

BMJ Open Validation of adverse events after hip arthroplasty: a Swedish multi-centre cohort study

Martin Magnéli,^{1,2} Maria Unbeck,^{3,4} Cecilia Rogmark,⁵ Ola Rolfson,⁶ Ami Hommel,^{7,8} Bodil Samuelsson,¹ Kristina Schildmeijer,⁹ Desirée Sjöstrand,¹⁰ Max Gordon,^{1,2} Olof Sköldenberg^{1,2}

To cite: Magnéli M, Unbeck M, Rogmark C, *et al.* Validation of adverse events after hip arthroplasty: a Swedish multi-centre cohort study. *BMJ Open* 2019;**9**:e023773. doi:10.1136/bmjopen-2018-023773

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-023773>).

Received 23 April 2018

Revised 16 January 2019

Accepted 18 January 2019

ABSTRACT

Objectives Preventing adverse events (AEs) after orthopaedic surgery is a field with great room for improvement. A Swedish instrument for measuring AEs after hip arthroplasty based on administrative data from the national patient register is used by both the Swedish Hip Arthroplasty Register and the Swedish Association of Local Authorities and Regions. It has never been validated and its accuracy is unknown. The aim of this study was to validate the instrument's ability to detect AEs, and to calculate the incidence of AEs following primary hip arthroplasties.

Design Retrospective cohort study using retrospective record review with Global Trigger Tool methodology in combination with register data.

Setting 24 different hospitals in four major regions of Sweden.

Participants 2000 patients with either total or hemi-hip arthroplasty were recruited from the SHAR. We included both acute and elective patients.

Primary and secondary outcome measures The sensitivity and specificity of the instrument. Adjusted cumulative incidence and incidence rate.

Results The sensitivity for all identified AEs was 5.7% (95% CI: 4.9% to 6.7%) for 30 days and 14.8% (95% CI: 8.2 to 24.3) for 90 days, and the specificity was 95.2% (95% CI: 93.5% to 96.6%) for 30 days and 92.1% (95% CI: 89.9% to 93.8%) for 90 days. The adjusted cumulative incidence for all AEs was 28.4% (95% CI: 25.0% to 32.3%) for 30 days and 29.5% (95% CI: 26.0% to 33.8%) for 90 days. The incidence rate was 0.43 AEs per person-month (95% CI: 0.39 to 0.47).

Conclusions The AE incidence was high, and most AEs occurred within the first 30 days. The instrument sensitivity for AEs was very low for both 30 and 90 days, but the specificity was high for both 30 and 90 days. The studied instrument is insufficient for valid measurements of AEs after hip arthroplasty.

BACKGROUND

Adverse events (AEs) following surgery are a major challenge in the field of orthopaedics. Hip arthroplasty is one of the most successful procedures in modern medicine, and the

Strengths and limitations of this study

- The use of one of the most sensitive method for identifying AEs (retrospective record review with Global Trigger Tool methodology).
- The multi-centre study design, which includes a large sample size comprising both acute and elective patients.
- The use of the Swedish personal number in combination with the national register ensured that no admissions were missed.
- Our results are only generalisable to healthcare systems where International Classification of Disease codes are used to measure AEs.

technical improvements since Charnley arthroplasty have been minor.¹

Preventing AEs is a field with great room for improvement. Complication rates after hip arthroplasty are between 3.4% and 27%.²⁻⁴ However, comparison of AE rates should be done with caution.⁵ Two reasons for this is (1) there are no globally accepted definitions of AEs after hip arthroplasty⁶ and (2) there are many different methods for identifying AEs, which complicate comparisons.⁷

The method that has been proven to be most sensitive compared with others is retrospective record review (RRR) by trained reviewers.⁸⁻¹⁰ Another method for identifying and measuring AEs is by using administrative data and International Classification of Diseases (ICD) codes.¹¹

The Swedish Hip Arthroplasty Register (SHAR) issues a yearly report that includes the AE rate after hip arthroplasty.¹² This AE rate is generated from an instrument that uses administrative data with a set of selected AE ICD-10 codes (see online supplementary table A1), which are found in the Swedish National Patient Register (NPR).¹³ Thus this report is not based on SHAR data but on NPR data, and the same instrument is used by the



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Martin Magnéli;
martin.magneli@sil.se

Swedish Association of Local Authorities and Regions in a public accessible web application named Healthcare in Numbers (HIN).¹⁴ The major difference about HIN and SHAR concerns the definition of the population. HIN is based on NPR procedure codes and SHAR is based on hospitals recording of interventions into the register.

The instrument only uses codes that are registered during discharge from readmissions. AEs that occur during the index admission are not included.

Despite this widespread usage, we know nothing of its sensitivity and specificity. While NPR's primary ICD-codes are known to be accurate (but with some variation between diagnoses),¹⁵ we do not know the accuracy for secondary codes. We also do not know how well this set of codes and their selection are suited for detecting AEs.

The aim of this study was to validate the instrument's ability to detect AEs, and to calculate the incidence of AEs following primary hip arthroplasties.

METHODS

Study design

This is a retrospective multi-centre cohort study on prospectively collected data from medical records and register data from SHAR and NPR.

Study size

The calculated sample size was estimated to be 2000 patients, assuming 5%–10% inconclusive records, using an alpha level of 0.05 and a power minimum of 80%. The main assumptions regarding the instrument's rate of failure to register a correct ICD-10 code for an AE was set to 15% (the sensitivity), and the rate for incorrectly coded non-event was set to 5% (the specificity).

