Outcomes following primary total hip arthroplasty

With the focus on the surgeon & surgeons' perceptions of feedback

Per Jolbäck

Department of Orthopaedic
Institute of Clinical Sciences
Sahlgrenska Academy
University of Gothenburg

Gothenburg, Sweden, 2020

UNIVERSITY OF
GOTHENBURG

Cover illustration by
Pontus Andersson/Pontus Art Production

Outcomes following primary total hip arthroplasty – With the focus on the surgeon & surgeons' perceptions of feedback

per.jolback@vgregion.se

ISBN 978-91-7833-776-7 (Print)
ISBN 978-91-7883-777-4 (PDF)
http://hdl.handle.net/2077/62220

Printed in Gothenburg, Sweden 2020
BrandFactory

"If you feel safe in the area you're working in, you're not working in the right area. Always go a little further into the water than you feel you're capable of being in. Go a little bit out of your depth. And when you don't feel that your feet are quite touching the bottom, you're just about in the right place to do something exciting."

– David Bowie

To Lotta, Nora & Alma

Contents

List of papers

This thesis is based on the following papers:

Paper I

High annual surgeon volume reduces the risk for adverse events following primary total hip arthroplasty. A registry-based study of 12,100 cases in Western Sweden

Jolbäck P, Rolfson O, Cnudde P, Malchau H, Odin D, Lindahl H, Mohaddes M.

Acta Orthop. 2019;90(2):153–58.

Paper II

Does surgeon experience affect patient-reported outcomes 1 year after primary total hip arthroplasty? A register-based study of 6,713 cases in western Sweden

Jolbäck P, Rolfson O, Mohaddes M, Nemes S, Kärrholm J, Garellick G, Lindahl H.

Acta Orthop. 2018;89(3):265–71.

Paper III

Surgeons' perceptions of being provided with their own results following primary total hip arthroplasties – a phenomenographic study

Jolbäck P, Mohaddes M, Lindahl H, Klaeson K.

Submitted

Paper IV

Few outliers among orthopaedic surgeons performing primary total hip arthroplasties in western Sweden. An observational study based on 9,482 cases and 208 surgeons in year 2011-2016

Jolbäck P, Nauclér E, Bülow E, Lindahl H, Mohaddes M.

Submitted

Abbreviations

Abbreviation	Definition
ASA	American Society of Anaesthesiologists
BMI	Body Mass Index
CUSUM	Cumulative Sum of Outcomes
CI	Confidence Interval
EQ-5D	The Five-dimension Self-rated Assessment Tool (EuroQol)
ICD-10	International Statistical Classification of Diseases and Related Health Problems 10th Revision
NOMESCO	Nordic Medico-Statistical Committee Classification of Surgical Procedures Codes for Interventions
VAS	Visual Analogue Scale

Abstract

Background

Total hip arthroplasty is considered to be one of the most successful orthopaedic interventions of its generation and it has been proclaimed as "the operation of the century". In this thesis, the focus will be on the surgeon and surgeons' perceptions of individual surgeon feedback from a national quality register. Individual surgeon feedback following hip arthroplasties provided by quality registers already exists in some countries. These feedback programmes use different statistical methods to detect outliers following arthroplasties. There is a limited number of published studies (to the author's knowledge) trying to explore and describe the surgeons' perception of being provided with feedback from a quality register or the effects following these feedback programmes.

Over the past few decades, several studies investigating the association between annual surgeon volumes or between the experience of the surgeon and the outcomes following primary total hip arthroplasties have been published. The results of the studies of annual surgeon volume are almost uniform in their conclusions; higher annual surgical activity improves the outcomes following primary total hip arthroplasties. No such association can be shown between a longer term of surgical experience and improved patient-reported outcomes. There may be some variation between countries in both the orthopaedic education and training programmes and the healthcare organisation and there might also be a difference at patient level that could influence these outcomes. None of these earlier studies of annual surgeon volume or surgeons' experience has been performed in Scandinavia. So, there is a lack of knowledge relating to information on whether the same associations can be shown in a Swedish setting as in the rest of the world.

Objectives

The overall aim of this project is to develop a methodology for providing feedback to individual surgeons following primary total hip arthroplasties and to explore the potential effect of this methodology being implemented. We posed the following questions:

- Are adverse events and mortality dependent on the number of surgeries performed each year by the surgeon?
- Are patient-reported outcomes dependent on the surgeon's experience?
- Which perceptions exist in the orthopaedic society about the phenomenon of feedback of individual surgeon's results from a national quality register?
- Which factors need to be taken into account to develop a feedback system in a Swedish setting?

Patients and method

Papers I, II and IV are register-based studies. For Papers I and IV, the data were retrieved from primary total hip arthroplasties due to osteoarthritis from ten local hospital records in the region of western Sweden, a regional patient record and the Swedish Hip Arthroplasty Register. Paper I comprise surgeries performed between 2007 and 2016. Paper IV comprises surgeries performed between 2011 and 2016.

For Paper II, data on primary total hip arthroplasties due to osteoarthritis performed between 2007 and 2012 were retrieved from the same ten local hospitals as in Papers I and IV and were linked to patient-reported outcomes from the Swedish Hip Arthroplasty Register. Information on the surgeons' year of specialist certification in orthopaedics or a licence to practise was obtained from the publicly available data from the Swedish National Board of Health and Welfare register of licensed health-care professionals.

For Paper III, all surgeons (both orthopaedic specialist and trainees) employed at one of the orthopaedic departments reporting to the Swedish Hip Arthroplasty Register were invited to participate in this

phenomenographic qualitative study. Based on the purposive sampling of informants, the informants were recruited for a semi-structured interview.

Results

Paper I comprised 12,100 primary total hip arthroplasties. The mean risk of an adverse event within 90 days postoperatively was seven per cent. A simple logistic regression demonstrated that, if the annual surgeon volume increased by ten primary total hip arthroplasties, the risk of an adverse event decreased by ten per cent and, in adjusted multiple regression, the corresponding number was eight per cent. The mortality rate in the study was low (zero point two per cent) and we were unable to find any association between 90-day mortality and annual surgeon volume.

For Paper II, 6,713 primary total hip arthroplasties were included in the analysis. The findings in Paper II showed that there was a statistically significant difference in patient age, American Society of Anaesthesiologists classification, Charnley classification, diagnosis and fixation technique associated with the surgeon's experience. At the one-year follow-up, there were no statistically significant differences in patient-reported outcomes among the subgroups of orthopaedic specialists. Patients operated on by orthopaedic trainees reported less satisfaction with the result of the surgery compared with surgeons with more than 15 years as orthopaedic specialists.

In Paper III, 19 interviews with orthopaedic surgeons and trainees from 15 hospitals were conducted. The analysis of the collected material outlined four categories of description expressed by the informants: 1) progression in the profession, 2) exposing the surgeons to inaccurate criticism, 3) might lead to impaired patient utility, 4) not contributing to enhanced feedback to surgeons.

In the last study, Paper IV, 9,482 primary total hip arthroplasties performed by 208 surgeons were included. In the observed funnel plots, the percentage of outliers was small for both adverse events within 90 days (zero to five per cent) and re-operations within two years (zero to one per cent). In the standardised models, the corresponding num-

bers were even lower (adverse events zero to three per cent/re-operations zero to one per cent). A small number of surgeons were outliers for adverse events within 90 days or re-operations within two years following primary total hip arthroplasties in a Swedish setting.

Conclusion

The findings in this thesis show that the outcomes in the form of short-term adverse events are dependent on the surgeons' operation volume and small number of surgeons will be outliers compared with their peers in a Swedish setting. However, patients can expect similar health improvements, pain relief and satisfaction at one year following surgery, irrespective of the surgeons' experience as orthopaedic specialists. There are a limited number of perceptions among Swedish orthopaedic specialists and trainees of the phenomenon of individual surgeon feedback. These perceptions vary between an opportunity to improve care for the patient through professional development to jeopardising patient safety by focusing more on the numbers than on the patients.

Summary in Swedish

Den övergripande målsättningen med det här avhandlingsarbetet var att utveckla en metodologi som förser ortopedkirurgen med information om sina egna resultat efter primär höftprotesoperation samt att undersöka effekterna om en sådan metod skulle införas.

Avhandlingen baseras på fyra delarbeten där vi i delarbete undersöker om oönskade händelser och död efter primär höftprotesoperation kan vara associerat med operatörens årliga volym? Och i delarbete II undersöker vi om det finns skillnader i det patientrapporterat utfallet baserat på operatörens erfarenhet (hur länge man varit specialist i ortopedi eller om man är ST-läkare)? Det tredje delarbetet är en kvalitativ studie där vi undersöker vilka uppfattningar det finns bland specialister i ortopedi eller ST-läkare i Sverige om individuell återkoppling av operationsresultat. I det fjärde och sista delarbetet i avhandlingen undersöker vi hur många operatörer som blir ”outliers” med avseende på oönskade händelser och risk för reoperation inom två år efter primär total höftprotes i Sverige.

Delarbete I, II och IV är registerbaserade studier baserad på data från Svenska Höftprotesregistret, den regionala vårdgivardatabasen i Västra Götalandsregionen (VEGA) (inte använt i delarbete II), lokal sjukhusadministrativa system samt det öppna Registret över legitimerad hälso- och sjukvårdspersonal hos Socialstyrelsen. Delarbete III är en kvalitativ intervjustudie där samtliga specialister i ortopedi eller ST-läkare anställda vid någon av det enheter som rapporterar till Svenska Höftprotesregistret inbjöds att delta i studien.

Syftet med delstudie I är att undersöka om operatörens årliga volym av primär total höftprotes pga. artros före operationen är associerad medoönskade händelser och död inom 90 dagar efter indexoperationen. Operatörens årliga volym är beräknat enligt formeln; antalet primära totala höftproteser 365 dagar innan operationen indexoperationen. 12 100 primär total höftprotes utförda på tio sjukhus i Västra Götalandsregionen utförda under åren 2007 till 2016 av 268 operatörer är inkluderade i analysen.

I denna studie har vi använt oss av logistisk regression (ojusterad och justerad) och i den justerade regressionsmodellen har vi justerat för patientdemografiska faktorer (ålder, kön, BMI, samsjuklighet), operationsfaktorer (orsaksdiagnos för operationen, typ av snitt, fixationsmetod) men också för sjukhus-och operatörsspecifika faktorer (sjukhusets årliga volym, operatörens antal år som specialist i ortopedi).

Resultatet i denna studie visar att om den årliga volymen av primär total höftprotes ökar med tio primära höftproteser reduceras risken för önskad händelse med tio procent och efter justering reduceras risken till åtta procent. För att beräkna hur risken för att drabbas av önskad ändelse inom 90 dagar efter operationen har ett 95 % prediktions intervall beräknats (Tabell 1).

Tabell 1. Predikterad risk för oönskad händelse inom 90 dagar efter index-operationen beroende på operatörens årliga volym.

Operatörens årliga volym, antal	Medelrisk, %*	95 % prediktions intervall*
0	8	7–10
10	8	6–9
20	7	5–9
30	6	5–8
40	6	4–7
50	5	4–7

*justerat för patienten ålder, kön, BMI, samsjuklighet, orsaksdiagnos, snitt, sjukhusets årliga volym.

Mortalitetstalen i studien var låga (0.2 %) och där kunde vi inte finna någon association mellan dödsfall inom 90 dagar efter indexoperationen och operatörens årliga volym. Resultaten i denna studie skiljer

sig inte ifrån tidigare utförda studier angående om årlig operationsvolym per operatör

I studie II var syftet med studien att undersöka om det fanns någon association mellan operatörens erfarenhet och patientrapporterat utfall ett år efter primär total höftprotesoperation. I denna studie har vi använt oss av operationer utförda under 2007 till 2012 vid samma tio sjukhus i Västra Götalandsregionen som i delstudie I. Operatörens erfarenhet är definierad i denna studie som antal år efter specialistbevis i ortopedi eller om inget specialistbevis är utfärdat klassificerades operatören som ST-läkare. Erfarenheten kategoriserades därefter i fyra grupper; 1) ST-läkare, 2) Specialist med mindre än åtta års erfarenhet sedan specialistbevis i ortopedi, 3) Specialist med åtta till femton års erfarenhet sedan specialistbevis i ortopedi, 4) Specialist med mer än femton års erfarenhet sedan specialistbevis i ortopedi.

I denna studie har vi använt oss av linjär regression (ojusterad och justerad) och i den justerade regressionsmodellen har vi justerat för; ålder, kön, BMI, ASA-klassificering, orsakdiagnos för operationen och Charnley-klassificering ett år postoperativt. Specialister med specialist med mer än 15 års erfarenhet har använts som referensgrupp i den linjära regressionen.

Resultatet i denna studie visade att det finns signifikanta skillnader i patientdemografi och i val av fixationsmetod mellan erfarenhetsgrupper. Denna skillnad patientdemografi och val av fixationsmetod är förväntad, då ST-läkare i Sverige först lär sig cementerad fixationsmetod (Golden Standard i Sverige) och övriga fixationsmetoder lär man sig först senare under sin ortopediska yrkeskarriär.

Den linjära regressionsmodellen visade efter justering att det inte finns någon skillnad i patientrapporterat utfall mellan specialister i ortopedi. Dock fann vi även efter justering att patienter opererade av ST-läkare rapporterar en lägre nöjdhet med operationsresultatet än patienter opererade av specialister med längst erfarenhet.

Delstudie III i avhandling hade vi som syfte att undersöka svenska operatörers uppfattningar om att bli försedd med sina egna resultat efter primär total höftprotesoperation. Denna delstudie är en fenomenografiska kvalitativ studie där vi har använt oss av enskilda intervjuer som datainsamlingsmetod. Alla intervjuer har skrivits ut ordagrant

och analyserats enlig de fenomenografiska analysstegen. I denna studie har vi bjudit in alla specialister i ortopedi samt ST-läkare anställda vid någon av det enheter som rapporterar till Svenska Höftproteser att delta. För att maximerat antalet uppfattningar som kunde finnas om fenomenet individuell återkoppling av operationsresultat i Sverige gjordes ett strategiskt urval av informanter. Detta urval baserat på vissa antaganden som vi i förväg trodde kunde påverka antalet uppfattning;

- Sjukhusnivå (Länsdelssjukhus, Länssjukhus, Universitets- eller regionsjukhus samt Privatsjukhus)
- Operatörens erfarenhet dvs. antal år sedan specialistbevis i ortopedi eller ST-läkare (samma uppdelning som i delstudie II).
- Operatörens kön (Man/Kvinna)

Sammanlagt utfördes det 19 intervjuer med specialister i ortopedi eller ST-läkare anställda vi 15 olika sjukhus i Sverige. Fyra uppfattningar om att bli försedd med sina egna resultat efter primär total höftprotesoperation: 1) Något som ger en möjlighet till individuella utveckling inom yrke, 2) Något som kan utsätta operatörerna för obefogad kritik, 3). Något som kan leda till försämrad patientsäkerhet, 4) Bidrar inte till förbättrad återkoppling till operatörerna.

