BMJ Open

Validation of adverse events after hip arthroplasty: a multicentre cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023773
Article Type:	Research
Date Submitted by the Author:	23-Apr-2018
Complete List of Authors:	Magnéli, Martin; Karolinska Institutet, Department of Clinical Sciences, Danderyd Hospital, Division of Orthopaedics Unbeck, Maria; Karolinska Institutet Department of Molecular Medicine and Surgery Rogmark, Cecilia; Skanes universitetssjukhus Malmo Rolfson , Ola ; Institute of Clinical Sciences, Department of Orthopaedics, Sahlgrenska Academy, University of Gothenburg Hommel, Ami; Lunds Universitet, Samuelsson, Bodil; Karolinska Institutet, Department of Clinical Sciences at Danderyd Hospital, Division of Orthopaedics Schildmeijer, Kristina; School of Health and Caring Sciences, Linnaéus University Sjöstrand, Desirée; Skånevård Kryh, 11Education Unit Gordon, Max; Karolinska Institutet, Department of Clinical Sciences at Danderyd Hospital, Division of Orthopaedics Sköldenberg, Olof; Karolinska Institutet, Department of Clinical Sciences at Danderyd Hospital, Division of Orthopaedics
Keywords:	Adverse events < THERAPEUTICS, ORTHOPAEDIC & TRAUMA SURGERY, Hip < ORTHOPAEDIC & TRAUMA SURGERY

Validation of adverse events after hip arthroplasty: a multicentre cohort study

Martin Magnéli[1,2], Maria Unbeck[3,4], Cecilia Rogmark[5,6], Ola Rolfson[6,7], Ami Hommel[8,9], Bodil Samuelsson[2], Kristina Schildmeijer[10], Desirée Sjöstrand[11], Max Gordon[1,2], Olof Sköldenberg[1,2]

[1]Department of Orthopaedics, Danderyd University Hospital, SE-182 88 Stockholm, Sweden

[2]Department of Clinical Sciences, Danderyd Hospital, Division of Orthopaedics, Karolinska Institutet, SE-182 88 Stockholm, Sweden

[3]Theme Trauma and Reparative Medicine, Karolinska University Hospital, SE-171 79 Stockholm, Sweden

[4]Department of Molecular Medicine and Surgery, Karolinska Institutet, SE-171 77 Stockholm, Sweden

[5]Skane University Hospital Malmö, Lund University, SE-205 02 Malmö, Sweden

[6]Swedish Hip Arthroplasty Register, Centre of Registers Västra Götaland, SE-413 45 Gothenburg, Sweden

[7]Sahlgrenska Academy, SE-405 30 Gothenburg, Sweden

[8]Skane University Hospital, Department of Orthopaedics, Sweden, Malmö University

[9]Swedish Hip Fracture Register, SE-221 85 Lund, Sweden

[10]School of Health and Caring Sciences, Faculty of Health and Life Sciences, Linnaeus University, S-391 82 Kalmar, Sweden

[11]Education Unit, Skånevård Kryh, Region Skåne, SE-291 89 Kristianstad, Sweden

Corresponding author: Martin Magnéli, Department of Orthopaedics, Danderyd Hospital Mörbygårdsvägen 88, SE-182 88 Stockholm, Sweden, E-mail: martin.magneli@sll.se, Telephone: +46(0) 737 886 557

Word count: 3968

Number of figures: 1

Number of tables: 4

Number of references: 24

ABSTRACT

Objectives

Preventing adverse events (AEs) after orthopaedic surgery is a field with great room for improvement. There are two major instruments for measuring AEs after hip arthroplasty surgery. Both are based on administrative data, and neither is validated. The aim of this study is to validate the sensitivity and specificity of the two instruments and to calculate the adjusted cumulative incidence and incidence rate of AEs following hip arthroplasty.

Design

Retrospective cohort study using retrospective record review (RRR) (Global Trigger Tool) in combination with register data.

Setting

Twenty-four different hospitals in 4 major regions of Sweden.

Participants

Two thousand patients with either total or hemi hip arthroplasty were recruited from the Swedish Hip Arthroplasty Register. We included acute patients with hip fractures and elective patients with degenerative joint disease.

Primary and secondary outcome measures

The sensitivity and specificity of the instruments. Adjusted cumulative incidence and incidence rate.

Results

The sensitivity for all identified AEs was 6% (95% CI: 5 – 7) for 30 days and 15% (95% CI: 8 – 25) for 90 days, and the specificity was 95% (95% CI: 94 – 97 for 30 days and 92% (95% CI: 90 – 94) for 90 days. The adjusted cumulative incidence for all AEs was 28% (95% CI: 25 – 33) for

30 days and 30% (95% CI: 26 – 34) for 90 days. The incidence rate was 0.43 AEs per person-month (95% CI: 0.39 – 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 instruments are insufficient for valid measurements of AEs after hip arthroplasty.

ARTICLE SUMMARY

Strengths and limitations of this study

- The use of retrospective record review and the Global Trigger Tool for data collection is the method that identifies the most adverse events (AEs).
- This is a multicentre study that 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.

KEYWORDS

Orthopaedics, Adverse events, Hip arthroplasty, Validation, Global Trigger Tool

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 technical improvements since Charnley arthroplasty have been minor.[1]

Complication rates after hip arthroplasty are between 3.4% – 57%[2–5] and preventing AEs is a field with great room for improvement. AEs after hip arthroplasty, including dislocation or periprosthetic joint infection, are typically connected to the implant, but patients also sustain more general AEs, such as pulmonary embolism and pneumonia.[6, 7]

There are many different methods for identifying AEs. The method that has been proven to be most sensitive compared to 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]

Sweden has a rich history of quality registers and today, there are 96 publicly founded quality registers in Sweden. The Swedish Hip Arthroplasty Register (SHAR) issues a yearly report that includes the AE rate after hip arthroplasty. The AE rate is generated from administrative data with selected ICD-10 codes indicating AEs. The codes are found in the Swedish National Patient Register (NPR).[12] The SHAR instrument only uses codes that are registered during discharge from readmissions. AEs that occur during the index admission are not included. The instrument algorithm used by SHAR is also used by the Swedish Association of Local Authorities and Regions in a public access web application named the healthcare in numbers (HIN). The SHAR and HIN are both used for comparing the performances of different caregivers in Sweden. None of these instruments have been validated, and their sensitivity and specificity are unknown.

The aim of this study was to validate the SHAR and HIN by studying their sensitivity and specificity, and to calculate the adjusted cumulative incidence and incidence rate of AEs following hip arthroplasty in patients with both planned and acute surgery.

METHODS

Study design

This is a retrospective multicentre 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 HIN and SHAR´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 Västra Norrland) 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). 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. Ten different selection groups were created as follows.

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. 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.

The weighted samples, the ten selection groups and the ICD-10 codes are shown in Table A.1, Appendix.

Patient involvement

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

Data sources

Data on the primary procedures were received from the SHAR and cross-linked with data from the NPR, using the Swedish personal identity numbers. The coverage for total hip arthroplasties and for hemi arthroplasties in the SHAR was 98.3% and 97.5%, respectively, in 2015.[13]

All admissions and readmissions were received from cross-linking SHAR data with data from NPR. Medical records were obtained as paper copies or were reviewed on location at the hospital. Death data were available from the national death register.

Review teams and the RRR method

The review team consisted of ten reviewers (registered nurses, medical students and physicians) with a record review experience ranging from novice to expert. All reviewers received obligatory one-day training by two of the senior researchers (MG and MU).

We used the Swedish adaptation of the Global Trigger Tool (GTT),[14] named Markörbaserad journalgranskning,[15] as the RRR method for collecting all 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,[16] 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.

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.

The 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 predefined triggers that indicated a potential AE. The triggers were divided into 5 modules: general triggers (n=18), laboratory triggers (n=5), surgical triggers (n=7), medication triggers (n=3) and intensive care triggers (n=5).

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 4 were included, and the reviewer made a full assessment that included evaluations of preventability, 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 AEs.

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.[17] 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 hospitalization, G) contributed to or resulted in permanent harm, H) required intervention necessary to sustain life within 60 minutes and I) contributed or resulted in the patient´s death.

Reliability and validity

Inter-rater reliability was evaluated through the double review of 6 percent of the records to assess 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

To validate the SHAR and HIN, we used data from the NPR as test data that were validated with the AE data from the from the RRR. The SHAR and HIN use a set of selected

ICD-10 codes as definitions of AEs (Table A.2, Appendix). All admissions from the NPR that had one of the selected primary codes as the main diagnosis or one of the selected general codes as the primary or secondary diagnosis, or had a death date after the index admission within 90 days after surgery were considered to be test positive in the sensitivity and specificity analysis. We analysed the sensitivity and We excluded the index admissions (i.e., the admission for the primary surgery) and all admissions over 90 days after surgery, and we only analysed the readmissions.

Statistical methods

Adjusted sensitivity and specificity were calculated for both 30 days and 90 days. We compared the test positive results from the NPR with both the results from all identified AEs from the RRR and from the AEs identified during readmissions. We also compared the results with preventable AEs in the RRR and major AEs (NCC MERP F and above).

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. Bootstrap samples (n=2000) were used to calculate the 95% confidence intervals.

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 ten 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. Bootstrap samples (n=3000 for all AEs and preventable AEs and n=1500 for serious AEs) were used to calculate the 95% confidence intervals.

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.[18]

We used R (v 3.4.2) and packages dplyr, boot, irr and Gmisc.

RESULTS

Participants

The study population consisted of 21 774 patients. We included 2000 patients weighted according to the selection group table (Table A.1, Appendix). 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.

The study cohort comprised 667 acute hip fracture patients and 1331 elective patients, and 63% of the patients were female. The hip fracture group comprised more women, contained older patients, and had a longer length of stay during the index admission. Demographic baseline data are shown in Table 1.