Setting

The study comprises hip arthroplasty patients from four major county councils in Sweden (Stockholm, Skåne, Västra Götaland and Västerbotten) in 24 different hospitals (six university hospitals, five central county council hospitals, seven county council hospitals and six private hospitals who have agreements/contracts with the county councils, one private hospital treats both acute and elective patients. Patients underwent surgery between January 2009 and December 2011.

Participants

All patients 18 years of age or older whose data were recorded in the SHAR for either a hemi or total hip arthroplasty were eligible for inclusion. Both acute surgery for hip fractures and elective surgery for degenerative joint disease were included.

To increase the probability of selecting medical records with an AE and avoiding excess RRR on records without AEs, we used a weighted sample. Twenty different selection groups for acute and elective arthroplasties were created as follows (see online supplementary table A2).

1. We constructed three groups with lengths of primary stay in percentiles divided as 0%–50%, 51%–80% and 81%–100%. The three groups were further divided based on whether there was an ICD-10 code indicating an AE in the NPR (see online supplementary table A3). Overall, six groups were generated.
2. A selection was made for patients who had readmissions in the NPR. The readmission groups were divided in readmission within 2–30 days and within 31–90 days after surgery. The two groups were further divided based on whether there was an ICD-10 code indicating an AE in the NPR, generating a total of four groups.

This created a total of ten selection groups and we sampled according to the table (see online supplementary table A2) both from acute and elective patients yielding a total of 20 groups.

Patient and public involvement

This is a register and record-based retrospective study with no patient involvement.

Data sources

From the SHAR we collected data on the primary procedures that were cross-linked with data from the NPR, using the Swedish personal identity numbers. From the NPR, we collected data on all admissions from the primary procedure and 90 days post-operatively. With the NPR data, we could create a timeline with all admissions for each patient. This timeline was used as a template to know which admissions to review with the RRR. The NPR data also contained ICD codes that were used in the validation of the instrument. Death data that was used in the validation of the instrument were available from the national death register (NDR). Medical records were obtained as paper copies or were reviewed on location at the hospital.

Review teams and the RRR method

The review team consisted of 10 reviewers with a record review experience ranging from novice to expert (see online supplementary table A4). The more experienced reviewers performed both stage one and two of the review. All reviewers received obligatory 1 day training by two of the senior researchers (MG and MU).

We used the Swedish adaptation of the Global Trigger Tool (GTT),¹⁶ named marker-based record review,¹⁷ as the RRR method for collecting AE data. A study-specific manual was created and included definitions, inclusion criteria, exclusion criteria and all alterations and clarifications from the GTT.

Definitions

An AE was defined as suffering, physical harm or disease as well as death related to the index admission and as a condition that was not an inevitable consequence of the patient's disease or treatment.

Based on the terminology in the Swedish Patient Safety Act,¹⁸ a preventable AE was defined as an event that could

have been prevented if adequate actions had been taken during the patient's contact with healthcare.

The index admission was defined as the orthopaedic admission when the patient had hip arthroplasty surgery. If the patient was discharged directly to a geriatric or rehabilitation clinic, this admission was also considered to be a part of the index admission.

AEs related to acts of either omission or commission were included.

Inclusion and exclusion criteria

We included and performed RRR on all inpatient care and all unplanned outpatient care in all Swedish hospitals from the index admission date up to 90 days after surgery. We included AEs that occurred during index admission and AEs that occurred during readmissions that originated from the index admission. AEs that were identified during unplanned outpatient visits at a hospital (accidents and emergencies visits) and originated from the index admission were also included.

We excluded AEs that were unrelated to the index admission and AEs that originated from the care of another AE. For example, if a patient was admitted because of a periprosthetic joint infection and sustained a fracture from falling in the ward, the infection was included as an AE, and the fracture was not included. We did not include planned outpatient visits at hospitals or planned or unplanned outpatient visits outside of hospitals, such as with a general practitioner.

Review process

The GTT consisted of a two-stage review process.

Review stage 1

All medical records, including notes from different professionals, were reviewed. The reviewers screened the record, searching for any of the 38 pre-defined triggers that indicated a potential AE. The triggers were divided into five modules: general triggers (n=18), laboratory triggers (n=5), surgical triggers (n=7), medication triggers (n=3) and intensive care triggers (n=5) (see online supplementary table A5).

A summary of the RRR and all identified triggers with a free text description of the trigger/event were documented in a database (Microsoft Access 2007). All records with a potential AE went forward to review stage 2.

Review stage 2

All identified triggers deemed as positive for a potential AE were assessed in stage 2.

Each potential AE was then assessed if it was caused by the healthcare service using a 4-point Likert scale graded as follows: (1) the AE was not caused by the index admission, (2) the AE was probably not caused by the index admission, (3) the AE was probably caused by the index admission and (4) the AE was caused by the index admission.

AEs graded as 1 or 2 were excluded and AEs graded as 3 or four were included, and the reviewer made a full assessment that included evaluations of preventability,

type of AE (71 different types in 15 different categories), severity and whether or not the AE was ICD-10 coded.

Preventability was assessed using a similar 4-point Likert scale as follows: (1) the AE was not preventable, (2) the AE was probably not preventable, (3) the AE was probably preventable and (4) the AE was preventable. AEs that were graded 3 or 4 were classified as preventable.