I delstudie IV som är den sista delstudien i avhandlingen är syftet beskriva frekvensen av operatörerna som blir outliers på grund av oönskade händelser inom 90 dagar och reoperationer inom två år vid primär total höftprotesoperation i en svensk miljö samt undersöka effekten av en standardisering. I delstudie IV ingår operationer utförda under åren 2011–2016 vid samma tio sjukhus i Västra Götalandsregionen. Alla operationer som inkluderades i studien skall vara utförda pga. artros och operationen skall vara utförd med antingen cementerad-, hybrid-, -omvänd hybrid eller ocementerad fixationsmetod. I analysen inkluderades total 9482 primära totala höftproteser utförda av 208 operatörer. I denna studie har funnel plots för att visualiserad och plotta ut outliers. För varje operatör beräknas en standardiserad andel av ”utfallet” enligt formeln: antalet observerad oönskade händelser/förväntade antalet önskade händelser multiplicerat med det totala antalet händelser. För att beräkna det för används en logistisk regressionsmodell för ett bestämma sannolikheten att en händelse in-

träffar baserat på ett antal kovariater. Vi har använt oss av fem möjliga kovariater (patients ålder, kön, ASA-klassificering, BMI och orsaksdiagnos för operationen) som finns registrerade i Svenska Höftprotesregistret.

Resultatet i denna studie visar att andelen operatörer som blir outliers i Sverige är lågt både för oönskade händelser inom 90 dagar och reoperationer inom två år efter primär total höftprotesoperation. Vi har också i studien genomfört en subanalys där vi endast tog med operatörer som utfört mer än tio primära höftproteser och resultat av denna subanalys visade att alla outliers försvann efter denna standardisering.

Konklusioner från avhandlingen:

- En hög årlig operationsvolym per operatör är associerad med en reducerad risk för önskad händelse inom 90 dagar.
- Patienterna kan förvänta sig samma hälsorelaterade vinster, smärtlindring och nöjdhetsgrad ett år efter primär total höftprotesoperationen oberoende av operatörens antal år som specialist i ortopedi.
- Svenska ortopedspecialister och ST-läkare uppfattar individuell återkoppling av operationsresultat efter primär total höftprotesoperation från ett kvalitetsregister på flera sätt. Dels som ett system som kan bidra till individuella förbättringar och utveckling i yrket genom att varje operatör får vetskap om sina styrkor och svagheter. En farhåga är att individuell återkoppling av operationsresultat kan skada operatören om uppgifterna om operatörerna hamnar i fel händer eller misstolkas av patienterna. Återkoppling av operationsresultat uppfattas också som något som kan försämra patientnyttan och som onödig då all värdefull information redan kommer till operatörens kännedom.
- Det är ett lågt antal operatörer som skulle bli outliers avseende oönskade händelser inom 90 dagar och reoperationer inom två år efter primär total höftprotesoperation i en svensk miljö.

Background and introduction

History of hip arthroplasty

Total hip arthroplasty is considered to be one of the most successful orthopaedic interventions of its generation and has been proclaimed as "the operation of the century" in an article in *The Lancet* (Learmonth et al. 2007). The first attempt at hip replacement was made in Germany in 1891 by Professor Glück. In this first attempt, ivory was used to replace a hip joint destroyed by tuberculosis. In 1925, an American surgeon, Marius Smith-Peterson, created the first mould arthroplasty from glass. Despite being a fairly solid material, the glass failed to withstand the great force through the hip joint and shattered. Over the years, attempts were made to create hip arthroplasties from different materials, but, in 1953, an English surgeon, George McKee, used a metal-on-metal prosthesis on a regular basis. The English orthopaedic surgeon, Sir John Charnley (*1911-†1982), is regarded as the father of the modern total hip arthroplasty (Jackson J. 2011). In the early 1960s, his low-friction total hip arthroplasty was designed and this design is still used in principle today. The development of the prosthesis, the optimisation of the patient before surgery and the care after surgery has been an ongoing process since the days of Sir John Charnley. This progress and development have improved the outcomes following total hip arthroplasty and have been described in several scientific publications but also in annual reports from national quality registers. Despite all the efforts that have been made to reduce the risk of complications and revisions following total hip arthroplasties, some patients will still suffer. A successful outcome following total hip arthroplasty can be described from a surgical point of view as surgery without complications and any further surgery on the hip in question. However, from the patient's point of view, successful surgery takes place when the pain is reduced in the hip and patient mobility is restored (Bozic K J & Rubash H E. 2004).

The Swedish Hip Arthroplasty Register

Ernest Amory Codman (*1869-†1940) is regarded as the father of patient registers (Brand R A. 2009, Brand R A. 2013). His development of a sarcoma register (Codman E A. 2009) at the beginning of the 20th century was the first known medical register around the globe. In 1975, the Swedish Knee Arthroplasty Register was initiated and was the first national register of its type and was followed by the Swedish Hip Arthroplasty Register established in 1979. After the millennium, the Swedish government rediscovered the importance of registers and their use in the maintenance of quality control. Consequently, it began to support the start of new registers and produced structured mechanisms to be applied to financing and to register certification. This has resulted in more than 100 healthcare registers at different stages of establishment in Sweden.

The aim of the Swedish Hip Arthroplasty Register is to register all primary total hip arthroplasties, revisions and re-operations performed in Sweden. The coverage has been 100 per cent over the last 25 years and the completeness of primary total hip arthroplasties has been around 98 per cent during the last ten years (Kärrholm et al. 2017). The data that are collected in the Swedish Hip Arthroplasty Register contain information about the patient (age, gender, American Society of Anaesthesiologists classification, height, weight, Body Mass Index (BMI) and smoking habits), as well as surgical data (diagnostic indication for implantation, fixation technique, surgical approach, type of cement and type of implants). Data collection has been extended over the years.

The Swedish Hip Arthroplasty Register is widely recognised and its acceptance in the orthopaedic community has led to changes in practice that have resulted in the revision rates following total hip arthroplasties in Sweden being among the lowest in the world. The Swedish Hip Arthroplasty Register has three main assignments, like all other national quality registers in Sweden:

1. Analysing healthcare institutions and their activities
2. Stimulating continuous clinical improvements
3. Performing clinical research

In addition to these three overall assignments, the Swedish Hip Arthroplasty Register has a fourth assignment related to implants; post-marketing surveillance (Kärrholm et al. 2016).

In 2002, there was a development of the Swedish Hip Arthroplasty Register, when the patient-reported outcome measurement programme (Rolfson et al. 2011) was initiated. The success of the Swedish Hip Arthroplasty Register has led to the development of both regional and national quality registers around the globe (Table 1). This development of registers in other countries has led to extended collaboration between different registers and the establishment of more formal networks like the Nordic Arthroplasty Registers Association introduced in 2007 (Mäkelä et al. 2019) and the global International Society of Arthroplasty Registries established in 2005 (The International Society of Arthroplasty Registries. 2018).

Table 1. Examples of Hip arthroplasty registries around the globe.

Name	Country	Established in year
Swedish Hip Arthroplasty Register (SHAR)	Sweden	1979
Finnish Arthroplasty Register (FAR)	Finland	1980
Norwegian Arthroplasty Register (NAR)	Norway	1987
Danish Hip Arthroplasty Register (DHR)	Denmark	1995
New Zealand Joint Registry (NZJR)	New Zealand	1997
Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR)	Australia	1999
Romanian Arthroplasty Register (RAR)	Romania	2001

Kaiser Permanente National Implant Registries	USA	2001
National Joint Registry for England, Wales, Northern Ireland and the Isle of Man (NJR)	United Kingdom	2003
Slovak Arthroplasty Register (SAR)	Slovak Republic	2003
Dutch Arthroplasty Register (LROI)	Netherland	2007
Swiss National Joint Registry (SIRIS)	Switzerland	2007
American Joint Replacement Registry (AJRR)	USA	2009
Lithuanian Arthroplasty Register (LSER)	Lithuania	2010

Validity of the Swedish Hip Arthroplasty Register

A register-based study is never better than the register itself and is dependent on the quality of the data in that register. The validity of a (intervention) register is dependent on the coverage, the completeness (i.e. are all individuals included in the register) and the validity of the variables included (i.e. is all information on the individual collected and registered correctly). Coverage refers to the proportion of units reporting to the register and completeness refers to the proportion of individuals in the target population within the scope of the register (e.g. primary total hip arthroplasties), which are correctly included in the register. To evaluate the completeness in a register, there are several options:

- Compare the register of interest with another data source believed to be complete
- Compare the aggregated number of cases in a register with the total number in another source or, alternatively, calculate

the expected number of cases based on demographic data for a similar population

- Review patient charts to evaluate whether all patients are correctly classified in the register (a time-consuming and expensive method).

However, the demand for high completeness and validity depends on the research question(s). In studies investigating associations between an exposure and an outcome, the high validity of the outcome is more important than completeness, but, in descriptive studies and follow-up studies, both high completeness and validity are important.

Both the coverage and the completeness of the Swedish Hip Arthroplasty Register are of a high standard, with a reported coverage of 100 per cent and a completeness of approximately 98 to 99 per cent since 2010 (Kärrholm et al. 2019). The Swedish Hip Arthroplasty Register validates its completeness with the Swedish National Patient Register founded in 1964 (Ludvigsson et al. 2011) using the formula "*All total hip arthroplasties recorded in the Swedish Hip Arthroplasty Register divided by the total number of unique total hip arthroplasties recorded in the Swedish Hip Arthroplasty Register and the National Inpatient Register*".

The validity of the variables may be more difficult to evaluate in a large-scale register with multiple variables. The Swedish Hip Arthroplasty Register has implemented several steps to increase the validity of the variables. Examples are mandatory boxes in the web-based registration, automatically generated reports when one or more variable is missing or inconsistent, all units are supplied with a report on surgeries to reporting units to enable comparison and, if needed correction against the local hospital real databases.

The proportion of missing values in the Swedish Hip Arthroplasty Register is low, between one and two per cent of the elective total hip arthroplasties have missing data on ASA classification, BMI, fixation and articulation (Kärrholm et al. 2019).

For the patient reported outcomes measures programme, the pooled respondent rate for the preoperative questionnaire and the one-year follow-up questionnaire is between 68 per cent and 76 per cent for 2014 to 2017 for elective total hip arthroplasties with arthrosis as the

reason for surgery (Kärrholm et al. 2019). In 2011, (Rolfson et al.) reported a 79 per cent pooled respondent rate for both questionnaires.

Surgeon-specific outcome data and surgeon feedback programmes

Public reporting on the quality of the healthcare outcomes, providers' performance and patient experience varies between countries (Rechel et al. 2016). The public reporting of surgeon-specific outcome data was introduced by the National Health Service in the United Kingdom in June 2013. At the time of the introduction, nine surgical specialities and one non-surgical speciality were selected and this has been expanded to include other specialities since the introduction. The Society of Cardiothoracic Surgeons of Great Britain and Ireland had already started to publish the surgical activity and mortality rates of all consultants undertaking adult cardiac surgery in the United Kingdom in 2004, following high death rates in paediatric cardiac surgery at Bristol Royal Infirmary, where two surgeons had significantly higher mortality rates than their colleagues at comparable units.

One rationale for the publication of surgeon-specific outcomes could be to facilitate the patients' choice of surgeons in order to optimise their own care (Walker et al. 2013). The publication of surgeon-specific data is, however, controversial and it has both advocates and opponents. The advocates believe that the reporting of surgeon-specific outcome facilitates transparency and allows patients to make informed choices which might result in qualitative improvements. Critics argue that surgeons might become less interested in offering surgical treatment to high-risk patients (Radford et al. 2015). Data may also be manipulated to increase a patient's predicted risk or to make patients ineligible for public reporting, often referred to as "gaming" (Vallance et al. 2018). Some claim that there is a risk that the surgeon alone will become accountable for the outcome, not taking account of the impact of the entire team involved in the care of the patient. Furthermore, surgeons who treat high-risk patients might become discredited, if no case-mix adjustments are made, especially when the number of cases per surgeon is too low to enable a fair comparison of outcomes on a statistically basis (Walker et al. 2013).

Surgeon feedback programmes have already been introduced in some national arthroplasty quality registers around the world. The Scottish Arthroplasty Project surgeon feedback programme was introduced in 2003 (Macpherson et al. 2011). The aim of the Scottish Arthroplasty Project feedback programme is to encourage continual improvement in the quality of care given to joint replacement patients in Scotland by promoting positive change in individual surgeons. Since the introduction of the programme and until 2009, the Scottish Arthroplasty Project used Shewhart control chart methodology to present recent complication data and to identify variations. The use of a Shewhart control chart means that surgeons are presented with historical information relating to their performance during the presented years. In 2010, the Scottish Arthroplasty Project started using cumulative sum of outcomes (CUSUM) (Noyez L. 2009) methodology for feedback in real time. The statistical modelling used by the Scottish Arthroplasty Project only has access to routine data, so its case-mix adjustment is minimal and only takes account of age, gender, osteoarthritis and rheumatoid arthritis. A surgeon can therefore be identified as an outlier depending on complex case-mix factors.

The National Joint Registry for England, Wales, Northern Ireland and the Isle of Man programme for surgeon feedback was piloted in 2013 as a part of National Health Service England's surgeon-specific outcomes publication. The aim of the National Joint Registry for England, Wales, Northern Ireland and the Isle of Man Clinician Feedback is to provide orthopaedic surgeons with data and information relating to the care that has been delivered both by them and on their behalf and it enables an analysis of data within the wider context of hospital, sector (National Health Service and independent healthcare) and national benchmarks. The programme has been developed and expanded since the introduction (NJR Orthopaedic Consultant Outcomes Publication 2015) and is currently using risk-adjusted (age, gender and ASA classification) funnel plots as a statistical method (Noyez L. 2009) for visualising surgeons' performance and enabling surgeons to view trends made clear by the data held by the National Joint Registry for England, Wales, Northern Ireland and the Isle of Man.