Table 1 Baseline data on patients

		Hip Fracture	
	Total n=1998	Yes n=667	No n=1331
Age*, median	77 (68 – 84)	84 (79 – 89)	73 (64 – 80)
Female, n (%)	1 250 (62.6)	444 (66.6)	806 (60.6)
Length of stay*, days	7 (4 – 11)	11 (6 – 15)	6 (4 – 8)

*Interquartile range

Identified AEs and rate of correct ICD-10 codes

In total, we found 2116 AEs in 1171 (58.6%) patients. Of these, 1604 AEs (75.8%) in 975 (48.8%) patients were classified as preventable AEs, and 1378 AEs (65.1%) in 935 (46.8%) patients were classified as major AEs (NCC MERP F or higher).

Of the 2116 AEs, a correct ICD-10 code was found in 1145 (54.1%) records, in 879 (54.8%) of the 1604 preventable AEs and in 1004 (72.9%) of the 1378 major AEs.

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

The group of AEs that had the highest rate of correct ICD-10 codes was thrombosis and embolism, at 91.4%. 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 correct coding was pressure ulcers (5.3%), followed by skin and superficial vessel damage (6.3%) and neurological AEs (14.6%). All groups of AEs and their rates of correct ICD coding are displayed in Table 2.

Table 2 Groups of adverse events (AEs) and rates of correct International Classification of Diseases (ICD) codes

AE group	N of correct ICD codes	N of incorrect ICD codes	Total n of ICD codes	% of correct ICD codes
Thrombosis/embolus	106	11	116	91.4%
AEs related to the surgical procedure	353	111	464	76.1%
Haemorrhage, not during surgery	28	9	37	75.7%
Hospital-acquired infections	430	228	658	65.3%
Falls	53	30	83	63.9%
Miscellaneous	55	65	120	45.8%
Deterioration in vital signs	57	69	126	45.2%
Pain	9	19	28	32.1%
Allergic reactions	8	19	27	29.6%
Late detection of urine retention	19	63	82	23.2%
AEs related to anaesthesia care	2	7	9	22.2%
Neurological AEs	7	41	48	14.6%
Skin and superficial vessel AEs	8	119	127	6.3%
Pressure ulcers	10	180	190	5.3%
Total	**1145**	**971**	**2116**	

The single individual AE that had the highest rate of correct coding was acute myocardial infarction with 100% of the ICD-10 codes being correct, followed by the next top four correctly coded AEs, 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 correctly coded at all and had rates of 0%. Among them were AEs caused during anaesthesia care (awareness and laryngospasm), thrombophlebitis, small blood vessel damage, soft tissue damage, muscle weakness, respiratory arrest, genital infection and pressure ulcers of unknown category or category 4.

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. The cumulative incidences for preventable AEs and major AEs are displayed in Table 3.

Table 3 Cumulative incidence (95% confidence interval) for all, preventable and major adverse events (AEs)

Cumulative incidence	All AEs	Preventable AEs	Major AEs
30 days	28.4 (24.7 – 32.5)	22.2 (19.2 – 25.7)	17.5 (15.0 – 20.4)
90 days	29.5 (25.8 – 33.6)	23.4 (20.3 – 27.1)	18.9 (16.2 – 21.9)

The incidence rate for all AEs was 0.43 AEs per person-month (95% CI: 0.39 – 0.47). For preventable AEs, the incidence rate was 0.32 (95% CI: 0.29 – 0.35), and for major AEs, the incidence rate was 0.22 (0.20 – 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. This was the comparison of the test positive readmissions from the NPR data versus all identified AEs in the RRR that included both index admissions and readmissions and represented the true performance of the SHAR and HIN. Table

4 shows the comparisons of these instruments versus all identified preventable and major AEs and of the instruments versus AEs identified only during readmissions.

Table 4 Sensitivity and specificity with 95% confidence intervals for all adverse events (AEs), preventable AEs and major AEs.

All admissions *vs* readmissions	**All AEs**	**Preventable AEs**	**Major AEs**
Sensitivity 30 days	5.7 (5.0 – 6.7)	5.9 (5.0 – 7.0)	6.2 (5.1 – 7.6)
Specificity 30 days	95.2 (93.6 – 96.6)	95.0 (93.3 – 96.2)	95.1 (93.4 – 96.3)
Sensitivity 90 days	14.8 (8.2 – 24.5)	18.0 (8.6 – 30.9)	15.4 (8.0 – 31.2)
Specificity 90 days	92.1 (90.0 – 93.7)	92.0 (90.0 – 93.5)	91.5 (89.5 – 93.3)
Readmissions *vs* readmission	**All AEs**	**Preventable AEs**	**Major AEs**
Sensitivity 30 days	4.2 (2.9 – 5.5)	3.0 (2.9 – 3.1)	3.0 (2.9 – 3.1)
Specificity 30 days	93.8 (92.4 – 94.9)	93.9 (92.4 – 95.1)	93.8 (92.4 – 95.0)
Sensitivity 90 days	21.6 (6.4 – 64.4)	22.1 (5.1 – 65.5)	22.0 (5.1 – 65.5)
Specificity 90 days	90.4 (88.3 – 92.1)	90.5 (88.3 – 92.2)	90.4 (88.3 – 92.0)

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 κ=0.828 and 0.965, respectively. The inter-rater reliability for whether the record was to be forwarded to secondary review was κ=0.965. The inter-rater reliability values for the identification of a specific event or whether that event was a potential AE were κ=0.65 and 0.873, respectively.

DISCUSSION

In this retrospective multicentre cohort study using RRR on 1998 patients who had undergone hip arthroplasty surgery, we validated two instruments based on administrative data for measuring AEs (the SHAR and HIN). We found a low overall rate of correct ICD-10 codes for all and preventable AEs (55%) and a higher rate for major AEs (73%).

The incidence of AEs was high, and more than every fourth patient sustained an AE, more than every fifth patient sustained a preventable AE, and slightly less than every fifth patient sustained a major AE within 30 days. The difference in incidences between 30 and 90 days was below 2% for all AEs, preventable AEs and major AEs. In the validation of the existing instruments, we found the sensitivity to be very low for both 30 and 90 days after surgery.

The specificity was very high for all AEs, preventable AEs and serious AEs, both 30 days and 90 days after surgery. We conclude that the two instruments have a low capacity for detecting AEs but are reliable when an AE has not yet occurred.

Huddleston et al.[3] found a 5.8% rate of AEs after total hip arthroplasty within 30 days. Our study found a much higher adjusted rate, and this was probably because we found more AEs with the RRR method than with the data abstraction from Medicare records. Studies on AEs in mixed orthopaedic patients have shown rates of 15–30%.[19, 20] The use of administrative data for measuring AEs after orthopaedic surgery has been studied by Sebastien et al.[21] The authors compared the Agency for Healthcare Quality and Research´s Patient Safety Indicators (AHRQ-PSI), an ICD code-based instrument, with the Agency for Healthcare Quality and Research National Surgical Quality Improvement Program (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.[22] compared the ACS-NSQIP with administrative data for AEs after surgery and found similar results to those of our study. The authors found that the sensitivity was > 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.

These two examined instruments are used to compare hospitals in Sweden related to the Orthopaedic departments´ quality of care. With regards to their low sensitivity to detect AEs, their validity is 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. The perfect rates (100%) - for coding specific AEs, such as acute myocardial infarction indicate that the method can be used as a powerful instrument for measuring some specific AEs. In our study, 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. A plausible explanation for the low rate of correct ICD-10 codes is that the ICD-10 codes are not only used for medical purposes but also for economic ends. This can lead to Diagnosis Related Groups (DRG) creep, when hospitals choose codes with a higher reimbursement for economic purposes.[23]

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.[24] 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.

Conclusion

The conclusions from this study are that the incidence of AEs after hip arthroplasty is high and that the tested instruments 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.

Acknowledgments

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.

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.

Author contributions

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.

Ethical 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.

Data-sharing statement

No additional data are available.

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, Callaghan JJ, et al. Adverse Outcomes in Hip Arthroplasty: Long-Term Trends. *J Bone Joint Surg Am.* 2012 Jul 18;94(14):e103.
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 Am* 2017;99:1365–72.
5. Gagnier JJ, Morgenstern H, Kellam P. A retrospective cohort study of adverse events in patients undergoing orthopaedic surgery. *Patient Saf Surg* 2017;11:15.
6. Phillips CB, Barrett JA, Losina E, et al. Incidence rates of dislocation, pulmonary embolism, and deep infection during the first six months after elective total hip replacement. *J Bone Joint Surg Am* 2003;85-A:20–6.
7. Hansson S, Rolfson O, Akesson K, et al. Complications and patient-reported outcome after hip fracture. A consecutive annual cohort study of 664 patients. *Injury* 2015;46:2206–11.
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. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. [Internet]. 1992. http://www.who.int/classifications/icd/en/bluebook.pdf (accessed 18 April 2018).

12. Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011;11:450.

13. Kärrholm K, Lindahl H, Malchau H, et al. The Swedish Hip Arthroplasty Register. Annual Report 2016. Sweden: Ola Rolfson 2016.

14. Griffin F, Resar R. IHI Global Trigger Tool for Measuring Adverse Events. Cambridge, MA: Institute for Healthcare Improvement 2009.

15. Swedish Association of Local Authorities and Regions (SALAR). Marker Based Record Review to Identify and Measure Harm in Healthcare (in Swe: Markörbaserad Journalgranskning – för Att Identifiera Och Mäta Skador i Vården). Stockholm: Sveriges kommuner och landsting 2012.

16. Swedish Government. Patient Safety Act SFS 2010:659 (in Swe: Patientsäkerhetslag SFS 2010:6 59). Ministry of Health and Social Affairs. [Internet]. 2010. https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/patientsakerhetslag-2010659_sfs-2010-659 (accessed 18 April 2018)

17. National Coordinating Council for Medication Error, Reporting and Prevention. NCC MERP index for categorizing medication errors. [Internet]. 2017. http://www.nccmerp.org/types-medicationerrors/. (accessed 18 April 2018).

18. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46.

19. 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.

20. 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.

21. 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.e1.

22. 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 Mar 1;194(3):257–66.