The severity of the AEs was evaluated using a slightly modified version of the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) index.¹⁹ NCC MERP index categories E–I were included, and the categories indicated the following: (E) contributed to or resulted in temporary harm, (F) contributed to or resulted in temporary harm that required outpatient or inpatient care or prolonged hospitalisation, (G) contributed to or resulted in permanent harm, (H) required intervention necessary to sustain life within 60 min and (I) contributed or resulted in the patient's death.

Reliability and validity

Inter-rater reliability was evaluated through the double review of 6% of the records to assess the agreement between the primary reviewers' judgements concerning whether at least one trigger or potential AE was identified in the record, whether the record was to be forwarded to secondary review, whether the reviewer identified the same specific event and whether this event was a potential AE.

The review process was monitored by an RRR expert (MU) who also was available for questions from the reviewers. The completeness and adherence to the study manual in stages 1 and 2 were monitored closely. All questions or discrepancies were given as written feedback to the reviewers for resolution. If needed, clarifying discussions were held with the respective reviewer.

Validation

The instrument is based on a set of 13 specific ICD codes and one code category (I-codes: diseases of the circulatory system) defining AEs (see online supplementary table A1). Five of the specific codes and the code category has to be used as the primary diagnosis and the remaining eight can be either as primary or secondary code. In the validation of the instrument, test was defined as positive for an AE if the patient had:

1. Any of these code criteria in any readmission within 90 days after surgery (data source=NPR). Or
2. A death date after discharge from the primary admission and within 90 days after surgery (data source=NDR).

We used the results from the RRR as gold standard when we performed the sensitivity and specificity analysis. To give a nuanced study of the performance of the instrument, we divided the AEs found with RRR into four categories.

1. All AEs (all found AEs with causality Likert scale ≥ 3).
2. Preventable AEs (all AEs with preventability Likert scale ≥ 3).
3. Major AEs (preventable AEs with NCC MERP $\geq F$).

4. Selected AEs (AEs types that correspond to the set of 'AE' ICD-codes).

We did two different validations for the four AE categories:

1. AEs found (with RRR) during both index and readmissions versus the instrument (only readmissions).
2. AEs found (with RRR) during only readmissions versus the instrument.

We performed the two separate validations for all AE categories for all patients and with the subsets of acute and elective patients. The rationale for the multiple validations was to test different nuances of the instrument.

Statistical methods

Adjusted sensitivity and specificity were calculated for both 30 and 90 days. The sensitivity and specificity were calculated in each sample group and multiplied by the group proportion (population group/total population). The products of all groups were summed, and the result was the adjusted sensitivity and specificity for the population.

The adjusted cumulative incidence for 30 and 90 days was calculated by dividing the number of patients with an AE in each group with the group sample size, generating a rate for that group. This rate was multiplied by the group proportion (population group/total population). The products of all 10 groups were summed to provide the adjusted cumulative incidence. The same method was used to calculate the adjusted cumulative incidence of preventable AEs and serious AEs.

We used the selection group tables for acute and elective patients separated for the analysis of sensitivity and specificity for acute and elective patients and the two tables pooled together for the analysis of all patients.

The incidence rate was calculated by taking the total sum of the identified AEs within 30 days after surgery for each selection group and dividing it with the sample group size and then multiplying it with the group proportion. The sum was the incidence rate in AEs/person-month.

Cohen's kappa was calculated for inter-rater reliability between the primary reviewers.²⁰ Bootstrap samples (n=2000) were used to calculate the 95% CIs.

We used R (V. 3.5.2) and packages dplyr, boot, irr, htmlTable and Gmisc.

RESULTS

Participants

The study population consisted of 21 774 patients. We included 2000 patients weighted according to the selection group table (see online supplementary table A2). Two patients were excluded. The first patient had no available medical record, a short primary admission, no readmissions and was unlikely to have sustained an AE. The second patient had a hip fracture treated with internal fixation, with an assumingly faulty registration in the SHAR. After exclusion, 1998 patients with a total of 5422 inpatient admissions and outpatient visits in 69 hospitals were reviewed and included in the analysis (figure 1).

The study cohort comprised of 667 acute hip fracture patients and 1331 elective patients, and 63% of the