The Australian Joint Register, Australian Orthopaedic Association National Joint Replacement Registry, introduced a feedback programme in 2017 with the aim of providing surgeons with an annual downloadable report on their individual outcomes which will assist them to understand their practice and help identify how they are performing compared with their peers. The Australian Orthopaedic Association has recommended access to a surgeon's individual reports for continuous professional development for those surgeons performing joint replacement. The statistical method used by the Australian Orthopaedic Association National Joint Replacement Registry is similar to the National Joint Registry for England, Wales, Northern Ireland and the Isle of Man's methodology, funnel plots adjusted for age and gender (Australian Orthopaedic Association National Joint Replacement Registry. 2017, de Steiger R N & Graves S E. 2019).

The effect on the quality of the results of the implementation of an individual surgeon feedback programme following total hip arthroplasties may be difficult to estimate.

The Scottish Arthroplasty Project, the register with the longest experience of surgeon-specific feedback, reports the number of outlier notifications in its Annual Report. In the 2019 Annual Report, it is stated that 44 surgeons received individual CUSUM outlier responses for 2018 (Scottish Arthroplasty Project 2019). This has provided them with an opportunity to reflect on their own practice and, if obviously justified, make the necessary changes.

However, to our knowledge, there are limited number of published reports aiming to investigate the effect of an individual surgeon feedback programme or how surgeons wish to be provided with their own results or the usefulness of the information with which they are provided. This lack of follow-ups of the feedback programme may be due to the relatively short period since the introduction of the programmes in the registers providing surgeons with information on their own results.

Surgeon volume

There are at least two ways of calculating surgeon volume. The most common methodology for calculating volume is to use "period methodology" (i.e. the number of surgeries performed during a specific period). All surgeries during the measurement period are attributed with the same volume regardless of whether it is the first or the last surgery. This period methodology of calculation has been used in several publications over the years. In 2014, Ravi et al. introduced alternative methodology for calculating surgeon volume and they defined annual surgeon volume as the number of procedures performed by the operating surgeon in the 365 days before the index procedure, Ravi et al.'s methodology for calculating volume might be slightly more correct if volume is regarded as a perishable but also when there is a large variation in producing surgeries between surgeons. The "period methodology for calculating volume" is also able to handle volume as a perishable if the investigated period is divided into shorter periods, weeks or months, and not annual volume; this shorter period might require a high proportion of orthopaedic surgeons performing a large number of surgeries every week.

The calculation of surgeon volume is not the only difficulty when it comes to comparing studies in the field of surgeon volume. There are also differences in follow-up periods, measured outcomes and covariates used for adjustment. The follow-up period varies between a few days and up to ten years. The measured outcome stretches from adverse events including diagnoses in the field of internal medicine (pulmonary embolism, infarction and so on) or local complication (dislocation, deep joint infection and so on) to re-operations and revisions related to the hip in question. Almost all earlier studies conclude that there is a relationship between surgeon volume and better outcomes following primary total hip arthroplasties (Lavernia C J & Guzman J F. 1995, Kreder et al. 1997, Solomon et al. 2002, Katz el al. 2001, Losina et al. 2004, Battaglia et al. 2006, Yasunaga et al. 2009, Paterson et al. 2010, Camberlin et al. 2011, Ravi et al. 2014, Ravi et al. 2014, Kurtz et al. 2016, Le Cossec et al. 2017, Koltsov et al. 2018, Murphy et al. 2019, Kishimoto et al. 2019), apart from Kreder et al.'s (1998) study where they did not find any association between annual surgeon volume and complications requiring hospital re-admission or mortality.

Surgeons' experience

The training of future orthopaedic surgeons is important in the healthcare system in order to supply hospitals with surgeons to meet the increased demand for total hip arthroplasties in the future (Kurtz et al. 2007, Nemes et al. 2014, Culliford et al. 2015, Inacio et al. 2017, Pilz et al. 2018). Concerns have been raised that "training surgeries" and the training of future orthopaedic surgeons may be associated with poorer outcomes and increased healthcare costs. There are some earlier publications dealing with the research question of whether the outcomes of surgery are dependent on the surgeons' level of experience (Moran et al. 2004, Palan et al. 2009, Inglis et al. 2013, Reidy et al. 2016, Wilson et al. 2016, Weber et al. 2017) and one systematic meta-analysis (Singh et al. 2019). In these studies, the surgeons' level is categorised and entitled supervised or unsupervised trainee, senior or junior trainee, consultant, senior surgeon and trainer, for example. The outcome measurement in these studies is not only patient-reported outcomes (Oxford Hip Score, Harris Hip Score, Western Ontario and McMaster Universities Osteoarthritis Index or EQ-5D) and complications but also radiological findings (component alignment, cementation), operation time and length of stay. The follow-up period varies between six and 120 months in these studies.

Almost all the studies revealed no significant association between the surgeon's level of experience and patient-reported outcome measurements, apart from the study by Inglis et al. (2013), which detected a significant difference in the Oxford Hip Score in favour of consultants compared with trainees. Palan et al. (2009) reported a superior Oxford Hip Score at five years in the consultant group compared with the trainee group. In this study, the consultants' patients reported higher scores on the Oxford Hip Score preoperatively.

Quantitative methods

Study design – pros and cons

Observational studies such as register studies obtain data from groups which have or have not been exposed to the subject of interest. The investigator does not intervene but simply "observes" and assesses the strength of the relationship between an exposure and an outcome variable. A register study can be categorised as a retrospective observational study graded as a level two (Burns et al. 2011) study in the Levels of Evidence Based Medicine. Prospective register studies are preferable when investigating the *effects of predictive risk factors* on an outcome. To study the *effects of an intervention*, randomised clinical trials are the gold standard. In randomised clinical trials, the participants are assigned to either an intervention or a control/placebo, preferably using blinded random selection. So, the choice of study design might depend on the research question, but there are naturally both advantages and disadvantages to these different study designs. In randomised clinical trials, the advantages are the unbiased distribution of confounders, the opportunity to blind the researcher and the fact that randomisation enables statistical analyses. The advantages of register studies are that they might be ethically more advantageous, due to the immediate availability of the data. This study design is comparatively less costly and shorter than randomised clinical trials (Song J W & Chung K C. 2010). Randomised clinical trials often include too few observations to investigate rare events. Conducting randomised clinical trials with sufficient statistical power to draw definitive conclusions is regarded as challenging (Lochner et al. 2001, Bhandari et al. 2002). The larger amount of data is one of the strengths of register studies, together with the opportunity to match study groups. Register studies also investigate the performance or effect of an intervention in the real world and not only in a specific clinical or laboratory environment within a selected group of patients. They therefore reduce the risk of performance bias and the results are often more generalisable. The disadvantages of register studies are that there are no

controls and there may be hidden confounders. In observational studies, there is also always the risk of bias: selection, detection and reporting bias (Song J W & Chung K C. 2010).

In Sweden, there is a long tradition of register-based studies linking different Swedish registers by the unique identifier, the personal identity number introduced in 1947, since when every individual residing in Sweden on a permanent basis has been assigned a personal identity number (Ludvigsson et al. 2009). The Swedish personal identity number is a useful tool for linkages between medical registers and enables 100 per cent coverage of the Swedish healthcare system.

Confounding factors can obscure the real effect of the exposure and it is therefore adequate to address confounding to make valid causal inferences from observational data. The term *confounding* refers to a situation in which a spurious association is found or a true association is missed between an exposure variable and an outcome variable as a result of a third factor or a group of factors, referred to as confounding factor/factors (Braga et al. 2012). A confounding variable is a factor associated with both the explanatory variable and the outcome variable. The problem with confounding is essentially a major limitation in observational studies. In large-scale, well-designed randomised clinical trials, the problem with confounding is partially solved by the inclusion criteria (baseline characteristics) in the study and by balancing both measured and unmeasured confounders between the treatment groups. However, problems with confounding factors in randomised clinical trials might occur, especially in small randomised clinical trials or if the inclusion criteria are poorly designed (i.e. not familiar with confounding factors).

A conventional method for identifying a confounder is to check whether the assumed confounder is associated with both the outcome variable and the explanatory variable and, secondly, to compare the association before and after adjusting for the confounding factors. Another possible method to help understand whether bias is potentially reduced or increased when adjusting for covariates is the graphical representation of causal effects between different covariates using directed acyclic graphs. However, the use of directed acyclic graphs in identifying confounding still relies on prior knowledge and assumed causal effects. As a result, it does not say anything about the truth of

your assumptions. It may well be possible that different researchers have different beliefs about which factor causes the other and this may result in different choices regarding factors to adjust for. Directed acyclic graphs can aid in this discussion among researchers by providing a visual representation to discuss causal research questions by making the underlying assumptions about causal mechanisms explicit.

How can a confounder be dealt with in observational studies? A simple answer to this question is: during the planning phase of the study. There are several ways of dealing with the confounding factor when planning a study; at the design stage and at the analysis stage. In the design stage, two different methods can be used to control for confounders: restriction or matching and, in the analysis stage, there are other methods: multivariable analysis, stratification or propensity score matching.

Restriction is one of the possible opportunities for dealing with confounders in observational studies at the design stage and it is a method that partially eliminates the influence of a confounder. This method restricts the study population to individuals with a certain characteristic by tightening the inclusion criteria in the study (= limiting the number of confounders at the analysis stage). The disadvantage of using this method is that patients with a special characteristic are selected to create a more homogeneous study population, but the generalisability of the study might be lost. Another disadvantage is that we also lose the opportunity to examine the association between the variables used in the inclusion criteria and the outcome. So, in the restriction method, only confounders known to be associated with the outcome should be selected.

Matching is the second possible opportunity for dealing with confounders in observational studies at the design stage and it involves pairing the study groups for potential confounding factors (age, gender, comorbidities and so on) and could be used in case-control studies and cohort studies. Matching enables adjustment for multiple confounding factors, provided that appropriate control patients can be identified. This method may be useful when there are a limited number of exposed patients, compared with non-exposed control patients. Normally, in order to increase the power of the study, the

matching of one exposed patient to more than one patient in the controls is performed. The disadvantages of this method are, the difficulty finding suitable matching pairs for multiple confounders. Second, variables selected for matching can no longer be investigated in later stages. Matching should therefore be based on variables that have a confounding effect but in which there is no interest in investigating.

Multivariable regression analysis is the most commonly used method to deal with confounders in healthcare studies (Petrie A & Sabin C. 2009). Explained briefly, a regression analysis is a mathematical model that estimates the association between a number of independent variables (risk factors) and one dependent variable (the outcome). When using regression models to estimate the change in the association, it is the amplitude of the change that is important and not the statistical significance. Multivariable regression analysis can be difficult to use if the study population is fairly small and there are a large number of confounders, because the acceptable number of confounders is dependent on the sample size. The rule of thumb is to have ten or more observations per variable. Other disadvantages of using regression models to deal with confounding factors in observational studies may be limited statistical knowledge or the assumptions of the models not being fulfilled.

Stratification is another method for dealing with confounders at the analysis stage and divides data as strata, based on possible confounding variables. The data are stratified and analysed for each stratum. Stratification is effective when dealing with categorical confounding variables, because the data can be split into two or more different strata. This method is therefore more difficult to use for continuous variables, because they must first be divided into arbitrary strata. There is a risk of residual confounding with few strata and, to reduce this risk, it is necessary to increase the observations in each stratum. The main disadvantage of stratification is the inability to deal with multiple confounding factors simultaneously, so it is most suitable to have one to two confounders, otherwise each stratum will be very small or disappear.

Propensity score analysis is the fourth method for dealing with confounders at the analysis stage and is defined as the conditional probability

of being treated, given the patient's risk factors (confounding factors), and it can be used to balance the difference between groups and thereby reduce bias (Braga et. 2012). This method is difficult to use with continuous exposure variables. Binary regression analysis produces an estimate of the probability of belonging to one group or another group and, once this probability is calculated for each patient, this score can be used to estimate the adjusted effect, through matching, stratification or regression. The main disadvantage of propensity score analysis is that the score does not account for unknown or unmeasured confounders. This problem is shared with any of the methods used for dealing with confounders in observational studies.

Limitations and strengths of registered-based studies

There are several limitations and strengths when it comes to using data from registers. In traditional epidemiological studies, the researcher collects the data on her/his own, while, in a registered-based study, the collection of variables is performed by a register (e.g. someone else collecting the data and entering the data in the register) and is extracted by the researcher. This might limit the control of the data collection for the researcher. Moreover, if the aim of the data collection is other than research, for example, if the data is being collected for billing purposes, there might be a possibility that the data will be salted with a secondary diagnosis. The reason for this is that the diagnosis influences the payment for the treatment (recall bias). Another limitation of using register-based data is that the researcher is limited to using the variables that have already been collected and might therefore lack information on important confounders. There is also a risk of variations in coding between persons and institutions that might influence the quality of the data. The meaning of "missing values" in register data can be difficult for a researcher to manage.

The quality of the data in registers might also be a limitation. A common way to validate data quality is to perform a cross-tabulation between different registers (Ludvigsson et al. 2011).

One risk associated with the use of register-based data rather than a limitation is that unimportant differences may become statistically significant, depending on the large sample size that is made available for

research. It is therefore important to interpret not only the significance level but also the size of the estimates and evaluate the clinical relevance.

The strength of register-based data is that the data already exist, which makes data collection faster and less expensive. A second strength in register-based studies is the large sample size in this kind of study and the subsequent great statistical power which makes it suitable to investigate both rare exposures and rare outcomes. Another strength of register-based studies is that they investigate the real world and, if the register has a high percentage of completeness, it minimises the effect of selection bias due to non-response and loss to follow-up (attrition bias). Large data registers with high completeness make it possible to conduct studies with the emphasis on small subpopulations in specific areas. The independent collection of data is a fourth strength when using data from registers and this often reduces the different types of recall bias and minimises a possible Hawthorne effect (Sedgwick P & Greenwood N. 2015). The fifth strength is related to the period between the exposure and the outcome. It is possible to observe various health problems which might occur or perhaps occur a long time after the time of exposure. By using data that have already been collected in registers, this latency between exposure and outcome is easier to investigate. The sixth and final strength when using register-collected data is the opportunity to adjust for some confounders available for the whole population (for example, educational level, income). This information is often registered with a very high degree of completeness.

