect of the World Health Organization checklist on patient outcomes: a stepped wedge cluster randomized controlled trial. *Ann Surg* 2015; **261**: 821–828.
- 31 Lin WL, Yao G. Concurrent validity. In *Encyclopedia of Quality of Life and Well-Being Research*, Michalos AC (ed.). Springer Netherlands: Dordrecht, 2014; 1184–1185.
- 32 Streiner DL, Norman GR. *Validity* (4th edn). Oxford University Press: Oxford, 2008.
- 33 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–174.
- 34 Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal trends in rates of patient harm resulting from medical care. *N Engl J Med* 2010; **363**: 2124–2134.
- 35 Mevik K, Griffin FA, Hansen TE, Deilkås ET, Vonen B. 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. *BMJ Open* 2016; **6**: e010700.
- 36 Naessens JM, Campbell CR, Berg B, Williams AR, Culbertson R. Impact of diagnosis-timing indicators on measures of safety, comorbidity, and case mix groupings from administrative data sources. *Med Care* 2007; **45**: 781–788.
- 37 Office of the Auditor General of Norway. *Office of the Auditor General's Investigation of Medical Coding Practice within the Health Enterprises. Document 3:5 (2016–2017)*. <https://www.riksrevisjonen.no/en/Reports/Documents/CodingPracticeHealthEnterprises.pdf> [accessed 2 January 2018].
- 38 Jackson TJ, Michel JL, Roberts R, Shephard J, Cheng D, Rust J *et al.* Development of a validation algorithm for 'present on admission' flagging. *BMC Med Inform Decis Mak* 2009; **9**: 48.
- 39 Flaatten H, Brattebø G, Alme B, Berge K, Rosland JH, Viste A *et al.* Adverse events and in-hospital mortality: an analysis of all deaths in a Norwegian health trust during 2011. *BMC Health Serv Res* 2017; **17**: 465.

Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of the article.