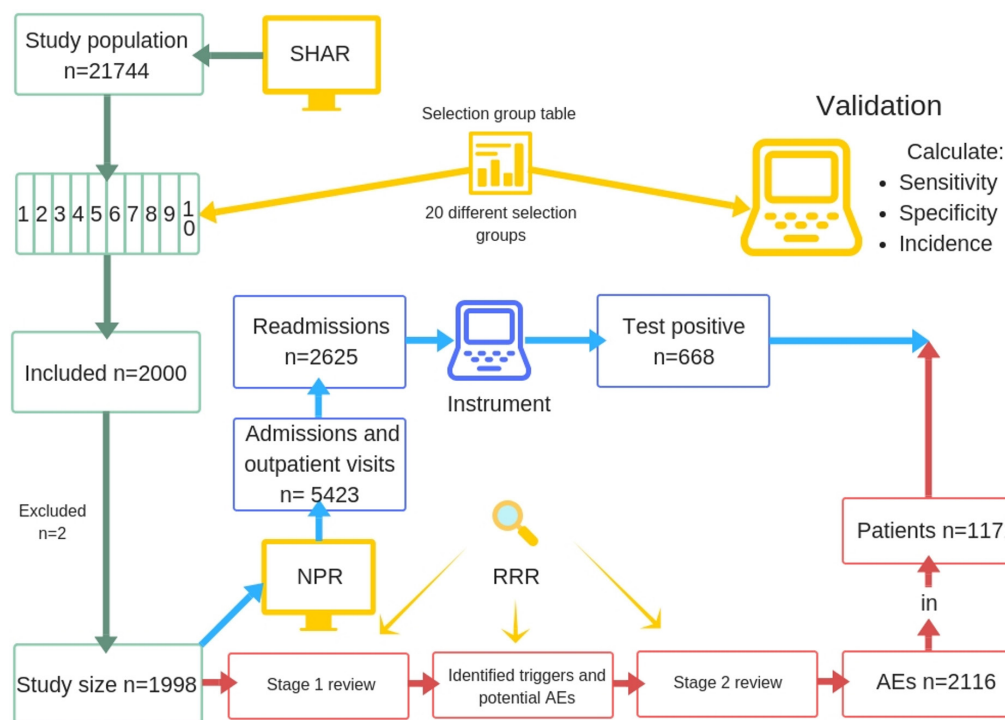


Figure 1 Flowchart of the study process. AEs, adverse events; NPR, National Patient Register; SHAR, Swedish Hip Arthroplasty Register; RRR, retrospective record review.

Table 1 Demographics

	Total n=1998	Acute n=667	Elective n=1331
Female n=	1250 (62.6%)	444 (66.6%)	806 (60.6%)
Male n=	748 (37.4%)	223 (33.4%)	525 (39.4%)
Age, median (IQR)	77.0 (68.0–84.0)	84.0 (79.0–89.0)	73.0 (64.0–80.0)
LOS, median (IQR)	7.0 (4.0–13.0)	14.0 (9.0–20.0)	5.0 (4.0–8.0)
Type of Hospital n=			
University	630 (31.5%)	295 (44.2%)	335 (25.2%)
Central county council	556 (27.8%)	180 (27.0%)	376 (28.2%)
County council	531 (26.6%)	109 (16.3%)	422 (31.7%)
Private	281 (14.1%)	83 (12.4%)	198 (14.9%)

patients were women. The hip fracture group comprised more women, had older patients and involved a longer length of stay during the index admission (table 1).

Identified AEs and rate of ICD-10 codes

In total, we found 2116 AEs in 1171 (58.6%) patients. Of these, 1605 AEs (75.9%) in 975 (48.8%) patients were classified as preventable AEs, 1066 AEs (50.4%) in 744 (37.2%) patients were classified as major AEs and 1206 (57.0%) in 829 (41.5%) patients were classified as selected AEs. The 667 acute patients sustained 981 (46.4%) of these and the elective patients sustained 1135 (53.6%). The acute patients sustained 758 (47.3%) of the preventable AEs and 431 (40.4%) of the major AEs.

Of the 2116 found AEs, an ICD-10 code for the AE was found in 1145 (54.1%) records, in 879 (54.8%) of the 1605 preventable AEs, in 787 (71.1%) of the 1066 major AEs and in 758 (62.9%) of the 1206 selected AEs.

The majority of AEs occurred during the index admission (n=1260, 59.5%), and 443 (35.2%) of them had an ICD-10 code. The number of AEs that occurred during readmission within 30 days after surgery was 590 (27.9%) and 476 (80.7%) had an ICD-10 code. The number of AEs that occurred during readmission within 90 days after surgery was 856 (40.5%), and 702 (82.0%) had an ICD-10 code.

The group of AEs that had the highest rate of ICD-10 codes was thrombosis and embolism, at 91.6%. AEs related to the surgical procedure, such as dislocation, had the second highest rate (76.1%), and bleeding that did not occur during the operation had the third highest rate (75.7%). The group of AEs that had the lowest rate of codes was pressure ulcers (5.3%), followed by skin and superficial vessel damage (6.3%) and neurological AEs (14.6%) (table 2).

The single AE type that had the highest rate of available ICD codes were acute myocardial infarction and stroke with 100% available codes, followed by the next top four, which were dislocation (98.5%), periprosthetic joint infection (96.0%), pulmonary embolism (95.3%) and fracture caused by falling (90.2%). Ten different individual types of AEs were not coded at all (see online supplementary table A6).

Adjusted cumulative incidence and incidence rate

The adjusted cumulative incidence for patients sustaining at least one AE was 28.4% for 30 days and 29.5% for 90 days (table 3). The acute patients had a higher incidence than the elective patients with 51.4% compared with 17.2% for 30 days and 52.1% compared with 18.6% for 90 days. The incidence of preventable AEs and major AEs

Table 2 Groups of adverse event types and frequency of ICD codes

	Available ICD code		Total n=	Rate %
	Yes n=	No n=		
Thrombosis or embolus	106	11	117	90.6
AEs related to the surgical procedure	353	111	464	76.1
Dislocation of prosthesis*	(270)	(4)	(274)	(98.5)
Tissue damage*	(11)	(7)	(18)	(61.1)
Bleeding, reoperation*	(3)	(2)	(5)	(60.0)
Bleeding, no reoperation*	(47)	(62)	(109)	(43.1)
Other AEs related to the surgical procedure*	(22)	(36)	(58)	(37.9)
Bleeding (not related to surgery)	28	9	37	75.7
Iatrogenic infections	430	228	658	65.3
Falls	53	30	83	63.9
Other AEs	112	134	246	45.5
Abnormal pain	9	19	28	32.1
Allergic reaction	8	19	27	29.6
Distended bladder	19	63	82	23.2
AE cause by anaesthesia	2	7	9	22.2
Neurological AEs	7	41	48	14.6
Skin and superficial vessel AEs	8	119	127	6.3
Pressure ulcer	10	180	190	5.3
Total n=	1145	971	2116	54.1%

ICD-10, the 10th revision of the International Classification of Diseases.