23. Zhan C, Miller M. Administrative data based patient safety research: a critical review. *Qual Saf Health Care* 2003;12:ii58–63.

24. Melvin JS, Karthikeyan T, Cope R, et al. Early failures in total hip arthroplasty -- a changing paradigm. *J Arthroplasty* 2014;29:1285–8.

1.

FIGURE LEGENDS

Figure 1. Flow chart of the study process.

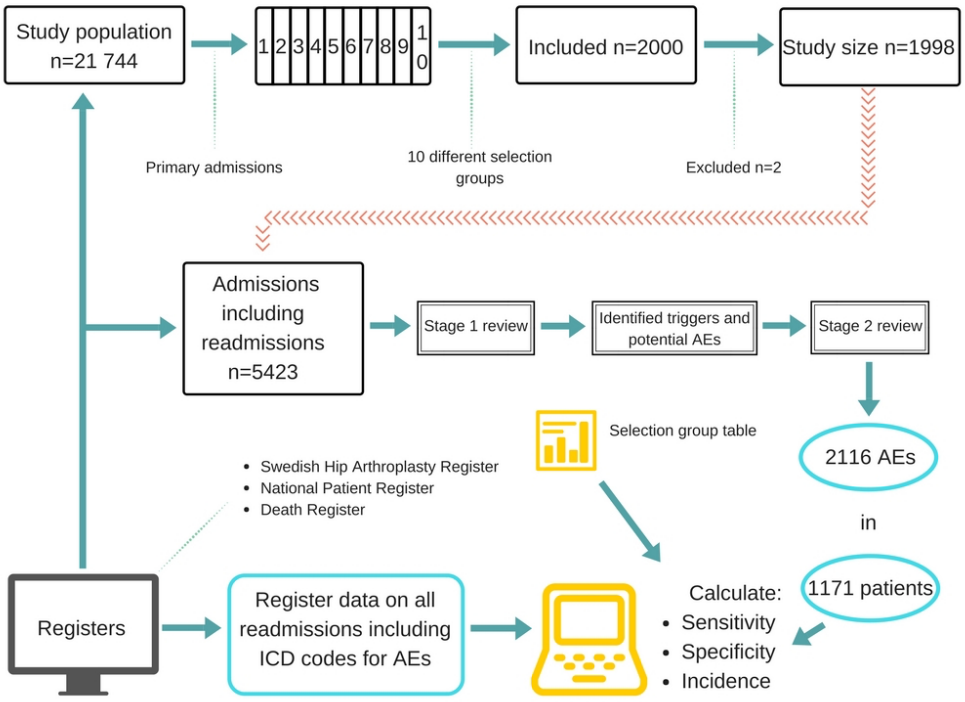

Figure 1. Flow chart of the study process.

270x203mm (96 x 96 DPI)

Table A1: Selection groups

		With AE code		No AE code	
		Population	Sample	Population	Sample
Percentiles (lenght of stay)	0 - 55 %	289	33	12 628	130
	56 - 80 %	206	49	3237	195
	81 - 100 %	537	74	2547	294
Readmission	20 - 30 days	630	294	631	442
	31 - 90 days	403	293	666	194

With AE code = Patients that had a recorded international classification of diseases (ICD) code indicating an AE. The following codes were used: as main diagnose: all I codes, J819, J13, J15, J18 and R33. As secondary diagnose I803, I269, L899, M243, M244, S730, T810, T813, T814, T840, T845 and T933.

Table A2: ICD-10 codes

ICD-10 code	As main code
I	Diseases of the circulatory system (All I codes)
J819	Pulmonary oedema
J15	Bacterial pneumonia
J18	Pneumonia, organism unspecified

J13	Pneumonia due to Streptococcus pneumoniae
R33	Urine retention
ICD-10 code	As secondary code
T810	Haemorrage and haematoma complication a procedure
T813	Disruption of operation wound
T814	Infection following a procedure
L899	Decubitus ulcer and pressure area, unspecified
T840	Mechanical complication of internal joint prosthesis
T845	Infection and inflammatory reaction due to internal joint prosthesis
S730	Dislocation of hip
T933	Sequele of dislocation, sprain and strain of lower limb

International classification of diseases (ICD) tenth revision codes used by the instrument for defining adverse events.

Table A.3: The 38 triggers with its five modules used in the study

Care module	Transfusion
	In-hospital stroke
	Cardiac arrest or deterioration in vital signs
	Unplanned dialysis
	Deep venous thrombosis or pulmonary embolus
	Fall
	Pressure ulcer
	Distended urinary bladder
	Thrombophlebitis or skin impairment
	Neurological impairment
	Abnormal temperature
	Positive blood culture
	Healthcare-associated infection
	Transfer to higher level of care
	Acute visit within 2 days after discharge from in-hospital care

	Readmission within 90 days
	Documentation of mistake
	Other
Laboratory module	Low haemoglobin value
	Low glucose value
	Increased creatinine value
	Abnormal potassium value
	Abnormal sodium value
Surgical and other invasive procedure module	Reoperation
	Change in procedure/organ harm
	Unplanned ventilation treatment
	Intra- or Post-Operative Death
	Postoperative increase of troponin
	Postoperative complication
	Anesthesia related impairment/harm
Medication module	Increased risk for haemorrhage
	Anaphylactic reaction
	Adverse drug event/adverse drug reaction
Intensive care module	Ventilator-associated pneumonia
	Readmission to the intensive care unit or other higher level of care
	Treatment within intensive care
	Intubation, reintubation, tracheotomy or coniotomy
	Intensive care unit syndrome

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study design with a commonly used term in the title or abstract.	1-4
		(*b*) In the abstract, provide an informative and balanced summary of what was done and what was found.	
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported.	6
Objectives	3	State specific objectives, including any prespecified hypotheses.	7
Methods			
Study design	4	Present key elements of the study design early in the paper.	7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection.	7
Participants	6	(*a*) Provide the eligibility criteria and the sources and methods of the selection of participants. Describe the methods of follow-up.	8
		(*b*) For matched studies, provide	

		matching criteria and the number of exposed and unexposed groups.	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Provide diagnostic criteria, if applicable.	9-10
Data sources/ measurement	8*	For each variable of interest, provide the sources of the data and details of the methods of assessment (measurement). Describe the comparability of the assessment methods if there is more than one group.	9
Bias	9	Describe any efforts to address potential sources of bias.	
Study size	10	Explain how the study size was determined.	7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why.	
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding.	13-14
		(*b*) Describe any methods used to examine subgroups and interactions.	
		(*c*) Explain how missing data were addressed.	

		(*d*) If applicable, explain how loss to follow-up was addressed.	
		(*e*) Describe any sensitivity analyses.	
Results			
Participants	13*	(a) Report the number of individuals at each stage of the study, such as the number of potentially eligible individuals, of those examined for eligibility, of those confirmed as eligible, of those included in the study, of those completing follow-up, and of those analysed.	14
		(b) Provide reasons for non-participation at each stage.	
		(c) Consider the use of a flow diagram.	
Descriptive data	14*	(a) Provide the characteristics of the study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders.	14-15
		(b) Indicate the number of participants with missing data for each variable of interest.	
		(c) Summarise the follow-up time (e.g., average and total duration).	
Outcome data	15*	Report numbers of outcome events or summary measures over time.	13-14
Main results	16	(*a*) Provide unadjusted estimates and, if applicable, confounder-adjusted	15-

		estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included.	16
		(*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 performed, such as analyses of subgroups and interactions as well as sensitivity analyses.	17-18
Discussion			
Key results	18	Summarise key results with reference to the study objectives.	19
Limitations	19	Discuss the limitations of the study, taking into account sources of potential bias or imprecision. Discuss both the direction and magnitude of any potential bias.	20-21
Interpretation	20	Provide a cautious overall interpretation of the results considering the objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	21
Generalisability	21	Discuss the generalisability (external validity) of the study results.	21
Other information			

Funding	22	Provide 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.	22

*Provide information separately for exposed and unexposed groups.

BMJ Open

Validation of adverse events after hip arthroplasty: a multicentre cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023773.R1
Article Type:	Research
Date Submitted by the Author:	03-Oct-2018
Complete List of Authors:	Magnéli, Martin; Karolinska Institutet, Department of Clinical Sciences, Danderyd Hospital, Division of Orthopaedics Unbeck, Maria; Karolinska Institutet Department of Molecular Medicine and Surgery Rogmark, Cecilia; Skanes universitetssjukhus Malmo Rolfson , Ola ; Institute of Clinical Sciences, Department of Orthopaedics, Sahlgrenska Academy, University of Gothenburg Hommel, Ami; Lunds Universitet, Samuelsson, Bodil; Karolinska Institutet, Department of Clinical Sciences at Danderyd Hospital, Division of Orthopaedics Schildmeijer, Kristina; School of Health and Caring Sciences, Linnaéus University Sjöstrand, Desirée; Skånevård Kryh, 11Education Unit Gordon, Max; Karolinska Institutet, Department of Clinical Sciences at Danderyd Hospital, Division of Orthopaedics Sköldenberg, Olof; Karolinska Institutet, Department of Clinical Sciences at Danderyd Hospital, Division of Orthopaedics
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Epidemiology
Keywords:	Adverse events < THERAPEUTICS, ORTHOPAEDIC & TRAUMA SURGERY, Hip < ORTHOPAEDIC & TRAUMA SURGERY, Orthopaedic & trauma surgery < SURGERY

Validation of adverse events after hip arthroplasty: a multicentre cohort study

Martin Magnéli[1,2], Maria Unbeck[3,4], Cecilia Rogmark[5,6,7], Ola Rolfson[7,8], Ami Hommel[9,10,11], Bodil Samuelsson[2], Kristina Schildmeijer[12], Desirée Sjöstrand[13], Max Gordon[1,2], Olof Sköldenberg[1,2]

[1]Department of Orthopaedics, Danderyd University Hospital, SE-182 88 Stockholm, Sweden

[2]Department of Clinical Sciences, Danderyd Hospital, Division of Orthopaedics, Karolinska Institutet, SE-182 88 Stockholm, Sweden

[3]Trauma and Reparative Medicine Theme, Karolinska University Hospital, SE-171 79 Stockholm, Sweden

[4]Department of Molecular Medicine and Surgery, Karolinska Institutet, SE-171 77 Stockholm, Sweden

[5]Skane University Hospital Malmö, SE-205 02 Malmö, Sweden

[6]Department of Orthopaedics, Lund University, SE-117 221 00 Lund, Sweden.