One unique strength of register-based studies based on Swedish register data is the opportunity for an exact linkage using the personal identity number. The number is unique and follows each person throughout his/her life and the same personal identity number is never given to a new person.

Statistical methods

Kruskal Wallis H test

The Kruskal Wallis H test is a non-parametric test which is used when there are more than two groups to compare and it is an extended version of the more frequently used Mann-Whitney U test (Sedgwick P. 2015). Statistical significance when using the Kruskal Wallis H test indicates that at least one group dominates another sample. A non-parametric test does not assume normally distributed data and can be used to analyse data measured on a continuous or ordinal scale. The Kruskal Wallis H test was used in Paper II in this doctoral thesis.

Pearson's chi-square test

Pearson's chi-square test is used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories. Pearson's chi-square test was used in Paper II in this doctoral thesis.

Linear regression

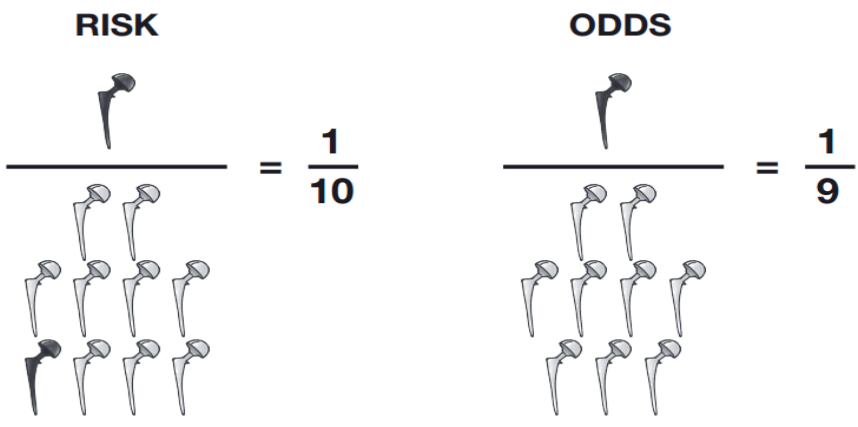
Linear regression (Lee at al. 2014) is used to describe the relationship between two continuous variables, a dependent (response) variable (y) and one or more independent (predictor or explanatory) variables. The regression coefficient (β) describes how the dependent variable relates to the explanatory one(s). As a result, the regression coefficient (β) denotes the expected change in the dependent (response) variable (y) for a one-unit change in the independent (predictor or explanatory) variable (x) when the other covariates are kept fixed. Linear regression was used in Paper II in this doctoral thesis.

Logistic regression

A logistic regression (Wiest et al. 2015) model is used when the outcome (dependent variable) is categorical (for example; dead/alive, win/lose, adverse events/no adverse events) and estimates the probability of a dependent (response) variable (y) based on one or more independent (predictor or explanatory) variable(s). It is more easily described as the probability of a binary response based on one or more predictor (or independent) variables (features) and it allows one to say that the presence of a risk factor increases the odds of a given outcome by a specific factor.

The results of the logistic regression can be presented in different ways. Outcomes can be presented as either a risk: the proportion of individuals exhibiting the outcome of interest in a particular group calculated by dividing the number of cases (typically participants with a poor outcome) by the total number of participants in each group or as odds; the probability of the event of interest occurring compared with the probability of the event not occurring calculated by dividing the number with a poor outcome by the number with a good outcome (odds = p/(one-p)) (Figure 1).

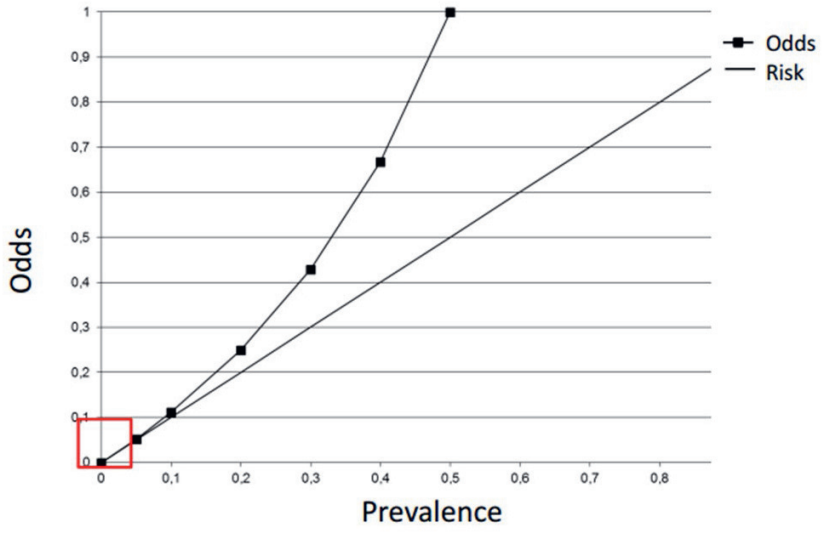
Figure 1. A visualization of different calculation methodology used for presenting outcomes



The **risk** of the condition is one in ten= ten per cent. The **odds** of the condition are one to nine=eleven point one per cent

The odds are approximately equal to risk in "rare events data". For odds of less than about 20 per cent, the odds are not greatly dissimilar to the risk, but, if the risk climbs above 50 per cent, the odds start to look very different (Davies et al. 1998) (Figure 2). Logistic regression was used in Papers I and IV in this doctoral thesis.

Figure 2. Odds is approximately equal to risk in "rare" events data



Confidence Interval

The confidence interval tells you how well the population mean has been determined (Perry et al. 2017). There are several options for calculating the confidence interval; Wald's, Copper-Pearson's and Wilson's and so on and some of these methods are used for calculating confidence interval when there is a small number of observations (Brown et al. 2001). This is visualized in Table 2. So, not choosing the most appropriate method depending on the number of observations might have an effect on the confidence interval limits. Confidence interval was used in Papers I, II and IV in this doctoral thesis. Wilson method is used for calculating the confidence interval in Paper IV.

Table 2. Differences in per cent for upper 95% CI with high and low number of observations (n) when using three different methods for calculating confidence interval, when the p-value is=0.04.

Method for calculating CI	Number of observations, n	Upper 95% CI limit	Difference in %
Wald's	50	0.094	
Wilsons	50	0.135	43.6
Copper-Pearson's	50	0.137	45.7
Wald's	2650	0.047	
Wilsons	2650	0.048	2.1
Copper-Pearson's	2650	0.048	2.1

CI=confidence interval

Prediction Interval

The prediction interval is an estimate of an interval in which future observations will fall, with a certain probability, given what has already been observed. The difference between prediction intervals and confidence intervals is that confidence intervals tell you how well you have determined the mean, while the prediction interval (a forecast) is an estimate of an interval in which future observations will fall, with a certain probability, given what has already been observed. The prediction interval is always wider than a confidence because the prediction intervals must account for both the uncertainty about knowing the value of the population mean, plus data scatter. The prediction interval was used in Paper I in this doctoral thesis.

Funnel plots

Funnel plots (Spiegelhalter D J. 2005) were used to visualise outliers. Constructing funnel plots involves plotting data on a scatter plot and then superimposing "control limits" around the data points. The control limits typically represent two and three standard deviations from

the mean. The control limits are dependent on the sample size; a small sample size increases the control limits, while a larger sample size reduces the limits (i.e. surgeons who undertake few surgeries will have a wider control limit). The X (horizontal) axis represents volume: the total number of primary total hip arthroplasties recorded for each surgeon during a period, while the Y axis is a measurement of performance given by the standardised proportion, for example; age, gender and ASA classification. Funnel plots were used in Paper IV in this doctoral thesis.

Directed acyclic graph

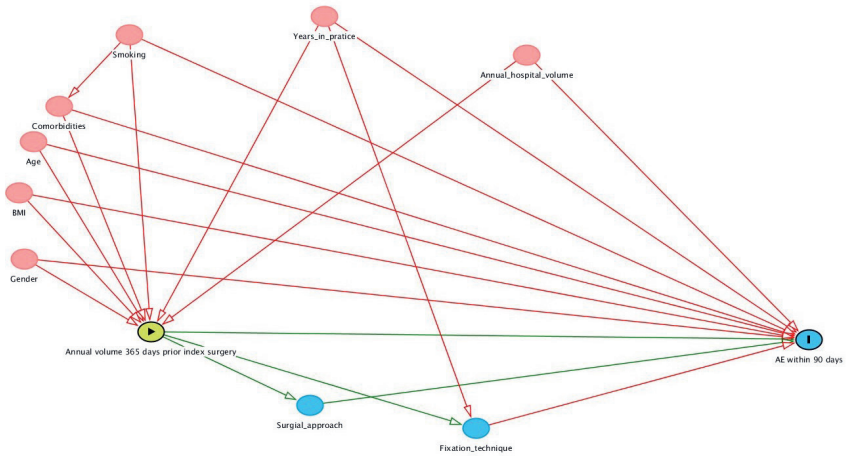
A graph is called directed if all the variables in the graph are connected by arrows. Arrows in directed acyclic graph represent direct causal effects of one factor on another, either protective or hazardous. Directed acyclic graphs can therefore help to identify the presence of confounders and one way to resolve the problem is to identify confounders. The arrows and their direction are based on à priori knowledge of the area of interest. When a directed acyclic graph contains all the relevant variables and their causal relationships (exposure, outcome and confounders), the presence of 'confounding' in general can be identified. The mechanisms in directed acyclic graphs (Suttorp el al. 2015) are as follows.

- A path is a sequence of arrows, irrespective of the direction of the arrows.
- A directed path is a sequence of arrows in which every arrow points in the same direction, representing the causal relationship.
- A backdoor path is a sequence of arrows from exposure to outcome that starts with an arrowhead towards the exposure and ends with an arrowhead towards the outcome.
- Two factors are associated if they are connected by an open path.
- A collider is a common effect; a factor on which two arrowheads collide and the collider blocks the path. However, a collider that has been conditioned no longer blocks a path but might introduce a form of selection bias.

- Blocked paths do not affect the direct causal relationship between the exposure and the outcome.
- Confounders are identified by an open backdoor path.
- The causal relationship between exposure and outcome will no longer be confounded if the only open paths from exposure to outcome are directed paths from exposure to outcome.

Directed acyclic graphs were used in Paper I in this doctoral thesis to visualise and determine covariates of interest (Figure 4)

Figure 4. Directed acyclic graph used in Paper I for visualising and determining covariates of interest.



Qualitative methods

Phenomenography

Phenomenography is a research approach developed by a team of researchers led by Professor Emeritus Ference Marton at the Department of Education at the University of Gothenburg in the 1970s. The aim of phenomenography is to discern and describe different ways of experiencing phenomena in the surrounding world. Phenomena are the central concept in phenomenographic research and can be divided into two words; phenomena (from the Greek word phaino= to bring into the light, cause to shine) and graphs (=describing). What is a phenomenon? Simply described, it is an object as it is presented to us. It could be a more concrete, like a hip prosthesis or fracture, but it could also be something more abstract, like postoperative pain. There is a distinction between a phenomenon and a situation, according to Marton (Marton F & Booth S. 1997). A situation is always experienced in a special context, a time and a place. A phenomenon, on the other hand, is experienced as being abstracted from or transcending an anchorage to time and space and is in a higher level of abstraction. Two other central words in phenomenographic research are *first- and second-order perspectives*. In the first-order perspective, the researcher is interested in how something really is and, in the second-order perspective, the researcher is primarily interested in how a phenomenon is conceived.

The outcomes of a phenomenographic study, according to Dahlgren & Fallsberg (1991), represent an attempt to describe and understand the nature of variations as regards the world as it is conceived.

In phenomenographic research, the preferred method of data collection is the semi-structured interview with a guide containing a few entry questions. All the interviews are recorded and should be transcribed verbatim. The qualitative analysis phase focuses heavily on reading and re-reading the transcribed material. Dahlgren & Fallsberg (1991), but also Sjöström & Dahlgren (2002) and Stenfors-Hayes

at al. (2013), describe a number of activities (steps) for making the sometimes difficult phenomenographic easier to handle.

The data analysis steps according to Sjöström & Dahlberg (2002) are as follows.

1. The first step can be called familiarisation and means that the researcher is introduced to the material by reading through the transcripts. Necessary for correcting errors in the transcript.
2. The second step involves the compilation of answers from all the respondents to a certain question. The main task here is to identify the most significant elements in the answers given by each informant.
3. The third step is a condensation or reduction of the individual answers to find the central parts of longer answers or a dialogue.
4. The fourth step contains a preliminary grouping or classification of similar answers.
5. The fifth step is a preliminary comparison of categories, where the researcher tries to establish borders between the categories.
6. The sixth step consists of naming the categories to emphasise their essence.
7. The seventh and last step is a contrastive comparison of categories, which contains a description of the unique character of every category, as well as a description of resemblances between categories.

Selection of informants

The recruitment of informants to a qualitative study is not based randomly (as it is in a quantitative study) and might be based on assumptions relating to attributes (age, gender, employment, years in practice and so on) of the informants to collect as broad a range of descriptions or experience of the phenomena as possible. It is more important to find a smaller number of informants with a maximum of variation in experience to obtain a rich description of the phenomena.

A larger homogeneous group will probably generate only huge data material, which makes it more complicated to analyse.

Different methods used in selection of informants in qualitative research are as follows.

- **Convenience sampling:** a method in which, for the sake of convenience, the study informants that happen to be available at the time of data collection are selected in the sample.
- **Purposive sampling:** a method that aims to select study informants which represent a wide range of variation in dimensions of interest.
- **Snowball sampling:** a method in which a researcher starts by identifying some (at least two) individuals who are relevant to the study and then asking them to locate other useful informants.
- **Sampling contrasting cases:** a method that is useful in comparative studies that aim to explain problems by establishing which factors are associated with them or cause them.
- **Stratified sampling:** a method used to capture major variations rather than to identify a common core.

Credibility, validity and reliability

Credibility in a phenomenographic study focuses largely on how the researcher is able to describe the relationship between the data and the categories for describing the phenomena. The credibility of the study is based on the how well each step in the research process is described (presentation of the interview questions, analysis process and the conclusion) to make it possible to replicate the study. One aspect of the validity of data analysis in qualitative studies is that if another researcher were to suggest the same categories using the same data material. To handle this issue of replicability, Marton F (1988) introduced intersubjective agreement. Another researcher applies the suggested categories and classifies the interview excerpts, after which a degree of agreement is determined. Generally, this degree of agreement varies between 60 and 100 per cent (Larsson S.1986).