*Sub-group, numbers in brackets are not included in total.

Table 3 Adjusted cumulative incidence of adverse events (AEs)

	All patients	Acute patients	Elective patients
All AEs			
Incidence 30 days	28.4 (25.0 to 32.3)	51.4 (44.0 to 59.5)	17.2 (14.0 to 21.1)
Incidence 90 days	29.5 (26.0 to 33.8)	52.1 (45.0 to 60.2)	18.6 (15 to 22.7)
Preventable AEs			
Incidence 30 days	22.2 (19.0 to 25.6)	40.6 (35 to 47.2)	13.9 (11.0 to 17.5)
Incidence 90 days	23.4 (20.0 to 26.8)	41.1 (36.0 to 48.1)	15.3 (12.0 to 19.2)
Major AEs			
Incidence 30 days	13.4 (11.0 to 15.6)	21.4 (18.0 to 25.7)	10.1 (8.0 to 13.1)
Incidence 90 days	14.7 (12.0 to 17.2)	22.1 (19.0 to 26.2)	11.6 (9.0 to 14.9)

All results are in %, 95% CI in brackets.

were also higher for the acute patients compared with the elective, both for 30 and 90 days.

The incidence rate for all AEs was 0.43 AEs per person-month (95% CI: 0.39 to 0.47). For preventable AEs, the incidence rate was 0.32 (95% CI: 0.29 to 0.35), and for major AEs, the incidence rate was 0.22 (0.20 to 0.25).

Adjusted sensitivity and specificity

Adjusted sensitivity and specificity for all AEs were 5.7% and 95.2%, respectively, at 30 days, and 14.8% and 92.1%, respectively, at 90 days (table 4). This was the comparison that used the widest definition of AEs that were found from surgery until 90 days post-operatively. The sensitivity and specificity for the narrowest definition of AE that only compared readmissions were 3.0% and 93.5%, respectively, at 30 days, 26.6% and 90.5%, respectively, at 90 days.

The acute patients had higher sensitivity but lower specificity compared with the elective patients, for all classes of AEs, for both 30 and 90 days.

Inter-rater reliability

The inter-rater reliability values of the primary reviewers' judgements concerning whether at least one trigger or potential AE was identified in the record were $\kappa=0.828$ and 0.965, respectively. The inter-rater reliability for whether the record was to be forwarded to secondary review was $\kappa=0.965$. The inter-rater reliability values for the identification of a specific event or whether that event was a potential AE were $\kappa=0.65$ and 0.873, respectively.

DISCUSSION

In this retrospective multi-centre cohort study using RRR on 1998 patients who had undergone hip arthroplasty surgery, we validated an instrument based on ICD-codes from NPR. We found a high incidence for AEs and more than every fourth patient sustained an AE. The incidence was higher for the acute patients and every other acute patient sustained an AE, compared with almost every fifth elective patient. Almost two-thirds of the AEs occurred

during the index admission and the difference between AEs within 30 days and 90 days was below two percentage.

We found a low overall rate of coded AEs for all and preventable AEs (55%) and a higher rate for major AEs (73%).

We validated different nuances of the instrument and found that sensitivity was low, and at best every fourth patient with an AE is detected. We found that for all different nuances the specificity was high with the best result of 97%. Maas *et al*²¹ compared ICD-codes with record review and also found low sensitivity and high specificity. When we compared found AEs (with RRR) during readmissions to the instrument the sensitivity was lower for all AEs within 30 days. This was due to the fewer total number of true positives and their distribution in fewer selection groups for the readmissions versus instrument.

The definition of AEs in this study is wide and can be considered as excessive by some. The rationale behind the choice of GTT as the method for identifying AEs was not that we wanted the instrument to fail or to imply that hip arthroplasty is a dangerous procedure. When we decided to do a record review validation, we wanted to use the method that has proven to identify the most AEs to ensure that we had the highest quality data possible. The range of severity of the found AEs is wide and it is easier to remove irrelevant AEs from a data-set than the opposite.

As expected, our definition and method for measuring AEs yielded higher rates than for example Huddleston *et al*,³ who used data abstraction from Medicare records and found a 30 day AE rate of 5.8% after total hip arthroplasty. Studies on AEs in mixed orthopaedic patients using the GTT have shown rates of 15%–30%.^{22 23}

The preventability can be a hard to assess in RRR. To ensure concordant assessments some AEs, as falls, prosthetic dislocation and pressure ulcers were always classed as preventable in the study. The combination of our inclusive definition of preventability and structured RRR might be an explanation that the rate of preventable AEs in elective patients were more than double than Jorgensen *et al*²⁴