[7]Swedish Hip Arthroplasty Register, Centre of Registers Västra Götaland, SE-413 45 Gothenburg, Sweden

[8]Sahlgrenska Academy, SE-405 30 Gothenburg, Sweden

[9]Department of Care Science, Faculty of Health and Society, Malmö University, SE-205 06 Malmö, Sweden

[10]Department of Orthopaedics, Skane University Hospital, SE-221 41 Lund, Sweden.

[11]Swedish Hip Fracture Register, SE-221 85 Lund, Sweden

[12]School of Health and Caring Sciences, Faculty of Health and Life Sciences, Linnaeus University, S-391 82 Kalmar, Sweden

[13]Education Unit, Skånevård Kryh, Region Skåne, SE-291 89 Kristianstad, Sweden

Corresponding author: Martin Magnéli, Department of Orthopaedics, Danderyd Hospital Mörbygårdsvägen 88, SE-182 88 Stockholm, Sweden, E-mail: martin.magneli@sll.se, Telephone: +46(0) 737 886 557

Word count: 4153

Number of figures: 1

Number of tables: 4

Number of references: 28

Tables in appendix: 6

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 that is based on administrative data from the national patient register (NPR), is used by both the Swedish hip Arthroplasty register (SHAR) 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 (RRR) with Global Trigger Tool methodology in combination with register data.

Setting

Twenty-four different hospitals in four major regions of Sweden.

Participants

Two thousand patients with either total or hemi hip arthroplasty were recruited from the SHAR. We included acute patients with hip fractures and elective patients with degenerative joint disease.

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: 4.9-6.7%) for 30 days and 14.8% (95% CI: 8.2-24.3) for 90 days, and the specificity was 95.2% (95% CI: 93.5-96.6%) for 30 days and 92.1% (95% CI: 89.9-93.8%) for 90 days. The adjusted cumulative incidence for all AEs was 28.4% (95% CI: 25.0 – 32.3%) for 30 days and 39.5% (95% CI: 26.0 – 33.8%) for 90 days. The incidence rate was 0.43 AEs per person-month (95% CI: 0.39 – 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.

ARTICLE SUMMARY

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 multicentre study design, that 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 generalizable to healthcare systems where International Classification of Disease codes are used to measure AEs.

KEYWORDS

Orthopaedics, Adverse events, Hip arthroplasty, Validation, Global Trigger Tool

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 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% – 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 arthroplasty[6] and 2) there are many different methods for identifying AEs which complicates comparisons.[7]

The method that has been proven to be most sensitive compared to 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 (Table A1, appendix), that 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 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 multicentre 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 2 000 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 Västernorrland) 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 (Table A2, appendix).

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 (Table A3, appendix). 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 (Table A2, appendix) 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 postoperatively. 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 ten reviewers with a record review experience ranging from novice to expert (Table A4, appendix). The more experienced reviewers performed both stage one and two of the review. All reviewers received obligatory one-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.

The 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 predefined 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).

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 4 were included, and the reviewer made a full assessment that included evaluations of preventability, type of AE (71 different types in 15 different categories, Table A5, appendix), 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 hospitalization, G) contributed to or resulted in permanent harm, H) required intervention necessary to sustain life within 60 minutes and I) contributed or resulted in the patient´s death.

Preventable AEs that were classed as NCC MERP F or higher were classified as major AEs.

Reliability and validity

Inter-rater reliability was evaluated through the double review of six percent of the records to assess 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 (Table A1, Appendix). Five of the specific codes and the code category has to be as primary diagnose and the remaining eight can be either as primary or secondary code. In the validation of the instrument, test positive for an AE was defined as 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 ≥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 days 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 ten 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=2 000) were used to calculate the 95% confidence intervals.

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 2 000 patients weighted according to the selection group table (Table A2, Appendix). 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, 1 998 patients with a total of 5 422 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 1 331 elective patients, and 63% of the patients were female. The hip fracture group comprised more women, contained older patients, and had a longer length of stay during the index admission (Table1).

FIGURE LEGENDS

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= 1 998	Acute n= 667	Elective n= 1 331
Female n=	1 250 (62.6%)	444 (66.6%)	806 (60.6%)
Male n=	748 (37.4%)	223 (33.4%)	525 (39.4%)
Age, median†	77.0 (68.0 - 84.0)	84.0 (79.0 - 89.0)	73.0 (64.0 - 80.0)
LOS, median†	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%)
County	556 (27.8%)	180 (27.0%)	376 (28.2%)
Rural	531 (26.6%)	109 (16.3%)	422 (31.7%)
Private	281 (14.1%)	83 (12.4%)	198 (14.9%)

†, Interquartile range

Identified AEs and rate of ICD-10 codes

In total, we found 2 116 AEs in 1 171 (58.6%) patients. Of these, 1 604 AEs (75.8%) in 975 (48.8%) patients were classified as preventable AEs, 1 066 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 1 135 (53.6%). The acute patients sustained 758 (47.3%) of the preventable AEs and 431 (40.4%) of the major AEs.

Of the 2 116 found AEs, an ICD-10 code for the AE was found in 1 145 (54.1%) records, in 879 (54.8%) of the 1 604 preventable AEs, in 787 (71.1%) of the 1 066 major AEs and in 758 (62.9%) of the 1 206 selected AEs.

The majority of AEs occurred during the index admission (n=1 260, 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 (Table A6, appendix).

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

	Available ICD-code			
	Yes n=	**No n=**	**Total n=**	**Rate %**
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

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 higher incidence than the

elective patients with 51.4% compared to 17.2% for 30 days and 52.1% compared to 18.6% for 90 days. The incidence of preventable AEs and major AEs were also higher for the acute patients compared with the elective, both for 30 and 90 days.

Table 3, adjusted cumulative incidence of adverse events (AEs)

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

All results are in %, 95% confidence interval in brackets.

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

postoperatively. 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.

Table 4, validation of instrument

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

AEs, adverse events
†, AEs found (with RRR) during index and readmissions versus instrument (readmissions only)
‡, AEs found (with RRR) during readmissions versus instrument (readmissions only)
95% confidence interval in brackets

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 κ=0.828 and 0.965, respectively. The inter-rater reliability for whether the record was to be forwarded to secondary review was κ=0.965. The inter-rater reliability values for the identification of a specific event or whether that event was a potential AE were κ=0.65 and 0.873, respectively.

DISCUSSION

In this retrospective multicentre cohort study using RRR on 1 998 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 2 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. compared ICD-codes with record review and also found low sensitivity and high specificity.[21] 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 by some be considered as excessive. 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. found in their study on total knee and hip replacements.[24] 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 (AHRQ-PSI), an ICD code-based instrument, with the Agency for Healthcare Quality and Research National Surgical Quality Improvement Program (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.

Acknowledgments

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 tremendesly.

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.

Author contributions

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.

Ethical 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.

Data-sharing statement

No additional data are available.

(2)

REFERENCES

1. Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. Lancet Lond Engl. 2007 Oct 27;370(9597):1508–19.

2. Wolf BR, Lu X, Li Y, Callaghan JJ, Cram P. Adverse Outcomes in Hip Arthroplasty: Long-Term Trends. J Bone Joint Surg Am. 2012 Jul 18;94(14):e103.

3. Huddleston JI, Wang Y, Uquillas C, Herndon JH, Maloney WJ. Age and obesity are risk factors for adverse events after total hip arthroplasty. Clin Orthop. 2012 Feb;470(2):490–6.

4. Millstone DB, Perruccio AV, Badley EM, Rampersaud YR. Factors Associated with Adverse Events in Inpatient Elective Spine, Knee, and Hip Orthopaedic Surgery. J Bone Jt Surg-Am Vol. 2017 Aug 16;99(16).

5. Grober ED, Bohnen JMA. Defining medical error. Can J Surg. 2005 Feb;48(1):39–44.

6. Healy WL, Iorio R, Clair AJ, Pellegrini VD, Della Valle CJ, Berend KR. Complications of Total Hip Arthroplasty: Standardized List, Definitions, and Stratification Developed by The Hip Society. Clin Orthop. 2016 Feb;474(2):357–64.

7. Thomas EJ, Petersen LA. Measuring errors and adverse events in health care. J Gen Intern Med. 2003 Jan;18(1):61–7.

8. Unbeck M, Muren O, Lillkrona U. Identification of adverse events at an orthopedics department in Sweden. Acta Orthop. 2008 Jun;79(3):396–403.

9. Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al. 'Global Trigger Tool' Shows That Adverse Events In Hospitals May Be Ten Times Greater Than Previously Measured. Health Aff (Millwood). 2011 Apr 1;30(4):581–9.

10. Naessens JM, Campbell CR, Huddleston JM, Berg BP, Lefante JJ, Williams AR, et al. A comparison of hospital adverse events identified by three widely used detection methods. Int J Qual Health Care. 2009 Aug 1;21(4):301–7.