Objectives

The overall aim of this project is to develop a methodology for providing feedback to individual surgeons following primary total hip arthroplasties and to explore the potential effect of this methodology being implemented. We posed the following questions:

- Are adverse events and mortality dependent on number of surgeries performed each year by the surgeon?
- Are patient-reported outcomes dependent on the surgeon's experience?
- Which perceptions exist in the orthopaedic society about the phenomenon of feedback of individual surgeon's results from a national quality register?
- Which factors need to be taken into account to develop a feedback system in a Swedish setting?

Patients and methods

Paper I

In this study, we collected information on primary total hip arthroplasties performed between 2007 and 2016 from ten hospitals in western Sweden. These data were linked with the Swedish Hip Arthroplasty Register and a regional patient register (Vega). A total of 15,086 surgeries were extracted from hospital medical records and 2,986 have since been excluded (reason for surgery not osteoarthritis, incision other than posterior or directly lateral, data on operating surgeon not available in local medical records, no information on volume 365 days prior to index total hip arthroplasty, missing data on BMI). 12,100 primary total hip arthroplasties performed due to osteoarthritis by 268 surgeons were identified and used in the analysis. The surgeon volume in the study is defined as the number of primary total hip arthroplasties performed by the operating surgeon in the 365 days prior the index procedure (Ravi et al. 2014).

Paper II

In this study, we collected patient characteristics and surgical information on primary total hip arthroplasties performed between 2007 and 2012 from ten hospitals in western Sweden. These data were linked with patient-reported outcomes (EQ-5D-3L, Satisfaction VAS, Pain VAS) in the Swedish Hip Arthroplasty Register. For each surgeon involved, data on the year of obtaining a license to practice and/or a specialist certificate in orthopaedics were obtained from publicly available data from the Swedish National Board of Health and Welfare register of licensed healthcare professionals. Eight thousand one hundred and fifty-eight primary total hip arthroplasties due to osteoarthritis were identified. We identified the surgeons' level of experience in 8,116 total hip arthroplasties. Data from the Swedish Hip Arthroplasty Register on pre- and postoperative patient-reported outcomes and satisfaction at one year were available for 6,713 total

hip arthroplasties. Our definition of experience of the surgeon is based on the years between orthopaedic specialist certification and the index total hip arthroplasty. The surgeons' level of experience was divided into four subgroups related to experience: less than eight years, eight to fifteen years and more than fifteen years of clinical practice after a specialist certificate and trainees. If no specialist certificate was obtained, the surgery was classified as trainee surgery. Surgeons with more than fifteen years' experience as an orthopaedic specialist were used as a reference group in the analyses.

Paper III

Paper III is qualitative study using phenomenography as the research method. All the surgeons in the specialist and trainee groups performing primary total hip arthroplasties in the 74 orthopaedic departments reporting primary total hip arthroplasties to the Swedish Hip Arthroplasty Register were invited to participate in the study. A purposive sampling of informants based on the experience of the surgeon (same sectioning as in Paper II), gender and hospital level (rural, county, private or university hospital) was made. If there was multiple interest in participating in the study, a secondary selection was made based on hospital location for geographic spread among the informants. Nineteen orthopaedic surgeons (15 orthopaedic specialist and four trainees) from 15 orthopaedic departments in Sweden were selected for a semi-structured interview. All the interviews were transcribed verbatim by the first author and were individually reviewed by the first author and an external reviewer to ensure the quality of the transcript before analysis. The data analysis procedure was followed step by step, as recommended in phenomenographic data analysis. The data analysis begins with reading through the interviews several times, thereby obtaining an overall idea of the material, as well as facilitating an understanding of the meaning of the statements. The next step in the data analysis was to look for statements containing different perceptions. Attention focused on qualitative similarities and differences in the data. By comparing differences, a perception becomes manifest and it is possible to detect the characteristic features in one perception thanks to its contrast to other perceptions.

Paper IV

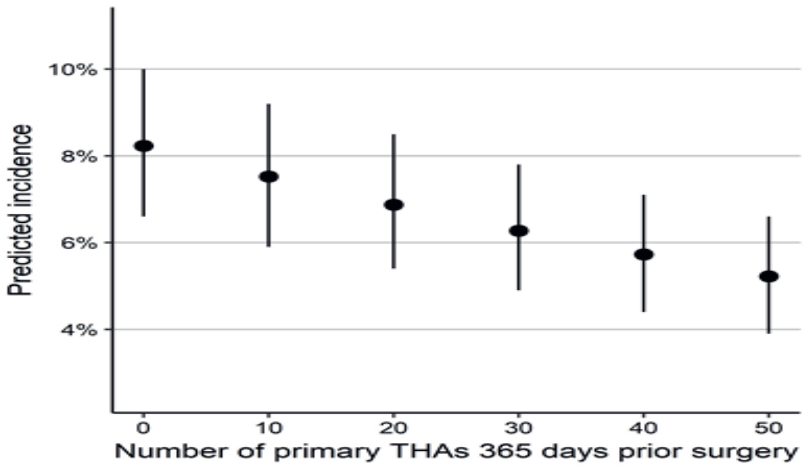
In this study, we used data from the same dataset as in Paper I. In the analysis, however, we used information on total hip arthroplasties performed in 2011-2016. 208 surgeons and 9,482 total hip arthroplasties were used in the analysis. A standardised proportion is calculated for each surgeon as the ratio of the number of negative observed events to the number of expected events, multiplied by the overall proportion of events. To calculate the expected number of events, a logistic regression model is used to determine the probability of an event based on a set of confounders (age, gender, ASA classification, BMI and diagnostic indication for implantation). Funnel plots with control limits based on 95% CI were used to detect outliers. The definition of outliers is a surgeon above the upper 95% CI.

Summary of results

Paper I

The median annual surgeon volume was 23 primary total hip arthroplasties (range zero to eighty-two) 365 days prior to the total hip arthroplasty of interest and the mean risk of adverse events within 90 days was seven per cent. If the annual volume increased by ten primary total hip arthroplasties in the simple logistic regression, the risk of adverse events decreased by ten per cent and, in the adjusted multiple regression, the corresponding number was eight per cent (Figure 4).

Figure 4. Adjusted prediction intervals for adverse events within 90 days.



The multiple regression has been adjusted for age, sex, BMI, comorbidities, years in practice as orthopaedic specialist at the time of the index total hip arthroplasty, fixation technique, diagnostic indication for implantation, surgical approach, and annual hospital volume.

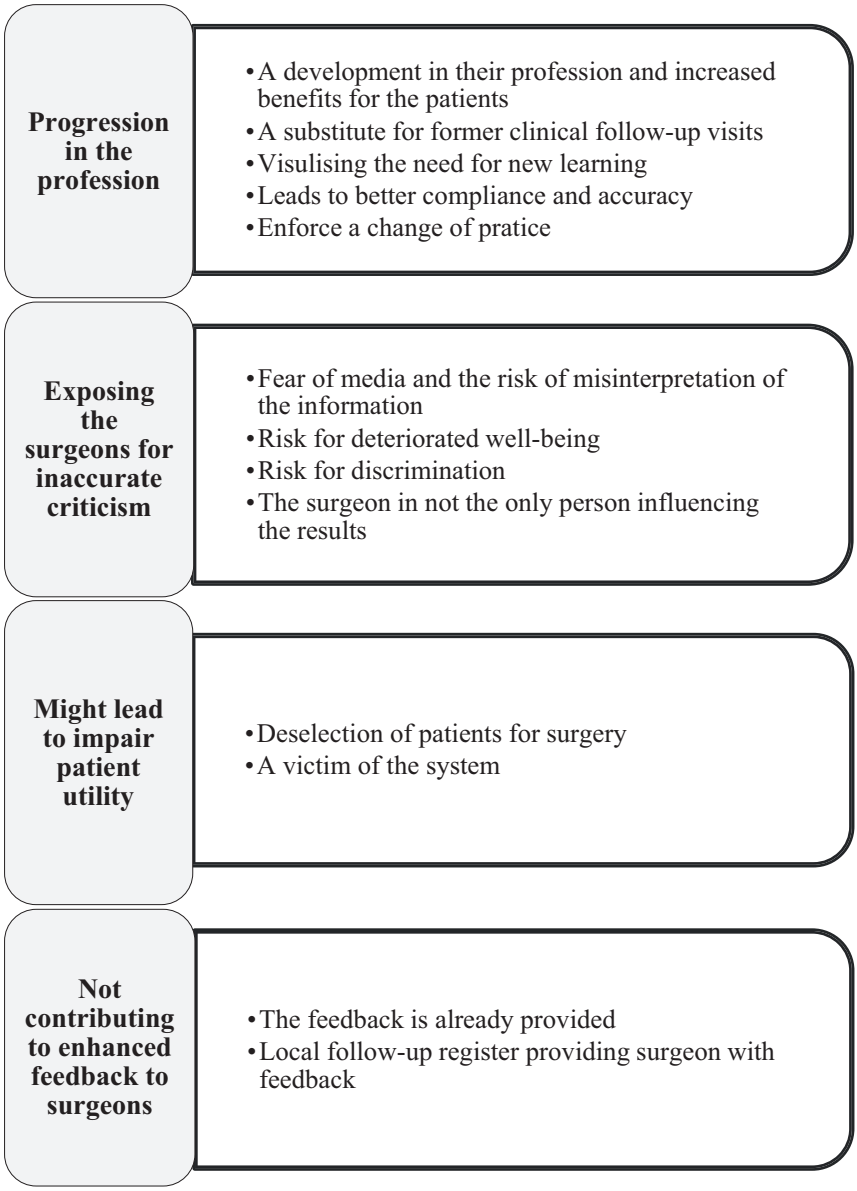
Paper II

We observed a statistically significant difference between the four groups of surgeons in terms of mean patient age, ASA classification, Charnley classification, diagnosis and fixation technique. At the one-year follow-up, there were no statistically significant differences in the Pain VAS, EQ-5D index, or EQ VAS between the subgroups of orthopaedic specialists.

Paper III

Nineteen orthopaedic surgeons and trainees from 15 hospitals were recruited and interviewed in the study. The phenomenographic data analysis of the collected and transcribed material outlined four categories of description on the phenomena individual surgeon feedback from a national quality register expressed by the informants: 1) progression in the profession; 2) exposing the surgeons for inaccurate criticism; 3) might lead to impaired patient utility; 4) not contributing to enhanced feedback to surgeons. These categories of description are presented in a in Figure 5.

Figure 5. Various way of understanding individual surgeon's feedback from a national arthroplasty quality register, formulated as descriptions categories.



Paper IV

In observed funnel plots, the percentage of outliers was low for both adverse events within 90 days (zero to five per cent) and re-operations within two years (zero to one per cent). In the standardised models, the corresponding numbers were even lower (adverse events zero to three per cent/re-operations zero to one per cent). A sub-analysis was conducted in which the limit was restricted to more than ten total hip arthroplasties annually in order to be evaluated. The result of this sub-analysis showed that the surgeons who are outliers for adverse events within 90 days are reduced by more than half or disappear, only depending on this restriction, despite the fact that no standardisation has been made. After standardisation, all the outliers disappeared. For re-operations within two years, all the outliers disappeared, apart from one surgeon in 2016. However, after standardisation, this outlier also disappeared.

Limitations, strengths and bias

Specific strengths and limitations in Paper I

Paper I share the same limitations as all observational studies using administrative data with regard to both changes in practice during the study period and also local trends and differences in the registration of complication diagnoses if/when they occur. Another limitation in Paper I is that we used a regional patient register as a data source for the calculation of both the comorbidity index and the adverse events. This regional patient register is not validated on its own but provides data to the Swedish National Patient Register of which the validated Swedish National In-patient Register is a part. The Swedish National In-patient Register contains 99 per cent of all hospital discharges (Ludvigsson et al. 2011) and, in this study, the adverse events definition is conditional on requiring a hospital admission. However, this regional patient register can also be a strength in the study, as it records all re-admission and healthcare for the inhabitants in the western region, regardless of healthcare providers.

In this study, we have not been able to adjust for patient smoking habits, which are known to influence adverse events negatively. However, we decided not to adjust for this known confounder in the multiple regression analysis because of the large number of missing values (70 per cent) for patient smoking habits.

One strength of this study is that we have been able to adjust for important confounders, fixation, surgical approach and the experience of the surgeon (i.e. years practising as an orthopaedic specialist) at the time of surgery. This is the first time the surgeons' experience calculated as years in practice has been used as a confounder in a multiple regression model. Earlier studies have used other definitions of surgeons' experience in their statistical model; years since graduating from medical school (Katz el al. 2001) or age of the surgeon (Ravi et al. 2014). In our definition, experience can be regarded as a proxy for

surgical skills accumulated through the experience of previous procedures during the surgeon's career and knowledge acquired of preparing patients both mentally and physically prior to the surgery. More experienced surgeons are likely to make more appropriate decisions regarding the indication for surgery, the surgical details (technical aspects) and other perioperative factors that could result in an improved outcome. Compared with the other two definitions of surgeon's experience, our calculation based on years as an orthopaedic specialist appears to be more reliable, as we know the exact number of years since the orthopaedic specialist obtained certification. Both the other definitions of surgeons' experience contain uncertainty relating to actual years as orthopaedic specialists.

There are also disadvantages to our definition. Surgeons might have longer or shorter experience of total hip arthroplasties before becoming an orthopaedic specialist. Some may start with total hip arthroplasty immediately after examination, whereas others may have worked within other orthopaedic subspecialities for years before starting total hip arthroplasty and some might have spent varying amounts of time on administrative work, reducing their time period as active arthroplasty surgeons.

Specific strengths and limitations in Paper II

Paper II shares the same limitations as all observational studies using administrative data. One specific limitation in Paper II is that we lack information on confounders known to influence patient-reported outcomes. We have not been able to adjust the multiple regression for socio-economic factors or patient educational level. These confounders are not available in the registers used in this study. This lack of information may affect satisfaction if there is a skewed distribution of patients with a good economic situation and a high educational level in any of the groups.