Table 4 Validation of the instrument

	All patients		Acute patients		Elective patients	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
All AEs						
Standard 30*	5.7 (4.9 to 6.7)	95.2 (93.5 to 96.6)	11.5 (9.4 to 13.8)	89 (82.0 to 93.4)	3.2 (2.9 to 3.5)	97.4 (97.1 to 97.7)
Standard 90†	14.8 (8.2 to 24.3)	92.1 (89.9 to 93.8)	20.6 (15.6 to 28.1)	82.6 (75.0 to 87.3)	15.5 (4.5 to 37.3)	95.1 (93.1 to 96.2)
Readmission 30†	4.2 (2.9 to 5.5)	93.8 (92.3 to 94.9)	4.5 (4.2 to 4.8)	86.6 (83 to 89.5)	3.1 (2.3 to 3.9)	96.8 (96.6 to 97)
Readmission 90†	21.6 (6.4 to 64.4)	90.4 (88.4 to 92.0)	12.9 (8.2 to 13.3)	80.1 (76.0 to 83.6)	20.8 (3.8 to 68.4)	94.7 (92.8 to 95.7)
Preventable AEs						
Standard 30*	5.9 (5 to 7)	94.9 (93.3 to 96.2)	11.8 (9.5 to 14.4)	89.6 (84.4 to 93.4)	3.1 (2.9 to 3.6)	97.3 (96.9 to 97.6)
Standard 90*	18.0 (8.6 to 31.4)	91.9 (89.9 to 93.6)	23.2 (15.8 to 36.1)	83.5 (77.7 to 87.7)	17.7 (4.6 to 69.0)	95.1 (93.1 to 96.2)
Readmission 30†	3.0 (2.9 to 3.1)	93.9 (92.4 to 95.1)	4.5 (4.2 to 4.9)	86.9 (83.4 to 89.9)	2.4 (2.3 to 2.5)	96.9 (96.7 to 97.2)
Readmission 90†	22.1 (5.1 to 65.5)	90.5 (88.4 to 92.2)	12.9 (8.0 to 13.4)	80.1 (76.1 to 83.5)	19.9 (3.8 to 57.3)	94.8 (92.9 to 95.8)
Major AEs						
Standard 30*	6.2 (5 to 7.8)	94.8 (93.2 to 96)	12.8 (9.2 to 17.5)	89.1 (85.0 to 92.3)	3.2 (2.9 to 3.6)	97.3 (96.9 to 97.6)
Standard 90*	18.9 (8.1 to 39.1)	91.2 (89 to 92.9)	18.2 (14.4 to 23.4)	81.9 (77.2 to 85.8)	21.2 (4.7 to 68.4)	95.1 (93.2 to 96.1)
Readmission 30†	3.0 (2.9 to 3.1)	93.9 (92.4 to 95.1)	4.5 (4.2 to 5)	87.0 (83.6 to 90.0)	2.4 (2.3 to 2.5)	96.9 (96.7 to 97.2)
Readmission 90†	19.6 (5.1 to 65.5)	90.5 (88.4 to 92.2)	8.4 (7.7 to 9)	80.1 (76.0 to 83.6)	19.9 (3.8 to 67.9)	87.2 (92.9 to 95.8)
Selected AEs						
Standard 30*	6.5 (5.3 to 8.1)	95.1 (93.4 to 96.2)	13.0 (10.1 to 16.2)	90.0 (85 to 93.7)	3.3 (3 to 3.6)	97.2 (96.9 to 97.5)
Standard 90*	14.3 (8.8 to 25.3)	91.6 (89.2 to 93.3)	19.5 (16.3 to 23.1)	83.4 (77.5 to 88.0)	17.9 (4.7 to 48.2)	95.0 (93.2 to 96.1)
Readmission 30†	3.0 (2.8 to 3.1)	93.9 (92.5 to 95.0)	4.5 (4.2 to 5.2)	87.0 (83.5 to 89.8)	2.3 (2.3 to 2.5)	96.9 (96.6 to 97.1)
Readmission 90†	26.6 (6.5 to 65.6)	90.5 (88.5 to 92.1)	13.3 (8.7 to 14.2)	80.4 (75.8 to 84.0)	26.2 (3.9 to 69.4)	87.3 (92.9 to 95.7)

95% CI in brackets.

*AEs found (with RRR) during index and readmissions versus instrument (readmissions only).

†AEs found (with RRR) during readmissions versus instrument (readmissions only).

AEs, adverse events.

found in their study on total knee and hip replacements. However, our incidence of preventability is in accordance with another national GTT study in orthopaedic care.²³

The use of administrative data for measuring AEs after orthopaedic surgery has been studied by Sebastien *et al.*²⁵ The authors compared the Agency for Healthcare Quality and Research's Patient Safety Indicators, an ICD code-based instrument, with the Agency for Healthcare Quality and Research National Surgical Quality Improvement Programme (ACS-NSQIP), a system that uses trained surgical clinical reviewers and well-defined criteria to identify AEs. In their study on mixed orthopaedic patients, the AHQR-PSI revealed an AE rate of 1%, and the ACS-NSQIP revealed an AE rate of 22%. The authors concluded that the instruments were unable to adequately assess AEs in orthopaedic surgery. Best *et al.*²⁶ compared the ACS-NSQIP with administrative data for AEs after surgery and found similar results to this study, a sensitivity of more than 50% in only 23% of the selected AEs. Classen *et al.*⁹ also compared the AHQR-PSI with the GTT and found that the AHQR-PSI fared very poorly.