11. World Health Organization. ICD - ICD-10-CM - International Classification of Diseases, Tenth Revision, Clinical Modification [Internet]. Geneva: World Health Organization; 2017. [cited 2018 Apr 18]. Available from: https://www.cdc.gov/nchs/icd/icd10cm.htm

12. Kärrholm J, Mohaddes M, Odin D, Vinblad J, Rogmark C, Rolfson O. Swedish Hip Arthroplasty Register, Annual report 2017 [Internet]. Gothenburg: Sweden; 2018. [cited 2018 Sep 26]. Available from: https://registercentrum.blob.core.windows.net/shpr/r/-rsrapport-2017-S1xKMzsAwX.pdf

13. Swedish National Board of Health and Welfare. Patientregistret [Internet]. [accessed 2018 Aug 29]. Available from: https://www.socialstyrelsen.se/register/halsodataregister/patientregistret/inenglish

14. Oönskade händelser efter knä- och total höftprotesoperation | Vården i siffror - Statistik för kvalitets- och förbättringsarbete [Internet]. Stockholm: Sveriges Kommuner och Landsting; 2018. (In Eng: Adverse events after knee and hip arthroplasty. Stockholm: Swedish Association of Local Authorities and Regions; 2018). [cited 2018 Sep 26]. Available from: https://www.vardenisiffror.se/indikator?datefrom=2013-01-01&metadatameasure=74903f74-ac82-4d5b-8320-2a409df71255&relatedmeasuresbyentry=keyword&relatedmeasuresbyid=ortopedi&units=se

15. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim J-L, Reuterwall C, et al. External review and validation of the Swedish national inpatient register. BMC Public Health. 2011 Jun 9;11:450.

16. Griffin F, Resar R. IHI Global Trigger Tool for Measuring Adverse Events (Second Edition). Cambridge, MA: Institute for Healthcare Improvement; 2009. (IHI Innovation Series white paper.).

17. Markörbaserad Journalgranskning – för att identifiera och mäta skador i vården [Internet]. 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). [cited 2018 Sept 26]. Available from : https://webbutik.skl.se/bilder/artiklar/pdf/7164-847-1.pdf

18. Patientsäkerhetslag (2010:659) [Internet]. Stockholm: Socialdepartementet; 2010. (In Eng: Patient safety act (2010:659). Stockholm: Swedish Parliament; 2010). [cited 2018 Sept 2018]. Available from: https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/patientsakerhetslag-2010659_sfs-2010-659

19. National Coordinating Council for Medication Error, Reporting and Prevention. NCC MERP Index for Categorizing Medication Errors. [Internet]. [cited 2018 sept 26]. Available from: http://www. nccmerp. org/ types- medicationerrors/

20. Cohen J. A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas. 1960 Apr;20(1):37–46.

21. Maass C, Kuske S, Lessing C, Schrappe M. 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 J Int Soc Qual Health Care. 2015 Aug;27(4):305–13.

22. Unbeck M, Schildmeijer K, Henriksson P, Jürgensen U, Muren O, Nilsson L, 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 Apr 15;7:10.

23. Rutberg H, Borgstedt-Risberg M, Gustafson P, Unbeck M. Adverse events in orthopedic care identified via the Global Trigger Tool in Sweden – implications on preventable prolonged hospitalizations. Patient Saf Surg [Internet]. 2016 Oct 26 [cited 2017 Mar 31];10. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5080833/

24. Jørgensen CC, Petersen MA, Kehlet H. Preoperative prediction of potentially preventable morbidity after fast-track hip and knee arthroplasty: a detailed descriptive cohort study. BMJ Open. 2016 Jan;6(1):e009813.

25. Sebastian AS, Polites SF, Glasgow AE, Habermann EB, Cima RR, Kakar S. Current Quality Measurement Tools Are Insufficient to Assess Complications in Orthopedic Surgery. J Hand Surg. 2017 Jan;42(1):10–15.e1.

26. Best WR, Khuri SF, Phelan M, Hur K, Henderson WG, Demakis JG, 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 1 1No competing interests declared. J Am Coll Surg. 2002 Mar 1;194(3):257–66.

27. Robertsson O, W-Dahl A, Lidgren L, Sundber. Annual Report 2017, Swedish Knee Arthroplasty Register. [Internet] Lund: Swedish Knee Arthroplasty Register; 2017. [cited 2018 Sept 26]. Available from: http://www.myknee.se/pdf/SVK_2017_Eng_1.0.pdf

28. Melvin JS, Karthikeyan T, Cope R, Fehring TK. Early failures in total hip arthroplasty -- a changing paradigm. J Arthroplasty. 2014 Jun;29(6):1285–8.

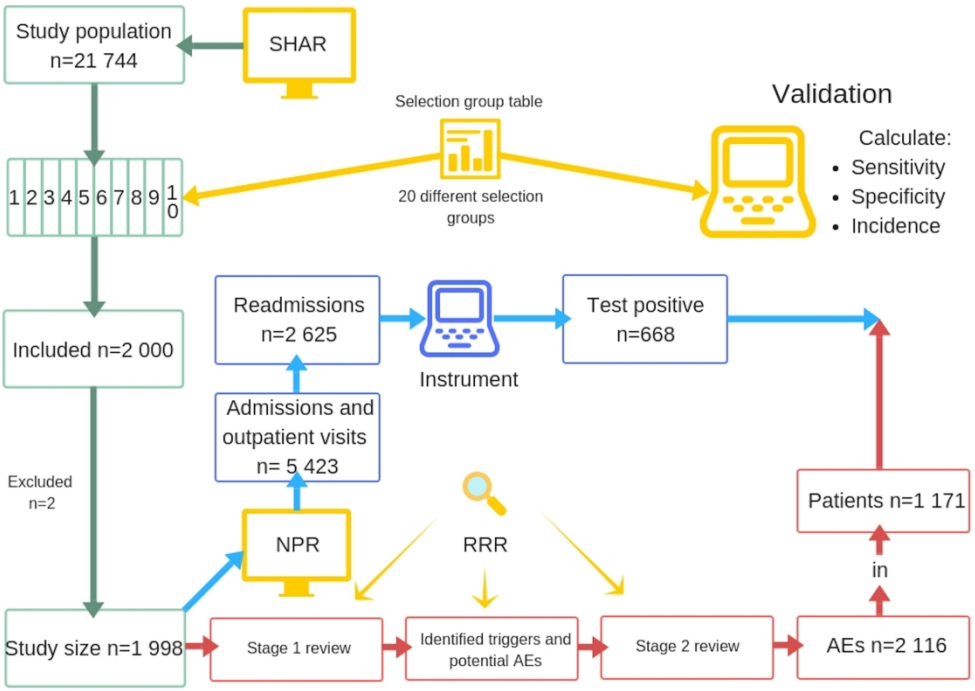

Figure 1, flowchart of the study process.
AEs, adverse events; NPR, National Patient Register; SHAR, Swedish Hip Arthroplasty Register; RRR, retrospective record review.

203x152mm (300 x 300 DPI)

APPENDIX

Table A1, the set of ICD-10 codes for the defining an adverse event by the instrument

As main diagnosis	
All I codes	Diseases of the circulatory system
J819	Pulmonary oedema
J13	Pneumonia due to Streptococcus pneumoniae
J15	Bacterial pneumonia, not elsewhere classified
J18	Pneumonia, organism unspecified
R33	Retention of urine
As main or secondary diagnosis	
L899	Decubitus ulcer and pressure area, unspecified
S730	Dislocation, sprain and strain of joint and ligaments of hip
T810	Haemorrhage and haematoma complicating a procedure, not elsewhere classified
T813	Disruption of operation wound, not elsewhere classified
T814	Infection following a procedure, not elsewhere classified
T840	Mechanical complication of internal joint prosthesis
T845	Infection and inflammatory reaction due to internal joint prosthesis
T933	Sequelae of dislocation, sprain and strain of lower limb

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

Table A2, selection groups used for the weighted sample

With a predefined ICD-10 code indicating an AE in the NPR		Acute		Elective	
		Population	Sample	Population	Sample
Percentiles of length of stay	0 - 55 %	194	11	95	22
	56 - 80 %	148	16	58	33
	81 - 100 %	302	25	235	49
Readmission	2 - 30 days	274	98	356	196
	31 - 90 days	199	98	204	195

Without a predefined ICD-10 code indicating an AE in the NPR		Acute		Elective	
		Population	Sample	Population	Sample
Percentiles of length of stay	0 - 55 %	2859	44	9769	86
	56 - 80 %	1167	65	2070	131
	81 - 100 %	766	97	1781	197
Readmission	2 - 30 days	294	147	337	295
	31 - 90 days	341	66	325	129
	Total	6544	667	15230	838

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

Table A3 , set of ICD-10 codes used in the selection of patients

As main diagnosis	
All I codes	Diseases of the circulatory system
J819	Pulmonary oedema
J13	Pneumonia due to Streptococcus pneumoniae
J15	Bacterial pneumonia, not elsewhere classified
J18	Pneumonia, organism unspecified
R33	Retention of urine
As main or secondary diagnosis	
I803	Phlebitis and thrombophlebitis of lower extremities, unspecified
I269	Pulmonary embolism without mention of acute cor pulmonale
L899	Decubitus ulcer and pressure area, unspecified
M243	Pathological dislocation and subluxation of joint, not elsewhere classified
M244	Recurrent dislocation and subluxation of joint
S730	Dislocation, sprain and strain of joint and ligaments of hip
T810	Haemorrhage and haematoma complicating a procedure, not elsewhere classified
T813	Disruption of operation wound, not elsewhere classified
T814	Infection following a procedure, not elsewhere classified
T840	Mechanical complication of internal joint prosthesis
T845	Infection and inflammatory reaction due to internal joint prosthesis
T933	Sequelae of dislocation, sprain and strain of lower limb

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

Table A4, characteristics of the reviewers and panel of specialists available for consultation during the review process.