Another limitation of this study is the lack of information on the quality of the preoperative information given to the patient. This is almost impossible to examine in an observational study based on register data.

We have no knowledge of the level of supervision of trainees in this study. The level of supervision may vary between completely unsupervised and fully supervised individuals. Earlier studies have, however, not reported any substantial difference in patient-reported outcomes after up to ten years of follow-up between unsupervised and supervised trainees (Reidy et al. 2016). It is therefore also possible that this potential confounder had a limited influence in our study.

Despite the fact that we have not been able to adjust for all the important confounders known to influence patient-reported outcomes, one strength of this study is that we have been able to adjust for both Charnley classification and comorbidity known to influence patient-reported outcomes (Rolfson et al. 2011).

Our definition of surgeon's experience is another strength of this study. It can be regarded as a proxy for surgical skills accumulated from the experience of previous procedures and knowledge acquired from preoperative assessments and observations of uneventful postoperative healing or treatment of more or less serious complications.

Specific strengths and limitations in Paper III

One limitation of Paper III is the lack of recruitment of informants at university or regional hospitals and private hospitals. This lack of recruitment might depend on the small number of hospitals at these two hospital levels compared with rural or county hospitals in Sweden. Another explanation of the lack of recruitment may be that these hospitals already have follow-up programmes for their patients depending on the case mix and/or more technically demanding total hip arthroplasties and may therefore not have reflections of register-based feedback programmes. According to Larsson & Holmström (Henricsson M (editor). 2018), 20 interviews are enough in a phenomenographic study, but their empirical knowledge reveals that even studies with fewer than 20 interviews are able to generate important results and no more perceptions are generally revealed after ten to twelve interviews.

One of the strengths in Paper III is the method used for data collection and it is suggested that a semi-structured interview is the correct method to use in a phenomenographic study. However, there are two

concerns attached to phenomenographic interviews. The first is the informant's motivation to participate in the study and the second is related to our understanding of what the informant is trying to convey. Regarding the second concern, there is a need during the interview immediately to interpret what the informant is saying in order to be able to decide about further questioning or exploration. Any misunderstanding in this interpretation may jeopardise the quality of the interview data.

Specific strengths and limitations in Paper IV

Paper IV shares the same limitations as all observational studies using administrative data. Paper IV also shares the same limitation as Paper I in using an unvalidated regional patient register as a data source and we used the same definition for adverse events as in Paper I. Some of the included surgeons may have had temporary or partial employment, having performed primary total hip arthroplasties outside the investigated region. This lack of information on total hip arthroplasties performed outside the region is a limitation in this study. There could also be a risk that an unknown confounder, apart from the five confounders evaluated in this study, might affect the outcomes following primary total hip arthroplasties.

The strength of Paper IV is the low frequency of missing values for confounders and the fact that we were able to adjust for confounders influencing adverse events and re-operations, known from earlier studies. To the best of our knowledge, this is the first time that standardisation for BMI and ASA classification have been made on funnel plots at surgeon level.

Ethical considerations

Register-based research does not require written consent from the participants. This lack of a need for written consent is based on both the EU directive (Data Protection 95/96) and the Swedish Patient Data Act (SFS 2008:35). The Swedish Patient Data Act obliges the healthcare providers to inform the patients that data will be registered and used for quality improvements and research. Written information

about the collection of data for quality registers is given to patients before the surgery. Patients are also informed that they can opt out at any time and that their data will be deleted from the register. Information on ongoing research is provided by the register in question and is sometimes posted on the quality register's webpage. However, there is a requirement for approval from an ethical review board to conduct register-based research equal to all other research.

Paper I

Ethical approval for Paper I was obtained from the Regional Ethical Board in Gothenburg (reference number 141-14) after an appeal to the Central Ethical Review Board (reference number Ö 9-2016).

Paper II

Ethical approval for Paper II was obtained from the Regional Ethical Board in Gothenburg (reference number 205-16) after an appeal to the Central Ethical Review Board (reference number Ö 11-2016).

Paper III

Ethical approval for Paper III was obtained from the Regional Ethical Board in Gothenburg (reference number 1057-18).

Paper IV

Ethical approval for Paper IV was obtained from the Regional Ethical Board in Gothenburg (reference number 141-14) after an appeal to the Central Ethical Review Board (reference number Ö 9-2016).

Discussion

Surgeon volume

There is no consensus regarding the optimum definition of a low volume surgeon and the best way to define this concept is open to debate. The present inconsistency in the categorisation of this parameter could therefore be expected. Differences between studies of the influence of surgeon volume in terms of follow-up periods and outcomes will further jeopardise comparisons. As a result, a low-volume category in one study may be a medium-volume category in another and in a third study, the same volume may be categorised as a high-volume category. This problem of definition between different volume categories is not unique in orthopaedic surgeon volume studies and has been highlighted in a systematic review by Chowdhury et al. (2007), exemplified by studies of angioplasty. Another inconsistency is the problem of different follow-up periods, which may vary up to four years. Finally, different outcomes such as adverse events, mortality, revision, payment or a specific complication have been used.

There are numerous possible scenarios to consider when defining low- and high-volume surgeons. A person could be a low-volume surgeon for primary total hip arthroplasties, but the same surgeon could be a high-volume producer of total hip arthroplasty revisions, or a low-volume surgeon for primary total hip arthroplasties but a high-volume producer of primary total knee surgery. So, determining the limit for the optimal surgeon volume of total hip arthroplasties is difficult, based on the findings in these studies. However, almost all studies investigating the association between annual surgeon volume and primary total hip arthroplasties conclude that increasing surgical volume is associated with a more or less enhanced quality of the outcome.

In the dataset used for Paper I, there is a trend towards a higher proportion of high-volume surgeons and fewer low volume surgeons

performing of primary total hip arthroplasties for year 2014-2016. This trend is in line with most previous studies. At present we do not know if there is a similar trend in other regions in Sweden, because this information is lacking. To obtain this information, we need access to all local hospital records in hospitals producing primary total hip arthroplasties. This might be feasible, although a significant administrative effort will be needed to merge the data into one dataset for analysis. The central registration of the surgeon in the Swedish quality registers for each primary total hip arthroplasties could be a solution to this problem.

Another question is whether Swedish orthopaedic surgeons need to acquire more surgical total hip arthroplasty skills by spending more time performing surgery and if so, how this should be organised? More time in the operating room will increase the annual volume and most probably surgical skills. One possible solution to increasing the annual volume and acquiring more skills could be to establish surgical training centres which are responsible not only for the training of individual surgeons but also for the entire surgical teams. These surgical training centres might contribute to a more uniform training programme for primary total hip arthroplasties but they should also be responsible for the development of surgical techniques and the care. To supply the surgical training centres with staff, there could be a mix of local employees and assigned staff from the hospital connected to the surgical training centres.

A third question of interest in the scope of annual surgeon volume is whether it is possible to maintain surgical skills if a surgeon has been a high-volume surgeon and is now a low-volume surgeon or does not produce any surgery at all for a period? This might be problematic to visualise in a research study. A high-volume surgeon has probably been exposed to more technical problems and has therefore been trained to solve them in situ. For how long a period will these skills last if the surgeon is not exposed to them anymore or how long will the "new" learning curve be for getting back on track? This is a million-dollar question without a good scientific answer, but most probably there will be a great variation between surgeons.

Limitations when quantifying surgeon experience

There are a limited number of studies examining the association between the experience of the surgeon and the patient-reported outcome measurements. Moreover, of these studies, there is only one presenting statistically significant results in favour of consultants (Inglis el al. 2013). One general limitation in some of these studies might be the categorisation of the supervision of trainees. The definition of supervision varies between the studies and is not explicitly described. Empirically, the supervision of students, trainees or in other education programmes may vary a great deal, depending on the experience and seniority of the trainee, as well as the relationship between trainer and trainee. Supervision might be a spectrum rather than a binary value. A better calculation for the categorisation of experience is the arbitrary heterogeneous calculation, depending, for example, on years as a trainee or a consultant. This arbitrary heterogeneous calculation may have other limitations; for example, there may be a variation in aims for each year between both individual and hospital trainee programmes and they might therefore not be completely comparable. There is also a difference in the covariates that are used in the studies of surgeons' experience and patient-reported outcome measurements. Some of these studies adjust for baseline (Moran et al. 2004, Reidy et al. 2016, Weber et al. 2017) data on the outcomes in question.

Limitations of confounders in observational studies

A statistical phenomenon (Tu et al. 2008) – *the reverse paradox* – might occur if the outcome and explanatory variables are the same and have important implications for the interpretation of evidence from observational studies. Simpson's paradox, Lord's paradox and suppression are different names for the same phenomenon and occur in variables of different types, but they share the same characteristic: the association between two variables can be reversed, diminished, or enhanced when another variable is controlled for statistically.

Furthermore, it is very difficult to compare results across studies where many varied attempts are made to control for different confounders, especially in the absence of any consistent reason being

given for the choice of confounders. In some situations, statistical adjustment may introduce bias rather than eliminate it. One possible solution to examining the introduced bias of *the reverse paradox* problem might be to compute two statistical analyses with or without the baseline data of the outcomes. Another solution to dealing with this problem might be to present the baseline data of the outcomes in order to visualise whether there is a difference in outcomes from the start, while a third possible solution might be to calculate the improvement (follow-up score minus baseline score = improvement) and compare these improvement values between the groups. In Paper II, the adjustment has not been made for the baseline data of outcomes (EQ-5D index, EQ VAS, Pain VAS) to avoid a Lord´s paradox.

In almost all retrospective observational studies, some confounding factors occur and are dealt with in the study design or are accounted for at the analysis stage. So, retrospective observational studies must, or at least should, have a section on the selection of used confounding factors stated in the manuscript. Some confounding factors can be more hazardous to use in studies with long-term follow-up periods, because they can change over time (for example: smoking habits, BMI and so on).

There is also an ongoing debate about whether or not there is any unobserved (and therefore unadjusted) confounder in these retrospective observational studies. However, it is important to ask whether all the important confounders are generally known in many areas and might therefore be unnecessary to mention as a limitation in a study? The unknown confounder line in an observational study manuscript should not be a substitute for the lack of information relating to an important confounder. Despite these limitations in both the quantification of surgeons' experience and differences in adjustment, the results in these studies of patient reported outcomes measures are similar and this can act as evidence to demonstrate that patient-reported outcome measurements following primary total hip arthroplasties are not or are only very weakly related to the surgeon's level of experience.

Qualitative method, informant sampling and interview difficulties

In Paper III, the purposive (stratified) sampling method was used, where assumptions relating to a possible variation in perceptions of being provided with surgeon-specific results from a national quality register might be dependent on surgeon-specific attributes; years in practice (divided into the categories used in Paper II), hospital level (based on the Swedish Hip Arthroplasty Register definition) and gender. This stratified sampling may not have been the optimal sampling method for this study because the number of possible informants in each category differs and the means of making contact with possible informants to inform them about the study might not be the best at university- and regional hospitals or there could be a low volume of primary total hip arthroplasties at university-and regional hospitals.

The number of possible hospitals differs between the categories of experience used in the study; there are far more rural hospitals (31) and country hospitals (23) compared with university and regional hospitals (nine) listed in the Swedish Hip Arthroplasty Register's Annual Reports (Kärrholm et al. 2018). This difference in the number of hospitals at each hospital level might have an impact on the possible number of informants, thereby making it easier to recruit informants. Using a different sample method, such as the convenience sampling method, the same problem of few hospitals and probably also few surgeons at the university and regional hospitals is likely to be encountered.

Another issue is the means of making contact with possible informants in larger administrative hospital organisations where the information might not be easy and smooth and therefore slips through the organisation. In this study, we addressed the study information to the head of the department and asked him/her to pass on the information in the organisation to the surgeons performing total hip arthroplasties. This "head of department method" might be more suitable in more compact hospital organisations. On-site visits might resolve this issue, but obtaining access to potential meetings to inform the surgeons may create the same problem, because the request for a visit has to be addressed to someone.

A fairly low annual volume of primary total hip arthroplasties at university-and regional hospitals (Kärrholm et al. 2018) might also explain the difficulty involved in recruiting informants at these hospitals. There might also be another case-mix of total hip arthroplasty at these hospitals that are more difficult intraoperatively and there are therefore a larger number of follow-up visits to the surgeon. So, the surgeons at university and regional hospitals already receive feedback on performance and are satisfied with this, in contrast to a high-volume hospital producing a large bulk of total hip arthroplasty surgeries where the priorities are first-time visits to decide whether or not to perform surgery and the follow-up visits to the surgeon are second on the list of priorities.

There are both advantages and disadvantages to using phenomenography as a research method. However, phenomenography focuses on people and how they experience phenomena in the world around them and not on the phenomena themselves – for example, the conception that total hip arthroplasty surgery might include "a surgery that will lead to pain relief", "something fearful", "something necessary for restoring mobility" or "a technically demanding surgery". All these are recognisable perceptions of total hip arthroplasty surgery which together provide a much richer and more colourful description of what total hip arthroplasty surgery is.

Sjöström & Dahlgren (2002) point out that there are some concerns associated with the phenomenographic interview. One of the first concerns is the informant's motivation to participate in the study. Is there a risk of bias in the informant sampling process used in Paper IV, are surgeons with a positive attitude to individual surgeon feedback more willing to participate than surgeons with a negative attitude? This latter question is almost impossible to answer easily. Surgeons who are willing to participate may have deeper thoughts about individual surgeon feedback and see an opportunity to communicate these thoughts and contribute to development, but this theory remains as a speculation.

Another consideration is the reliability of the phenomenographic result and whether other researchers would identify the same description categories as the original researcher. Moreover, would other

researchers recognise the perceptions identified by the original researcher through the description categories? If during the interviews the informants focus on a particular dimension influenced by their experience and, despite the fact that a structured questionnaire is being used, each interview might cover different aspects of the phenomenon, and this could pose a problem. This could be because the informant's comments can lead the researcher to ask for additional information in a certain way, or not to ask at all. The researcher's background and focus are likely to have an impact as well, even if the researcher endeavours to prevent this. These factors cannot be replicated. However, this problem is not purely associated with the phenomenographic research approach, it also applies to a great deal of qualitative research. The intersubjective agreement check suggested by Marton F (1988) might be one solution to increasing the reliability of the phenomenographic results from interviews that have already been conducted, but it is unable to handle the interpersonal relationship that might direct some of the interviews in one or another direction.