The examined instrument is used to compare the quality of care in different Swedish hospitals, and this is one of the quality indicators that determines economic reimbursement to the hospitals. With regards to the low sensitivity to detect AEs, their validity is questionable. The instrument algorithm is also used by the Healthcare in Numbers, and by the Swedish Knee Arthroplasty register to measure AEs following total knee arthroplasty.²⁷ The use of the ICD-instrument for knee arthroplasties have not yet been validated, but our results from the elective hip patients implies that the use of the instrument might be questionable.

The low overall rate of correct ICD-10 codes in only half of the cases is the largest obstacle for using administrative data with ICD-10 codes for measuring all AEs after hip arthroplasty. Furthermore, we found that the majority of the AEs, including one-fifth of the dislocations, occurred during the index admission, so excluding the index admissions in an instrument will decrease the sensitivity.

Strengths and limitations

To our knowledge, this is the largest study on AEs after hip arthroplasty that uses RRR and the only study that includes both acute hip fracture patients and elective surgery patients, thereby including both total and hemi hip arthroplasties. The study contains a large study population and a multicentre design with a wide range of patients of all ages and types of hospitals. The 90-day follow-up is long enough to detect all acute and subacute AEs. The Swedish personal identity numbers and the NPR enabled us to review admissions, and this, in combination with the RRR method, decreased the risk of missing an AE to approximately zero and resulted in high-quality data on the AEs. All kappa values were classified as near-perfect agreement except for one that was classified as good agreement, indicating the good quality of the RRR.

The study period of 90 days after surgery in this study makes this analysis a study on short-term AEs and does not address late-onset AEs, such as aseptic loosening, one of the most common causes of revision surgery.²⁸ The baseline data on the patients are from the registers, and information on patient factors, such as comorbidities and physical status, is lacking. Therefore, this study cannot identify risk factors for AEs. In addition, our results are only generalisable to healthcare systems where ICD codes are used to measure AEs. The weighted sample did not include type of hospital and we can therefore not calculate incidence for the different types of hospitals.

CONCLUSION

The conclusions from this study are that the incidence of AEs after hip arthroplasty is high and that the tested instrument cannot measure this correctly. Furthermore, because of the low reliability of the ICD-10 codes, an improved instrument needs to be based on robust variables, possibly in combination with ICD-10 codes, and also include AEs identified during index admission and a wider range of AE types.

Author affiliations

¹Department of Clinical Sciences at Danderyd Hospital, Division of Orthopaedics, Karolinska Institutet, Stockholm, Sweden

²Department of Orthopaedics, Danderyd University Hospital Corp., Stockholm, Sweden

³Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden

⁴Trauma and Reparative Medicine Theme, Karolinska University Hospital, Stockholm, Sweden

⁵Skanes universitetssjukhus, Malmö, Sweden

⁶Department of Orthopaedics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

⁷Department of Care Science, Faculty of Health and Society, Malmö University, Malmö, Sweden

⁸Department of Orthopaedics, Skåne University Hospital, Lund, Sweden

⁹School of Health and Caring Sciences, Linnaeus University, Kalmar, Sweden

¹⁰Education Unit, Skånevård Kryh, Kristianstad, Sweden

Acknowledgements The authors thank Marie Ax, Susanne Hansson, Ammar Jobory, Zara Hedlund, Mirta Stupin, Tim Hansson, Lovisa Hult-Ericson and Christina Jansson for valuable help in carrying out the study. We would also like to thank all department managers for access to the medical records and Per Nydert for help with the study database. Finally we would like to thank Christoffer C Jørgensen, who performed a splendid review of the manuscript, which improved this article tremendously.

Contributors MM collected, analysed and interpreted the data and contributed to the drafting of the work. MU contributed to the design of the study, collected data, contributed to the drafting of the work and revised the manuscript critically. CR contributed to the design of the study and the drafting of the work and revised the manuscript critically. OR contributed to the design of the study and to critically revising the work. AH collected data and critically revised the work. BS collected data and critically revised the work. KS collected data and critically revised the work. DS collected data and critically revised the work. MG contributed to the design of the study; collected, analysed and interpreted the data; contributed to the drafting of the work; and revised the manuscript critically. OS contributed to the design of the study, collected data, contributed to the drafting of the work and revised the manuscript critically. All authors have approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Funding This study was funded by institutional grants from the Karolinska Institutet, Department of Clinical Sciences, Danderyd Hospital, from the regional agreement on medical training and clinical research (ALF) between Stockholm

County Council and Karolinska Institutet, and from LÖF, the Swedish patient insurance programme. The grant providers were not involved in any part of the study, in the writing of the manuscript or in the decision to submit the manuscript for publication.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Ethical approval was provided by the Regional Ethics Committee of Gothenburg (516-13 and T732-13). Permission for data access for the reviewers was granted by the head of each respective unit. The patients did not provide an informed consent to the record review.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