Type of profession	Experience in RRR
Registered nurse	Expert in using different RRR methods including GTT
Registered nurse	Expert in using GTT
Registered nurse	Skilled in using GTT
Registered nurse	Skilled in using GTT
Registered nurse	Skilled in using GTT
Operating room nurse	Unfamiliar with GTT
Medical student	Unfamiliar with GTT
Medical student	Unfamiliar with GTT
Resident orthopedic surgeron	Used to structured review records but unfamiliar with GTT
Senior consultant orthopedic surgeon	Used to structured review records but unfamiliar with GTT
Senior consultant orthopedic surgeon	Specialist available for consultation
Senior consultant orthopedic surgeon	Specialist available for consultation
Senior consultant in internal medicine	Specialist available for consultation

GTT, Global Trigger tool; RRR, Retrospective record review

Table A5, the 38 triggers with its five modules used in the study

Care module	Transfusion
	In-hospital stroke
	Cardiac arrest or deterioration in vital signs
	Unplanned dialysis
	Deep venous thrombosis or pulmonary embolus
	Fall
	Pressure ulcer
	Distended urinary bladder
	Thrombophlebitis or skin impairment
	Neurological impairment
	Abnormal temperature
	Positive blood culture
	Healthcare-associated infection
	Transfer to higher level of care
	Acute visit within 2 days after discharge from in-hospital care
	Readmission within 90 days
	Documentation of mistake
	Other
Laboratory module	Low haemoglobin value
	Low glucose value
	Increased creatinine value
	Abnormal potassium value
	Abnormal sodium value
Surgical and other invasive procedure module	Reoperation
	Change in procedure/organ harm
	Unplanned ventilation treatment
	Intra- or Post-Operative Death
	Postoperative increase of troponin
	Postoperative complication
	Anesthesia related impairment/harm
Medication module	Increased risk for haemorrhage
	Anaphylactic reaction
	Adverse drug event/adverse drug reaction
Intensive care module	Ventilator-associated pneumonia
	Readmission to the intensive care unit or other higher level of care
	Treatment within intensive care
	Intubation, reintubation, tracheotomy or coniotomy
	Intensive care unit syndrome

Table A6, the ten individual types of adverse events without any ICD-10 code

Type of AE	Found n=
Thrombophlebitis	7
Pressure ulcer unknown category	7
Other AEs caused by anaesthesia	4
Respiratory arrest	3
Awareness during anaesthesia	2
Pressure ulcer category 4	2
Superficial vessel damage	1
Genital infection (vaginal candidiasis)	1
Neurological AE: Muscle weakness	1

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

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study design with a commonly used term in the title or abstract.	1-4
		(*b*) In the abstract, provide an informative and balanced summary of what was done and what was found.	
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported.	6
Objectives	3	State specific objectives, including any prespecified hypotheses.	7
Methods			
Study design	4	Present key elements of the study design early in the paper.	7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection.	7
Participants	6	(*a*) Provide the eligibility criteria and the sources and methods of the selection of participants. Describe the methods of follow-up.	8
		(*b*) For matched studies, provide	

		matching criteria and the number of exposed and unexposed groups.	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Provide diagnostic criteria, if applicable.	9-10
Data sources/ measurement	8*	For each variable of interest, provide the sources of the data and details of the methods of assessment (measurement). Describe the comparability of the assessment methods if there is more than one group.	9
Bias	9	Describe any efforts to address potential sources of bias.	
Study size	10	Explain how the study size was determined.	7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why.	
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding.	13-14
		(*b*) Describe any methods used to examine subgroups and interactions.	
		(*c*) Explain how missing data were addressed.	

		(*d*) If applicable, explain how loss to follow-up was addressed.	
		(*e*) Describe any sensitivity analyses.	
Results			
Participants	13*	(a) Report the number of individuals at each stage of the study, such as the number of potentially eligible individuals, of those examined for eligibility, of those confirmed as eligible, of those included in the study, of those completing follow-up, and of those analysed.	14
		(b) Provide reasons for non-participation at each stage.	
		(c) Consider the use of a flow diagram.	
Descriptive data	14*	(a) Provide the characteristics of the study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders.	14-15
		(b) Indicate the number of participants with missing data for each variable of interest.	
		(c) Summarise the follow-up time (e.g., average and total duration).	
Outcome data	15*	Report numbers of outcome events or summary measures over time.	13-14
Main results	16	(*a*) Provide unadjusted estimates and, if applicable, confounder-adjusted	15-

		estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included.	16
		(*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 performed, such as analyses of subgroups and interactions as well as sensitivity analyses.	17-18
Discussion			
Key results	18	Summarise key results with reference to the study objectives.	19
Limitations	19	Discuss the limitations of the study, taking into account sources of potential bias or imprecision. Discuss both the direction and magnitude of any potential bias.	20-21
Interpretation	20	Provide a cautious overall interpretation of the results considering the objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	21
Generalisability	21	Discuss the generalisability (external validity) of the study results.	21
Other information			

Funding	22	Provide 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.	22

*Provide information separately for exposed and unexposed groups.

BMJ Open

Validation of adverse events after hip arthroplasty: a Swedish multicentre cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023773.R2
Article Type:	Research
Date Submitted by the Author:	16-Jan-2019
Complete List of Authors:	Magnéli, Martin; Karolinska Institutet, Department of Clinical Sciences, Danderyd Hospital, Division of Orthopaedics Unbeck, Maria; Karolinska Institutet Department of Molecular Medicine and Surgery Rogmark, Cecilia; Skanes universitetssjukhus Malmo Rolfson , Ola ; Institute of Clinical Sciences, Department of Orthopaedics, Sahlgrenska Academy, University of Gothenburg Hommel, Ami; Lunds Universitet, Samuelsson, Bodil; Karolinska Institutet, Department of Clinical Sciences at Danderyd Hospital, Division of Orthopaedics Schildmeijer, Kristina; School of Health and Caring Sciences, Linnaéus University Sjöstrand, Desirée; Skånevård Kryh, 11Education Unit Gordon, Max; Karolinska Institutet, Department of Clinical Sciences at Danderyd Hospital, Division of Orthopaedics Sköldenberg, Olof; Karolinska Institutet, Department of Clinical Sciences at Danderyd Hospital, Division of Orthopaedics
Primary Subject Heading:	Surgery
Secondary Subject Heading:	Epidemiology
Keywords:	Adverse events < THERAPEUTICS, ORTHOPAEDIC & TRAUMA SURGERY, Hip < ORTHOPAEDIC & TRAUMA SURGERY, Orthopaedic & trauma surgery < SURGERY

Validation of adverse events after hip arthroplasty: a Swedish multicentre cohort study

Martin Magnéli[1,2], Maria Unbeck[3,4], Cecilia Rogmark[5,6,7], Ola Rolfson[7,8], Ami Hommel[9,10,11], Bodil Samuelsson[2], Kristina Schildmeijer[12], Desirée Sjöstrand[13], Max Gordon[1,2], Olof Sköldenberg[1,2]

[1]Department of Orthopaedics, Danderyd University Hospital, SE-182 88 Stockholm, Sweden

[2]Department of Clinical Sciences, Danderyd Hospital, Division of Orthopaedics, Karolinska Institutet, SE-182 88 Stockholm, Sweden

[3]Trauma and Reparative Medicine Theme, Karolinska University Hospital, SE-171 79 Stockholm, Sweden

[4]Department of Molecular Medicine and Surgery, Karolinska Institutet, SE-171 77 Stockholm, Sweden

[5]Skane University Hospital Malmö, SE-205 02 Malmö, Sweden

[6]Department of Orthopaedics, Lund University, SE-117 221 00 Lund, Sweden.

[7]Swedish Hip Arthroplasty Register, Centre of Registers Västra Götaland, SE-413 45 Gothenburg, Sweden

[8]Sahlgrenska Academy, SE-405 30 Gothenburg, Sweden

[9]Department of Care Science, Faculty of Health and Society, Malmö University, SE-205 06 Malmö, Sweden

[10]Department of Orthopaedics, Skane University Hospital, SE-221 41 Lund, Sweden.

[11]Swedish Hip Fracture Register, SE-221 85 Lund, Sweden

[12]School of Health and Caring Sciences, Faculty of Health and Life Sciences, Linnaeus University, S-391 82 Kalmar, Sweden

[13]Education Unit, Skånevård Kryh, Region Skåne, SE-291 89 Kristianstad, Sweden

Corresponding author: Martin Magnéli, Department of Orthopaedics, Danderyd Hospital Mörbygårdsvägen 88, SE-182 88 Stockholm, Sweden, E-mail: martin.magneli@sll.se, Telephone: +46(0) 737 886 557

Word count: 4153

Number of figures: 1

Number of tables: 4

Number of references: 28

Tables in appendix: 6

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 (NPR), is used by both the Swedish hip Arthroplasty register (SHAR) 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 (RRR) with Global Trigger Tool methodology in combination with register data.

Setting

Twenty-four different hospitals in four major regions of Sweden.

Participants

Two thousand 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: 4.9-6.7%) for 30 days and 14.8% (95% CI: 8.2-24.3) for 90 days, and the specificity was 95.2% (95% CI: 93.5-96.6%) for 30 days and

92.1% (95% CI: 89.9-93.8%) for 90 days. The adjusted cumulative incidence for all AEs was 28.4% (95% CI: 25.0 – 32.3%) for 30 days and 39.5% (95% CI: 26.0 – 33.8%) for 90 days. The incidence rate was 0.43 AEs per person-month (95% CI: 0.39 – 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.

ARTICLE SUMMARY

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 multicentre study design, that 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 generalizable to healthcare systems where International Classification of Disease codes are used to measure AEs.

KEYWORDS

Orthopaedics, Adverse events, Hip arthroplasty, Validation, Global Trigger Tool

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 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% – 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 arthroplasty[6] and 2) there are many different methods for identifying AEs which complicates comparisons.[7]

The method that has been proven to be most sensitive compared to 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 (Table A1, appendix), that 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 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 multicentre 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 2 000 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 Västernorrland) 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 (Table A2, appendix).

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 (Table A3, appendix). 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 (Table A2, appendix) 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 postoperatively. 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 ten reviewers with a record review experience ranging from novice to expert (Table A4, appendix). The more experienced reviewers performed both stage one and two of the review. All reviewers received obligatory one-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.

The 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 predefined 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) (Table A5, Appendix).

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 4 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 hospitalization, G) contributed to or resulted in permanent harm, H) required intervention necessary to sustain life within 60 minutes and I) contributed or resulted in the patient´s death.

Reliability and validity

Inter-rater reliability was evaluated through the double review of six percent of the records to assess 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 (Table A1, Appendix). Five of the specific codes and the code category has to be as primary diagnose and the remaining eight can be either as primary or secondary code. In the validation of the instrument, test positive for an AE was defined as 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 ≥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 days 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 ten 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=2 000) were used to calculate the 95% confidence intervals.