Future surgeon feedback in Sweden

All variations between surgeons in individual surgeons' programmes should be interpreted with some caution. We can and must expect some degree of variation between surgeons, depending on both annual volume (i.e. low-volume surgeons produce wider CI in the funnel plots) and factors unrelated to the surgeon's performance or scope. We must also take account of the case-mix profile for each surgeon which might influence the outcome of his/her performance. For example, a higher proportion of errors could be expected from surgeons performing more technically difficult procedures or operating on patients with more comorbidities.

To acquire credibility for an individual surgeon feedback programme in both the Swedish orthopaedic community and among individual surgeons, we must compare apples with apples and not with other fruits. In other words, we must make a fair adjustment to prevent falsely identifying a senior orthopaedic surgeon as a poor performer and the outcome is dependent on a high-risk patient case mix.

The Swedish Hip Arthroplasty Register has a long history of recording not only surgery-specific factors (type of implants, incision and so on) but also patient characteristics (age, gender, BMI, ASA classification, diagnostic indication for implantation and so on). These are open reported variables and are reported back to the hospitals at hospital level in the annual report. This hospital feedback struggles with the same case-mix problem as a future individual surgeon feedback programme; are we comparing apples with apples and not with other fruits? To deal with this possible problem with a fruit salad comparison, a standard patient (primary osteoarthritis, all genders, age 55 to 84 years, ASA classification one-two & BMI less than 30) (Kärrholm et al. 2019) has been created by the Swedish Hip Arthroplasty Register. This Swedish Hip Arthroplasty Register standard patient is a well-known concept among total hip arthroplasty surgeons in Sweden and has been suggested by a number of informants during interviews as a feasible case-mix adjustment.

In order to guarantee an open society with access to information about the work of the Riksdag (Swedish parliament), government and government agencies, the principle of public access to official documents has been incorporated into one of the fundamental laws, the Freedom of the Press Act.

> *"To encourage the free exchange of opinion and availability of comprehensive information, every Swedish citizen shall be entitled to have free access to official documents."*
>
> *(Chapter 2, Article 1, Freedom of the Press Act)*

This principle gives the general public the right to read official documents submitted to or drawn up by the authorities. All documents received or dispatched – letters, decisions and reports – are in principle public documents and must be made available to anyone to read. This openness also applies to some of the quality registers in Sweden and

(might) inhibit the development of a surgeon-specific feedback programme from quality registers. So, there is a risk that individual surgeon results are lost and do more harm rather than being useful to surgeons. Despite this obstacle, there are national quality registers in Sweden, such as the National Prostate Cancer Registry, which collect surgeon-specific data. During the 1990s, the number of regional registers of various aspects of prostate cancer care increased successively and they have been merged to create the National Prostate Cancer Registry. At Sahlgrenska University Hospital in Gothenburg, radical prostatectomy outcome data have been collected prospectively since 1988. Patient reported outcomes were added in January 2001. In 2015, a specific radical prostatectomy register was introduced. The surgeons fill in an on-line form immediately after the procedure. The surgeons' identity is coded and the code key is kept at the departments. All units have on-line access to all their own data and are able to compare the results for the individual surgeons at the clinic, as well as being able to compare their results with other units' aggregated results and the national average. The process of quality control locally at Sahlgrenska University Hospital, and nationally through the National Prostate Cancer Registry, has been long and fruitful (Stranne et al. 2019). At Sahlgrenska University Hospital, the open and direct feedback of all prostate surgeons' results on positive surgical margins and functional patient reported outcomes data have formed the basis for a discussion within the group of surgeons on how to improve the results and minimise the heterogeneity within the group with the aim of reducing the previously reported heterogeneous results for the individual surgeons (Carlsson et al. 2014). So, if individual surgeon feedback can influence the outcomes following urological procedures in Sweden, a programme for individual surgeon feedback following total hip arthroplasties might help to minimise the variation between surgeons and improve the results following these surgeries as well.

Conclusions

Paper I

High annual surgical activity is associated with a reduced risk of adverse events within 90 days. Based on these findings, healthcare providers should consider planning for increased surgeon volume.

Paper II

Patients can expect similar health improvements, pain reductions and satisfaction one year after a primary total hip arthroplasty operation, irrespective of years in practice after speciality certification as an orthopaedic surgeon.

Paper III

Orthopaedic surgeons and trainees in Sweden have multifarious perceptions on the meaning of individual surgeon feedback from a national arthroplasty quality register. The identification of these existing perceptions might contribute to a better awareness and understanding in the development of a future programme for individual surgeon feedback from at national quality register following total hip arthroplasties in Sweden. A future individual surgeon feedback programme from a national quality register needs to be developed in collaboration with the orthopaedic surgeons.

Paper IV

In a Swedish setting, the variation in surgeon performance, as measured by adverse events within 90 days and re-operations within two years following total hip arthroplasty, was small and there were few

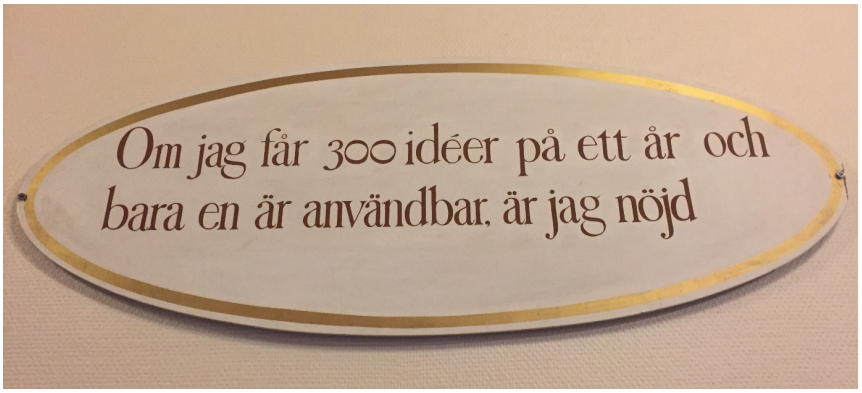
outliers (above the upper 95% CI). The probability of individual surgeons being regarded as outliers when creating surgeon-specific feedback in the Swedish Hip Arthroplasty Register is very low.

Overall conclusion

The findings in this thesis show that the outcomes in the form of short-term adverse events are dependent on the surgeons' operation volume and small number of surgeons will be outliers compared with their peers in a Swedish setting. However, patients can expect similar health improvements and pain relief at one year following surgery, irrespective of the surgeons' experience as orthopaedic specialists. There are a limited number of perceptions among Swedish orthopaedic specialists and trainees of the phenomenon of individual surgeon feedback. These perceptions vary between an opportunity to improve care for the patient through professional development to jeopardising patient safety by focusing more on the numbers than on the patients.

Future projects

There are so many interesting research fields just waiting to be explored in the future and to quote one of several important inventors in Sweden:



And if I get 300 ideas during a year and one is useful, I am happy

Alfred Nobel

One of the future projects might be involve investigating the influence and effect of the orthopaedic surgeon in the Swedish field of arthroplasty, which has only been touched upon. Surgeon volumes have been presented as one confounder that needs to be taken into account when conducting register-based studies. An extended version of "*High annual surgeon volume reduces the risk of adverse events following primary total hip arthroplasty. A register-based study of 12,100 cases in western Sweden*", including primary total hip arthroplasties in Sweden also adjusted for smoking, socio-economics and the annual volume of other knee and hip arthroplasties, might explore the influence of annual

surgeons' volumes on the outcomes. By adding the surgeon to the national quality register using a unique code, this could be done easily.

However, it is also important to investigate whether continuity between patients and surgeons is important to improve the outcomes following hip arthroplasties. This has been empirically suggested as being important for improved outcomes following hip arthroplasties. However, there is a lack of knowledge in this field in earlier publications. So, it might be worth investigating the effect of continuity from both the patient-surgeon angle and the surgeon-patient angle in the future.

Another interesting area to explore is whether the orthopaedic surgeon can be replaced by other hospital staff (physios, nurses and so on) in the decision on whether or not to perform surgery in standard patients with the same quality. A side-effect of the opportunity to replace the orthopaedic surgeon with other hospital staff is the time that is released for the orthopaedic surgeon to increase the time in the operating room. This project is already up and running at Sahlgrenska University Hospital, so we shall perhaps have the first results from this project in a couple of years.

Supplements

A published appendix of the ICD 10 and NOMESCO codes used for the definition of adverse events in Paper I and IV. The appendix is published online by Acta Orthopaedica 2018; 89 (DOI: 10.1080/17453674.2018.1554418).

Project collaborators

Supervisors

***Maziar Mohaddes*, MD, PhD** – Associate Professor, orthopaedic surgeon at the Department of Orthopaedics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg and affiliated with the Swedish Hip Arthroplasty Register

***Hans Lindahl*, MD, PhD –** orthopaedic surgeon at the Department of Orthopaedics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, and formerly affiliated with the Swedish Hip Arthroplasty Register

***Johan Kärrholm*, MD, PhD** – Professor Emeritus, orthopaedic surgeon at the Department of Orthopaedics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, and Director of the Swedish Hip Arthroplasty Register

***Göran Garellick*, MD, PhD** – Professor Emeritus at the Department of Orthopaedics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, and former Director of the Swedish Hip Arthroplasty Register

Collaborators and co-authors

***Ola Rolfson*, MD, PhD –** Professor, orthopaedic surgeon and head of the Department of Orthopaedics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, and Director of the Swedish Hip Arthroplasty Register

***Szilárd Nemes*, PhD –** Associate Professor, former biostatistician at the Swedish Hip Arthroplasty Register

***Henrik Malchau*, MD, PhD** – Professor, Harvard Medical School and co-director emeritus of the Harris Orthopaedic Laboratory at Massachusetts General Hospital, Boston, MA, USA, affiliated with

the Swedish Hip Arthroplasty Register and orthopaedic surgeon at the Department of Orthopaedics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg

***Peter Cnudde*, MD, PhD** – orthopaedic surgeon at the Department of Orthopaedics, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Department of Orthopaedics, Hywel Dda University Healthboard, Prince Philip Hospital,B ryngwynmawr, United Kingdom, and affiliated with the Swedish Hip Arthroplasty Register

***Kicki Klaeson*, RN, PhD** – nurse, Department of Oncology, Skaraborg Hospital, Skövde, Sweden

Daniel Odin – former biostatistician at the Swedish Hip Arthroplasty Register

Emma Nauclér – biostatistician at the Swedish Hip Arthroplasty Register

Eric Bülow – biostatistician at the Swedish Hip Arthroplasty Register

Acknowledgements

My thesis could not have been completed without great support and encouragement from other people around me. I will always be grateful to those people and I wish to address special thanks to:

Maziar Mohaddes, my main supervisor during the latter part of producing this thesis, for your remarkable support and for all the phone calls, mostly after the kids were put to bed. Without your efforts, this thesis would never have been completed. You know the thing about giving feedback without pushing it over the limit!

Hans Lindahl, my former main supervisor (and now one of my co-supervisors) at the start of this journey and the one who brought me into research. You are one of the kindest people I know and you will always have a very special place in my heart. I will always remember our discussion not only about research but also about music, sinkholes and life in general.

Johan Kärrholm, for all the good advice and support during this journey in science. In my eyes, you are one of the greatest researchers and, if I only had ten per cent of your knowledge, I would be grateful.

Göran Garellick, as the director of the Swedish Hip Arthroplasty Register at the time of the start of this thesis, you made a tremendous effort on the ethical approvals. I will always remember our battles with the "ladies" and the regional ethics board.

Ola Rolfson, for introducing me to the academic world and for the support in writing manuscripts.

Szilárd Nemes, only a phone call or an e-mail away to explain the simplest statistics.

Daniel Odin, "Mr Rapid" in statistics. You are the fastest man in R, a phone call begging for some new or a change to some old calculations, you solve it in no time.

Emma Nauclér, thanks for your kindness in always trying to explain my questions in statistics.

Lotta Kettil, my fiancée and the love of my life. Without your never-ending love and support, this thesis would never been completed.

Nora and Alma, my two beautiful daughters. You make my life complete.

My colleagues and friends at both the Department of Orthopaedics and the Research and Development Unit at Skaraborg Hospital. A special thanks to **Catrin Olofsson**, at the Department of Communications at Skaraborg Hospital for helping out with the layout of the thesis.

References

Australian Orthopaedic Association National Joint Replacement Registry, 2017 ANNUAL REPORT, 2017.

Bhandari M, Lochner H, Tornetta P 3rd. Effect of continuous versus dichotomous outcome variables on study power when sample sizes of orthopaedic randomized trials are small. Arch Orthop Trauma Surg. 2002;122(2):96-8.

Battaglia T C, Mulhall K J, Brown T E, Saleh K J. Increased surgical volume is associated with lower THA dislocation rates. Clin Orthop Relat Res. 2006;(447):28-33.

Bozic K J, Rubash H E. The painful total hip replacement. Clin Orthop Relat Res. 2004;(420):18-25.

Braga L H, Farrokhyar F, Bhandari M. Confounding: what is it and how do we deal with it? Canadian Journal of Surgery. 2012;55(2):132-8.

Brand R A. Ernest Amory Codman, MD, 1869-1940. Clin Orthop Relat Res. (2009) 467:2763–2765.

Brand R A. Biographical Sketch Ernest Amory Codman, MD (1869–1940). Clin Orthop Relat Res. (2013) 471:1775–1777.

Brown D L, Cai T T, DasGupta A. Interval estimation for a binomial proportion. Statistical Science. 2001;16(2):101-17.

Burns P B, Rohrich R J, Chung K C. The Levels of Evidence and their role in Evidence-Based Medicine. Plast Reconstr Surg. 2011;28(1):305-10.

Camberlin C, Vrijens F, De Gauquier K, Devriese S, Van De Sande S. Provider volume and short-term complications after elective total hip replacement: an analysis of Belgian administrative data. Acta Orthop Belg. 2011;77(3):311-19.

Chowdhury M M, Dagash H, Pierro A. A systematic review of the impact of volume of surgery and specialization on patient outcome. Br J Surg. 2007; 94(2):145-61.

Carlsson S, Berglund A, Sjoberg D, Khatami A, Stranne J, Bergdahl S, Lodding P, Aus G, Vickers A, Hugosson J. Effects of surgeon variability on oncologic and functional out-comes in a population-based setting. BMC Urol. 2014;14:25.