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

REFERENCES

1. Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. *Lancet* 2007;370:1508–19.
2. Wolf BR, Lu X, Li Y, et al. Adverse outcomes in hip arthroplasty: long-term trends. *J Bone Joint Surg Am* 2012;94:e103–8.
3. Huddleston JI, Wang Y, Uquillas C, et al. Age and obesity are risk factors for adverse events after total hip arthroplasty. *Clin Orthop Relat Res* 2012;470:490–6.
4. Millstone DB, Perruccio AV, Badley EM, et al. Factors Associated with Adverse Events in Inpatient Elective Spine, Knee, and Hip Orthopaedic Surgery. *J Bone Joint Surg* 2017;99:1365–72.
5. Grober ED, Bohnen JM. Defining medical error. *Can J Surg* 2005;48:39–44.
6. Healy WL, Iorio R, Clair AJ, et al. Complications of total hip arthroplasty: standardized list, definitions, and stratification developed by the hip society. *Clin Orthop Relat Res* 2016;474:357–64.
7. Thomas EJ, Petersen LA. Measuring errors and adverse events in health care. *J Gen Intern Med* 2003;18:61–7.
8. Unbeck M, Muren O, Lillkrona U. Identification of adverse events at an orthopedics department in Sweden. *Acta Orthop* 2008;79:396–403.
9. 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* 2011;30:581–9.
10. Naessens JM, Campbell CR, Huddleston JM, et al. A comparison of hospital adverse events identified by three widely used detection methods. *Int J Qual Health Care* 2009;21:301–7.
11. World Health Organization. *ICD-ICD-10-CM- International Classification of Diseases, Tenth Revision, Clinical Modification*. Geneva: World Health Organization, 2017. Available: <https://www.cdc.gov/nchs/icd/icd10cm.htm> (cited 2018 Apr 18).
12. Kärrholm J, Mohaddes M, Odin D, et al. *Swedish Hip Arthroplasty Register, Annual report 2017*. Gothenburg: Sweden, 2018. Available: <https://registercentrum.blob.core.windows.net/shpr/r/-rsrapport-2017-S1xKMzsAwX.pdf> (cited 2018 Sep 26).
13. Swedish National Board of Health and Welfare. Patientregistret. <https://www.socialstyrelsen.se/register/halsodataregister/patientregistret/inenglish> (Accessed 29 Aug 2018).
14. Önskad händelser efter knä- och total höftprotosoperation. *Vården i siffror - Statistik för kvalitets- och förbättringsarbete*. Stockholm: Sveriges Kommuner och Landsting, 2018. (In Eng: Adverse events after knee and hip arthroplasty. Stockholm: Swedish Association of Local Authorities and Regions; 2018). Available: <https://www.vardenisiffror.se/indikator?datefrom=2013-01-01&metadatameasure=74903f74-ac82-4d5b-8320-2a409df71255&relatedmeasuresbyentry=keyword&relatedmeasuresbyid=ortopedi&units=se> (cited 2018 Sep 26).
15. Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011;11:450.
16. Griffin F, Resar R. *IHI Global Trigger Tool for Measuring Adverse Events*. 2nd edn. Cambridge, MA: Institute for Healthcare Improvement, 2009. (IHI Innovation Series white paper).
17. Markörbaserad Journalgranskning. *För att identifiera och mäta skador i vården*. Stockholm: Sveriges Kommuner och Landsting, 2012. (In Eng: Marker based record review to identify and measure harm in healthcare. Stockholm: Swedish Association of Local Authorities and Regions; 2012). Available: <https://webbutik.skl.se/bilder/artiklar/pdf/7164-847-1.pdf> (cited 2018 Sept 26).
18. *Patientsäkerhetslag (2010:659)*. Stockholm: Socialdepartementet, 2010. (In Eng: Patient safety act (2010:659). Stockholm: Swedish Parliament; 2010). Available: https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/patientsakerhetslag-2010659_sfs-2010-659 (cited 2018 Sept 2018).
19. National Coordinating Council for Medication Error, Reporting and Prevention. NCC MERP Index for Categorizing Medication Errors. Available: <http://www.nccmerp.org/types-medicationerrors/> (cited 2018 sept 26).
20. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46.
21. Maass C, Kuske S, Lessing C, et al. Are administrative data valid when measuring patient safety in hospitals? A comparison of data collection methods using a chart review and administrative data. *Int J Qual Health Care* 2015;27:305–13.
22. Unbeck M, Schildmeijer K, Henriksson P, et al. Is detection of adverse events affected by record review methodology? an evaluation of the “Harvard Medical Practice Study” method and the “Global Trigger Tool.”. *Patient Saf Surg* 2013;7:10.
23. Rutberg H, Borgstedt-Risberg M, Gustafson P, et al. Adverse events in orthopedic care identified via the Global Trigger Tool in Sweden – implications on preventable prolonged hospitalizations. *Patient Saf Surg* 2016;10:23.
24. Jørgensen CC, Petersen MA, Kehlet H. Lundbeck Foundation Centre for Fast-Track Hip and Knee Replacement Collaborative Group. Preoperative prediction of potentially preventable morbidity after fast-track hip and knee arthroplasty: a detailed descriptive cohort study. *BMJ Open* 2016;6:e009813.
25. Sebastian AS, Polites SF, Glasgow AE, et al. Current quality measurement tools are insufficient to assess complications in orthopedic surgery. *J Hand Surg Am* 2017;42:10–15.
26. Best WR, Khuri SF, Phelan M, et al. Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: results from the department of veterans affairs national surgical quality improvement Program. *J Am Coll Surg* 2002;194:257–66.
27. Robertsson O, W-Dahl A, Lidgren L. *Annual Report 2017, Swedish Knee Arthroplasty Register*. Lund: Swedish Knee Arthroplasty Register, 2017. Available: http://www.myknee.se/pdf/SVK_2017_Eng_1.0.pdf (cited 2018 Sept 26).
28. Melvin JS, Karthikeyan T, Cope R, et al. Early failures in total hip arthroplasty – a changing paradigm. *J Arthroplasty* 2014;29:1285–8.