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 2 000 patients weighted according to the selection group table (Table A2, Appendix). 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, 1 998 patients with a total of 5 422 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 1 331 elective patients, and 63% of the patients were female. The hip fracture group comprised more women, contained older patients, and had a longer length of stay during the index admission (Table1).

FIGURE LEGENDS

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= 1 998	Acute n= 667	Elective n= 1 331
Female n=	1 250 (62.6%)	444 (66.6%)	806 (60.6%)
Male n=	748 (37.4%)	223 (33.4%)	525 (39.4%)
Age, median†	77.0 (68.0 - 84.0)	84.0 (79.0 - 89.0)	73.0 (64.0 - 80.0)
LOS, median†	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%)

†, Interquartile range

Identified AEs and rate of ICD-10 codes

In total, we found 2 116 AEs in 1 171 (58.6%) patients. Of these, 1 605 AEs (75.9%) in 975 (48.8%) patients were classified as preventable AEs, 1 066 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 1 135 (53.6%). The acute patients sustained 758 (47.3%) of the preventable AEs and 431 (40.4%) of the major AEs.

Of the 2 116 found AEs, an ICD-10 code for the AE was found in 1 145 (54.1%) records, in 879 (54.8%) of the 1 605 preventable AEs, in 787 (71.1%) of the 1 066 major AEs and in 758 (62.9%) of the 1 206 selected AEs.

The majority of AEs occurred during the index admission (n=1 260, 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 (Table A6, appendix).

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

	Available ICD-code			
	Yes n=	**No n=**	**Total n=**	**Rate %**
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

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 higher incidence than the

elective patients with 51.4% compared to 17.2% for 30 days and 52.1% compared to 18.6% for 90 days. The incidence of preventable AEs and major AEs were also higher for the acute patients compared with the elective, both for 30 and 90 days.

Table 3, adjusted cumulative incidence of adverse events (AEs)

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

All results are in %, 95% confidence interval in brackets.

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

postoperatively. 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.

Table 4, validation of instrument

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

AEs, adverse events
†, AEs found (with RRR) during index and readmissions versus instrument (readmissions only)
‡, AEs found (with RRR) during readmissions versus instrument (readmissions only)
95% confidence interval in brackets

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 κ=0.828 and 0.965, respectively. The inter-rater reliability for whether the record was to be forwarded to secondary review was κ=0.965. The inter-rater reliability values for the identification of a specific event or whether that event was a potential AE were κ=0.65 and 0.873, respectively.

DISCUSSION

In this retrospective multicentre cohort study using RRR on 1 998 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 2 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. compared ICD-codes with record review and also found low sensitivity and high specificity.[21] 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 by some be considered as excessive. 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. found in their study on total knee and hip replacements.[24] 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 (AHRQ-PSI), an ICD code-based instrument, with the Agency for Healthcare Quality and Research National Surgical Quality Improvement Program (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.

Acknowledgments

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.

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.

Author contributions

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.

Ethical 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.

Data-sharing statement

No additional data are available.

(2)

REFERENCES

1. Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. Lancet Lond Engl. 2007 Oct 27;370(9597):1508–19.

2. Wolf BR, Lu X, Li Y, Callaghan JJ, Cram P. Adverse Outcomes in Hip Arthroplasty: Long-Term Trends. J Bone Joint Surg Am. 2012 Jul 18;94(14):e103.

3. Huddleston JI, Wang Y, Uquillas C, Herndon JH, Maloney WJ. Age and obesity are risk factors for adverse events after total hip arthroplasty. Clin Orthop. 2012 Feb;470(2):490–6.

4. Millstone DB, Perruccio AV, Badley EM, Rampersaud YR. Fa(10–12)ctors Associated with Adverse Events in Inpatient Elective Spine, Knee, and Hip Orthopaedic Surgery. J Bone Jt Surg-Am Vol. 2017 Aug 16;99(16).

5. Grober ED, Bohnen JMA. Defining medical error. Can J Surg. 2005 Feb;48(1):39–44.

6. Healy WL, Iorio R, Clair AJ, Pellegrini VD, Della Valle CJ, Berend KR. Complications of Total Hip Arthroplasty: Standardized List, Definitions, and Stratification Developed by The Hip Society. Clin Orthop. 2016 Feb;474(2):357–64.

7. Thomas EJ, Petersen LA. Measuring errors and adverse events in health care. J Gen Intern Med. 2003 Jan;18(1):61–7.

8. Unbeck M, Muren O, Lillkrona U. Identification of adverse events at an orthopedics department in Sweden. Acta Orthop. 2008 Jun;79(3):396–403.

9. Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al. 'Global Trigger Tool' Shows That Adverse Events In Hospitals May Be Ten Times Greater Than Previously Measured. Health Aff (Millwood). 2011 Apr 1;30(4):581–9.

10. Naessens JM, Campbell CR, Huddleston JM, Berg BP, Lefante JJ, Williams AR, et al. A comparison of hospital adverse events identified by three widely used detection methods. Int J Qual Health Care. 2009 Aug 1;21(4):301–7.

11. World Health Organization. ICD - ICD-10-CM - International Classification of Diseases, Tenth Revision, Clinical Modification [Internet]. Geneva: World Health Organization; 2017. [cited 2018 Apr 18]. Available from: https://www.cdc.gov/nchs/icd/icd10cm.htm

12. Kärrholm J, Mohaddes M, Odin D, Vinblad J, Rogmark C, Rolfson O. Swedish Hip Arthroplasty Register, Annual report 2017 [Internet]. Gothenburg: Sweden; 2018. [cited 2018 Sep 26]. Available from: https://registercentrum.blob.core.windows.net/shpr/r/-rsrapport-2017-S1xKMzsAwX.pdf

13. Swedish National Board of Health and Welfare. Patientregistret [Internet]. [accessed 2018 Aug 29]. Available from: https://www.socialstyrelsen.se/register/halsodataregister/patientregistret/inenglish

14. Oönskade händelser efter knä- och total höftprotesoperation | Vården i siffror - Statistik för kvalitets- och förbättringsarbete [Internet]. Stockholm: Sveriges Kommuner och Landsting; 2018. (In Eng: Adverse events after knee and hip arthroplasty. Stockholm: Swedish Association of Local Authorities and Regions; 2018). [cited 2018 Sep 26]. Available from: https://www.vardenisiffror.se/indikator?datefrom=2013-01-01&metadatameasure=74903f74-ac82-4d5b-8320-2a409df71255&relatedmeasuresbyentry=keyword&relatedmeasuresbyid=ortopedi&units=se

15. Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim J-L, Reuterwall C, et al. External review and validation of the Swedish national inpatient register. BMC Public Health. 2011 Jun 9;11:450.

16. Griffin F, Resar R. IHI Global Trigger Tool for Measuring Adverse Events (Second Edition). Cambridge, MA: Institute for Healthcare Improvement; 2009. (IHI Innovation Series white paper.).

17. Markörbaserad Journalgranskning – för att identifiera och mäta skador i vården [Internet]. 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). [cited 2018 Sept 26]. Available from : https://webbutik.skl.se/bilder/artiklar/pdf/7164-847-1.pdf

18. Patientsäkerhetslag (2010:659) [Internet]. Stockholm: Socialdepartementet; 2010. (In Eng: Patient safety act (2010:659). Stockholm: Swedish Parliament; 2010). [cited 2018 Sept 2018]. Available from: https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/patientsakerhetslag-2010659_sfs-2010-659

19. National Coordinating Council for Medication Error, Reporting and Prevention. NCC MERP Index for Categorizing Medication Errors. [Internet]. [cited 2018 sept 26]. Available from: http://www. nccmerp. org/ types- medicationerrors/

20. Cohen J. A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas. 1960 Apr;20(1):37–46.

21. Maass C, Kuske S, Lessing C, Schrappe M. 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 J Int Soc Qual Health Care. 2015 Aug;27(4):305–13.

22. Unbeck M, Schildmeijer K, Henriksson P, Jürgensen U, Muren O, Nilsson L, 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 Apr 15;7:10.

23. Rutberg H, Borgstedt-Risberg M, Gustafson P, Unbeck M. Adverse events in orthopedic care identified via the Global Trigger Tool in Sweden – implications on preventable prolonged hospitalizations. Patient Saf Surg [Internet]. 2016 Oct 26 [cited 2017 Mar 31];10. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5080833/

24. Jørgensen CC, Petersen MA, Kehlet H. Preoperative prediction of potentially preventable morbidity after fast-track hip and knee arthroplasty: a detailed descriptive cohort study. BMJ Open. 2016 Jan;6(1):e009813.

25. Sebastian AS, Polites SF, Glasgow AE, Habermann EB, Cima RR, Kakar S. Current Quality Measurement Tools Are Insufficient to Assess Complications in Orthopedic Surgery. J Hand Surg. 2017 Jan;42(1):10–15.e1.

26. Best WR, Khuri SF, Phelan M, Hur K, Henderson WG, Demakis JG, 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 1 1No competing interests declared. J Am Coll Surg. 2002 Mar 1;194(3):257–66.

27. Robertsson O, W-Dahl A, Lidgren L, Sundber. Annual Report 2017, Swedish Knee Arthroplasty Register. [Internet] Lund: Swedish Knee Arthroplasty Register; 2017. [cited 2018 Sept 26]. Available from: http://www.myknee.se/pdf/SVK_2017_Eng_1.0.pdf

28. Melvin JS, Karthikeyan T, Cope R, Fehring TK. Early failures in total hip arthroplasty -- a changing paradigm. J Arthroplasty. 2014 Jun;29(6):1285–8.