Codman E A. The Registry of Bone Sarcomas as an Example of the End-Result Idea in Hospital Organization. Clin Orthop Relat Res. (2009) 467:2766–2770. (Reprint)

Culliford D, Maskell J, Judge A, Cooper C, Prieto-Alhambra D, Arden N K. Future projections of total hip and knee arthroplasty in the UK: results from the UK Clinical Practice Research Datalink. Osteoarthritis Cartilage. 2015;23(4):594-600.

Dahlgren L O, Fallsberg M. Phenomenography as a qualitative approach in social pharmacy research. Journal of Social and Administrative Pharmacy. 1991;8(4):150-56.

Davies H T, Crombie I K, Tavakoli M. When can odds ratios mislead? BMJ. 1998 Mar 28;316(7136):989-91.

de Steiger R N, Graves S E. Orthopaedic registries: the Australian experience. EFORT Open Rev. 2019;4 DOI: 10.1302/2058-5241.4.180071.

Henricsson M (editor), Vetenskaplig teori och metod, från idé till examination inom omvårdnad. Studentlitteratur, Lund, 2018. (In Swedish)

Inacio M C S, Graves S E, Pratt N L, Roughead E E, Nemes S. Increase in Total Joint Arthroplasty Projected from 2014 to 2046 in Australia: A Conservative Local Model With International Implications. Clin Orthop Relat Res. 2017;475(8):2130-37.

Inglis T, Dalzell K, Hooper G, Rothwell A, Frampton C. Does orthopedic training compromise the outcome in total hip arthroplasty? Journal of Surgical Education. 2013;70(1): 76-80.

The International Society of Arthroplasty Registries. https://www.isarhome.org/bylaws, Revised June 10, 2018.

Jackson J. Father of the modern hip replacement: Professor Sir John Charnley (1911–82). Journal of Medical Biography. 2011;19:151–156. DOI: 10.1258/jmb.

Katz J N, Losina E, Barrett J, Phillips C B, Mahomed N N, Lew R A, Guadagnoli E, Harris W H, Poss R, Baron J A. Association between hospital and surgeon procedure volume and outcomes of total hip replacement in the United States medicare population. JBJS The Journal of Bone & Joint Surgery. 2001;83-A(11):1622-29.

Kishimoto Y, Suda H, Kishi T, Takahashi T. A low-volume surgeon is an independent risk factor for leg length discrepancy after primary total hip arthroplasty: a case-control study. Int Orthop. E-pub Date 2019; 31 Oct. DOI: 10.1007/s00264-019-04435-6.

Koltsov J C B, Marx R G, Bachner E, McLawhorn A S, Lyman, S. Risk-Based Hospital and Surgeon-Volume Categories for Total Hip Arthroplasty. J Bone Joint Surg Am. 2018;100(14):1203-08.

Kreder H J, Deyo R A, Koepsell T, Swiontkowski M F, Kreuter W. Relationship between the volume of total hip replacements performed by providers and the rates of postoperative complications in the state of Washington. JBJS The Journal of Bone & Joint Surgery. 1997;79-A(4):485-94.

Kreder H J, Williams J I, Jaglal S, Hu R, Axcell T, Stephen, D. Are complication rates for elective primary total hip arthroplasty in Ontario related to surgeon and hospital volumes? A preliminary investigation. Can J Surg. 1998;40(6):431-7.

Kurtz S M, Lau E C, Ong K L, Adler E M, Kolisek F R, Manley M T. Hospital, Patient, and Clinical Factors Influence 30- and 90-Day Readmission After Primary Total Hip Arthroplasty. J Arthroplasty. 2016;31(10):2130-8.

Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg Am. 2007; 89(4):780-5.

Kärrholm J, Lindahl H, Malchau H, Mohaddes M, Rogmark C, Rolfson O. The Swedish Hip Arthroplasty Register Annual Report 2015. 2016.

Kärrholm J, Lindahl H, Malchau H, Mohaddes M, Rogmark C, Rolfson O. The Swedish Hip Arthroplasty Register Annual Report 2016. 2017.

Kärrholm J, Mohaddes M, Odin Daniel, Vinblad J, Rogmark C, Rolfson O. The Swedish Hip Arthroplasty Register Annual Report 2017. 2018.

Kärrholm J, Rogmark C, Nauclér E, Mohaddes M, Rolfson O. Svenska Höftprotesregistrets Årsrapport 2018. 2019 (In Swedish)

Larsson S. Kvalitativ Analys Exemplet Fenomenografi. Studentlitteratur, Lund, 1986. (In Swedish)

Lavernia C J, Guzman J F. Relationship of surgical volume to short-term mortality, morbidity, and hospital charges in arthroplasty. J Arthroplasty. 1995;10(2):133-40.

Learmonth I D, Young C, Rorabeck C. The operation of the century: total hip replacement. Lancet. 2007;370(9597):1508-19.

Le Cossec C, Colas S, Zureik M. Relative impact of hospital and surgeon procedure volumes on primary total hip arthroplasty revision: a nationwide cohort study in France. Arthroplasty Today. 2017;3(3):176-82.

Lee K J, Wiest M M, Carlin J B. Statistics for clinicians: an introduction to linear regression. J Paediatr Child Health. 2014;50(12):940-3.

Lochner H, Bhandari M, Tornetta P 3rd. Type-II error rates (beta errors) of randomized trials in orthopaedic trauma. J Bone Joint Surg Am. 2002;83(11):1650-5.

Losina E, Barrett J, Mahomed N N, Baron J A, Katz J N. Early failures of total hip replacement: effect of surgeon volume. Arthritis Rheum. 2004;50(4):1338-43.

Ludvigsson J F, Andersson E, Ekbom A, Feychting M, Kim J L, Reuterwall C, Heurgren M, Olausson P O. External review and validation of the Swedish national inpatient register. BMC Public Health. 2011;11:450.

Ludvigsson J F, Otterblad-Olausson P, Pettersson B U, Ekbom A. The Swedish personal identity number: possibilities and pitfalls in

healthcare and medical research. Eur J Epidemiol. 2009;24(11):659-67.

Macpherson G J, Brenkel I J, Smith R, Howie C R. Outlier analysis in orthopaedics: use of CUSUM: the Scottish Arthroplasty Project: shouldering the burden of improvement. J Bone Joint Surg Am. 2011;93 Suppl 3(E):81-8.

Marton F. Phenomenography: a research approach to investigating different understanding of reality. In Qualitative Research in Education: Focus and Methods (Sherman R.R. & Webb R.B. eds), Routledgefalmer, London, 1988.

Marton F, Booth S. The Antonomy of Awareness. F Marton & Booth S (red). Learning and Awareness, p 82-109. New Jersey: Lawrence Erlbaum Ass, 1997.

Moran M, Yap S L, Walmsley P, Brenkel I J. Clinical and radiologic outcome of total hip arthroplasty performed by trainee compared with consultant orthopedic surgeons. J Arthroplasty. 2004;19(7):853-7.

Murphy W S, Cheng T, Lin B, Terry D, Murphy S B. Higher Volume Surgeons Have Lower Medicare Payments, Readmissions, and Mortality After THA. Clin Orthop Relat Res. 2019:477(2):334-341.

Mäkelä K T, Furnes O, Hallan G, Fenstad A M, Rolfson O, Kärrholm J, Rogmark C, Pedersen A B, Robertsson O, W-Dahl A, Eskelinen A, Schrøder H M, Äärimaa V, Rasmussen J V, Salomonsson B, Hole R, Overgaard S. The benefits of collaboration: the Nordic Arthroplasty Register Association. EFORT Open Rev. 2019 Jun 3;4(6):391-400.

Nemes S, Gordon M, Rogmark C, Rolfson O. Projections of total hip replacement in Sweden from 2013 to 2030. Acta Orthop. 2014;85(3);238-43.

NJR Orthopaedic Consultant Outcomes Publication 2015: http://www.njrcentre.org.uk/njrcentre/Portals/0/Documents/England/COP/NJR%20COP%20Guidance%20for%20consultant%20orthopaedic%20surgeons%20July%202015%20FINAL.pdf?ver=2015-07-22-112032-157.

Noyez L. Control charts, Cusum techniques and funnel plots. A review of methods for monitoring performance in healthcare. Interactive CardioVascular and Thoracic Surgery. 2009;9(3):494–99.

Palan J, Gulati A, Andrew J G, Murray D W, Beard D J. The trainer, the trainee and the surgeons' assistant: clinical outcomes following total hip replacement. J Bone Joint Surg Br. 2009;91(7):928-34.

Paterson J M, Williams J I, Kreder H J, Mahomed N N, Gunraj N, Wang X, Laupacis A. Provider volumes and early outcomes of primary total joint replacement in Ontario. Can J Surg. 2010;53(3):175-83.

Petrie A, Sabin C. Bias and confounding. Medical statistics at a glance. 3rd ed. Oxford: Wiley-Blackwell, 2009.

Perry D C, Griffin X L, Dritsaki M, Costa M L, Parsons N. Becoming confident about confidence intervals. The Bone & Joint Journal. 2017;99-B(5):563-65.

Pilz V, Hanstein T, Skripitz R. Projections of primary hip arthroplasty in Germany until 2040. Acta Orthop. 2018;89(3):308-13.

Radford P D, Derbyshire L F, Shalhoub J, Fitzgerald J E F; Council of the Association of Surgeons in Training. Publication of surgeon specific outcome data: a review of implementation, controversies and the potential impact on surgical training. Int J Surg. 2015;13:211–216..

Ravi B, Croxford R, Austin P. C, Hollands S, Paterson J M, Bogoch E, Kreder H, Hawker G A. Increased surgeon experience with rheumatoid arthritis reduces the risk of complications following total joint arthroplasty. Arthritis Rheumatol. 2014;66(3):488-96.

Ravi B, Jenkinson R, Austin P C, Croxford R, Wasserstein D, Escott B, Paterson J M, Kreder H, Hawker G A. Relation between surgeon volume and risk of complications after total hip arthroplasty: propensity score matched cohort study. BMJ. 2014;348.

Rechel B, McKee M, Haas M, Marchildon G P, Bousquet F, Blümel M, Geissler A, van Ginneken E, Ashton T, Sperre Saunes I, Anell A, Quentin W, Saltman R, Culler S, Barnes A, Palm W, Nolte E. Public reporting on quality, waiting times and patient experience in 11 high-income countries. Health Policy. 2016;120(4):377–383.

Reidy M J, Faulkner A, Shitole B, Clift B. Do trainee surgeons have an adverse effect on the outcome after total hip arthroplasty?: a ten-year review. The Bone & Joint Journal. 2016; 98-B(3):301-6.

Rolfson O, Kärrholm J, Dahlberg L E, Garellick G. Patient-reported outcomes in the Swedish Hip Arthroplasty Register: results of a nationwide prospective observational study. J Bone Joint Surg Br. 2011 Jul;93(7):867-75.

Scottish Arthroplasty Project 2019. https://www.arthro.scot.nhs.uk/Reports/Visualisation.html

Sedgwick P. A comparison of parametric and non-parametric statistical tests. BMJ. 2015;35:2053.

Sedgwick P, Greenwood N. Understanding the Hawthorne effect. BMJ. 2015;351:h4672.

Singh P, Madanipour S, Fontalis A, Singh Bhamra J, Abdul-Jabar H B. A systematic review and meta-analysis of trainee- versus consultant surgeon-performed elective total hip arthroplasty. EFORT Open Rev. 2019;4:44-55.

Sjöström B, Dahlgren L O. Applying phenomenography in nursing research. Journal of Advanced Nursing. 2002;40(3):339–345.

Solomon D H, Losina E, Baron J A, Fossel A H, Guadagnoli E, Lingard E A, Miner A, Phillips C B, Katz J N. Contribution of hospital characteristics to the volume-outcome relationship: dislocation and infection following total hip replacement surgery. Arthritis Rheum. 2002;46(9):2436-44.

Song J W, Chung K C. Observational studies: cohort and case-control studies. Plast Reconstr Surg. 2010;126(6):2234-42.

Spiegelhalter D J. Funnel plots for comparing institutional performance. Statist Med. 2005;24:1185-1202.

Stenfors-Hayes T, Hult H, Dahlgren M A. A phenomenographic approach to research in medical education. Medical Education. 2013: 47: 261–270.

Stranne J, Axen E, Franck-Lissbrant L, Fransson P, Frånlund M, Hugosson J, Khatami A, Koss-Modig K, Lodding P, Nyberg M, Stattin P, Bratt O. Single institution followed by national implementation of

systematic surgical quality control and feedback for radical prostatectomy: a 20-year journey. World J Urol. 2019; published online 6 aug.

Suttorp M M, Siegerink B, Jager K J, Zoccali C, Dekker F W. Graphical presentation of confounding in directed acyclic graphs. Nephrol Dial Transplant. 2015; 30(9):1418-23.

Tu Y K, Gunnell D, Gilthorpe M S. Simpson's Paradox, Lord's Paradox, and Suppression Effects are the same phenomenon--the reversal paradox. Emerg Themes Epidemiol. 2008;5(5):2.

Vallance A E, Fearnhead N S, Kuryba A, Hill J, Maxwell-Armstrong C, Braun M, van der Meulen J, Walker K. Effect of public reporting of surgeons' outcomes on patient selection, "gaming," and mortality in colorectal cancer surgery in England: population based cohort study. BMJ. 2018;361:k1581.

Walker K, Neunberger J, Groene O, Cromwell D A, van den Meulen J. Public reporting of surgeon outcomes: low numbers of procedures lead to false complacency. Lancet. 2013;382:1674-77.

Wiest M, Lee K J, Carlin J B. Statistics for clinicians: An introduction to logistic regression. J. Paediatr. Child Health. 2015;7:670–3.

Wilson M D, Dowsey M M, Spelman T, Choong P F. Impact of surgical experience on outcomes in total joint arthroplasties. ANZ J Surg. 2016;86(12):967-72.

Weber M, Benditz A, Woerner M, Weber D, Grifka J, Renkawitz T. Trainee Surgeons Affect Operative Time but not Outcome in Minimally Invasive Total Hip Arthroplasty. Scientific Reports. 2017;7(1): 6152.

Yasunaga H, Tsuchiya K, Matsuyama Y, Ohe K. High-volume surgeons in regard to reductions in operating time, blood loss, and postoperative complications for total hip arthroplasty. J Orthop Sci. 2009;14(1):3-9.