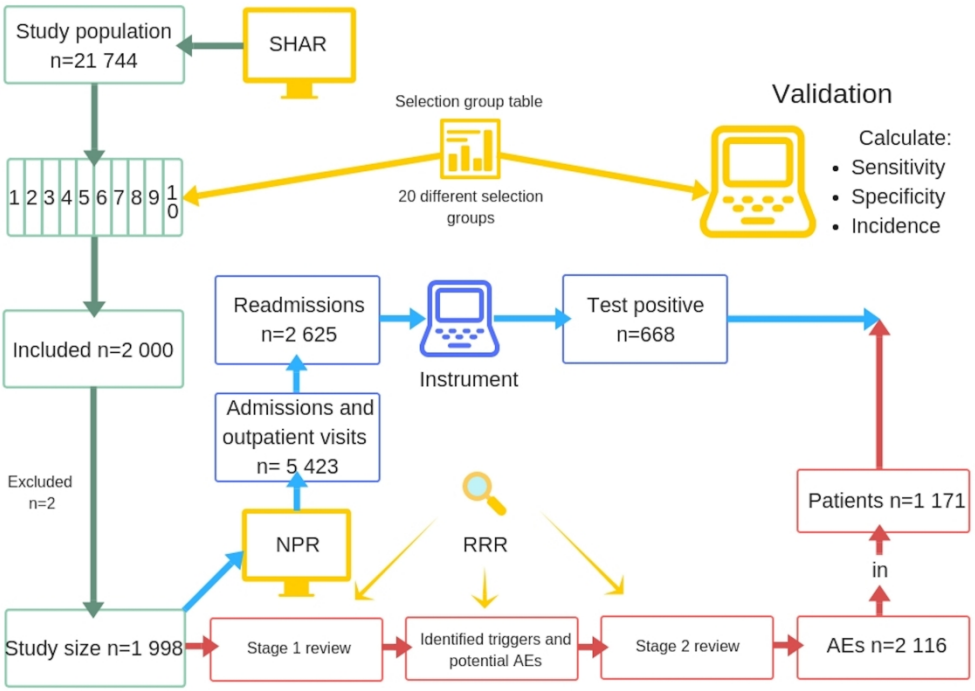

Figure 1, flowchart of the study process.
AEs, adverse events; NPR, National Patient Register; SHAR, Swedish Hip Arthroplasty Register; RRR, retrospective record review.

203x152mm (300 x 300 DPI)

APPENDIX

Table A1, the set of ICD-10 codes for the defining an adverse event by the instrument

As main diagnosis	
All I codes	Diseases of the circulatory system
J819	Pulmonary oedema
J13	Pneumonia due to Streptococcus pneumoniae
J15	Bacterial pneumonia, not elsewhere classified
J18	Pneumonia, organism unspecified
R33	Retention of urine
As main or secondary diagnosis	
L899	Decubitus ulcer and pressure area, unspecified
S730	Dislocation, sprain and strain of joint and ligaments of hip
T810	Haemorrhage and haematoma complicating a procedure, not elsewhere classified
T813	Disruption of operation wound, not elsewhere classified
T814	Infection following a procedure, not elsewhere classified
T840	Mechanical complication of internal joint prosthesis
T845	Infection and inflammatory reaction due to internal joint prosthesis
T933	Sequelae of dislocation, sprain and strain of lower limb

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

Table A2, selection groups used for the weighted sample

With a predefined ICD-10 code indicating an AE in the NPR		Acute		Elective	
		Population	Sample	Population	Sample
Percentiles of length of stay	0 - 55 %	194	11	95	22
	56 - 80 %	148	16	58	33
	81 - 100 %	302	25	235	49
Readmission	2 - 30 days	274	98	356	196
	31 - 90 days	199	98	204	195

Without a predefined ICD-10 code indicating an AE in the NPR		Acute		Elective	
		Population	Sample	Population	Sample
Percentiles of length of stay	0 - 55 %	2859	44	9769	86
	56 - 80 %	1167	65	2070	131
	81 - 100 %	766	97	1781	197
Readmission	2 - 30 days	294	147	337	295
	31 - 90 days	341	66	325	129
	Total	6544	667	15230	838

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

Table A3 , set of ICD-10 codes used in the selection of patients

As main diagnosis	
All I codes	Diseases of the circulatory system
J819	Pulmonary oedema
J13	Pneumonia due to Streptococcus pneumoniae
J15	Bacterial pneumonia, not elsewhere classified
J18	Pneumonia, organism unspecified
R33	Retention of urine
As main or secondary diagnosis	
I803	Phlebitis and thrombophlebitis of lower extremities, unspecified
I269	Pulmonary embolism without mention of acute cor pulmonale
L899	Decubitus ulcer and pressure area, unspecified
M243	Pathological dislocation and subluxation of joint, not elsewhere classified
M244	Recurrent dislocation and subluxation of joint
S730	Dislocation, sprain and strain of joint and ligaments of hip
T810	Haemorrhage and haematoma complicating a procedure, not elsewhere classified
T813	Disruption of operation wound, not elsewhere classified
T814	Infection following a procedure, not elsewhere classified
T840	Mechanical complication of internal joint prosthesis
T845	Infection and inflammatory reaction due to internal joint prosthesis
T933	Sequelae of dislocation, sprain and strain of lower limb

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

Table A4, characteristics of the reviewers and panel of specialists available for consultation during the review process.

Type of profession	Experience in RRR
Registered nurse	Expert in using different RRR methods including GTT
Registered nurse	Expert in using GTT
Registered nurse	Skilled in using GTT
Registered nurse	Skilled in using GTT
Registered nurse	Skilled in using GTT
Operating room nurse	Unfamiliar with GTT
Medical student	Unfamiliar with GTT
Medical student	Unfamiliar with GTT
Resident orthopedic surgeron	Used to structured review records but unfamiliar with GTT
Senior consultant orthopedic surgeon	Used to structured review records but unfamiliar with GTT
Senior consultant orthopedic surgeon	Specialist available for consultation
Senior consultant orthopedic surgeon	Specialist available for consultation
Senior consultant in internal medicine	Specialist available for consultation

GTT, Global Trigger tool; RRR, Retrospective record review

Table A5, the 38 triggers with its five modules used in the study

Care module	Transfusion
	In-hospital stroke
	Cardiac arrest or deterioration in vital signs
	Unplanned dialysis
	Deep venous thrombosis or pulmonary embolus
	Fall
	Pressure ulcer
	Distended urinary bladder
	Thrombophlebitis or skin impairment
	Neurological impairment
	Abnormal temperature
	Positive blood culture
	Healthcare-associated infection
	Transfer to higher level of care
	Acute visit within 2 days after discharge from in-hospital care
	Readmission within 90 days
	Documentation of mistake
	Other
Laboratory module	Low haemoglobin value
	Low glucose value
	Increased creatinine value
	Abnormal potassium value
	Abnormal sodium value
Surgical and other invasive procedure module	Reoperation
	Change in procedure/organ harm
	Unplanned ventilation treatment
	Intra- or Post-Operative Death
	Postoperative increase of troponin
	Postoperative complication
	Anesthesia related impairment/harm
Medication module	Increased risk for haemorrhage
	Anaphylactic reaction
	Adverse drug event/adverse drug reaction
Intensive care module	Ventilator-associated pneumonia
	Readmission to the intensive care unit or other higher level of care
	Treatment within intensive care
	Intubation, reintubation, tracheotomy or coniotomy
	Intensive care unit syndrome

Table A6, the ten individual types of adverse events without any ICD-10 code

Type of AE	Found n=
Thrombophlebitis	7
Pressure ulcer unknown category	7
Other AEs caused by anaesthesia	4
Respiratory arrest	3
Awareness during anaesthesia	2
Pressure ulcer category 4	2
Superficial vessel damage	1
Genital infection (vaginal candidiasis)	1
Neurological AE: Muscle weakness	1

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

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study design with a commonly used term in the title or abstract.	1-4
		(*b*) In the abstract, provide an informative and balanced summary of what was done and what was found.	
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported.	6
Objectives	3	State specific objectives, including any prespecified hypotheses.	7
Methods			
Study design	4	Present key elements of the study design early in the paper.	7
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection.	7
Participants	6	(*a*) Provide the eligibility criteria and the sources and methods of the selection of participants. Describe the methods of follow-up.	8
		(*b*) For matched studies, provide	

		matching criteria and the number of exposed and unexposed groups.	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Provide diagnostic criteria, if applicable.	9-10
Data sources/ measurement	8*	For each variable of interest, provide the sources of the data and details of the methods of assessment (measurement). Describe the comparability of the assessment methods if there is more than one group.	9
Bias	9	Describe any efforts to address potential sources of bias.	
Study size	10	Explain how the study size was determined.	7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why.	
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding.	13-14
		(*b*) Describe any methods used to examine subgroups and interactions.	
		(*c*) Explain how missing data were addressed.	

		(*d*) If applicable, explain how loss to follow-up was addressed.	
		(*e*) Describe any sensitivity analyses.	
Results			
Participants	13*	(a) Report the number of individuals at each stage of the study, such as the number of potentially eligible individuals, of those examined for eligibility, of those confirmed as eligible, of those included in the study, of those completing follow-up, and of those analysed.	14
		(b) Provide reasons for non-participation at each stage.	
		(c) Consider the use of a flow diagram.	
Descriptive data	14*	(a) Provide the characteristics of the study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders.	14-15
		(b) Indicate the number of participants with missing data for each variable of interest.	
		(c) Summarise the follow-up time (e.g., average and total duration).	
Outcome data	15*	Report numbers of outcome events or summary measures over time.	13-14
Main results	16	(*a*) Provide unadjusted estimates and, if applicable, confounder-adjusted	15-

		estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included.	16
		(*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 performed, such as analyses of subgroups and interactions as well as sensitivity analyses.	17-18
Discussion			
Key results	18	Summarise key results with reference to the study objectives.	19
Limitations	19	Discuss the limitations of the study, taking into account sources of potential bias or imprecision. Discuss both the direction and magnitude of any potential bias.	20-21
Interpretation	20	Provide a cautious overall interpretation of the results considering the objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	21
Generalisability	21	Discuss the generalisability (external validity) of the study results.	21
Other information			

Funding	22	Provide 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.	22

*Provide information separately for exposed and unexposed groups.