Annals of Epidemiology 76 (2022) 13–19

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: sciencedirect.com/journal/annals-of-epidemiology

Original article

The potential value of observational studies of elective surgical interventions using routinely collected data

Yassine Ochen, MD [a,b], Maaike GJ Gademan, PhD [a,c], Rob GHH Nelissen, MD, PhD [c], Rudolf W. Poolman, MD, PhD [c], Luke pH Leenen, MD PhD [b], R Marijn Houwert, MD, PhD [b], Rolf HH Groenwold, MD, PhD [a,d,*]

[a] *Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands*
[b] *Department of Surgery, University Medical Center Utrecht, the Netherlands*
[c] *Department of Orthopaedics, Leiden University Medical Center, Leiden, the Netherlands*
[d] *Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands*

ARTICLE INFO

Article history:
Received 20 June 2022
Revised 3 October 2022
Accepted 6 October 2022
Available online 14 October 2022

Keywords:
Observational studies
Randomized controlled trials (RCTs)
Routinely collected data
Total hip arthroplasty (THA)
Comparability
Dutch Arthroplasty Register (LROI)

ABSTRACT

Purpose: To assess the apparent validity of observational studies of elective arthroplasty interventions.
Methods: Data from the nationwide Dutch Arthroplasty Register were used. The first case study compared surgical approaches for total hip arthroplasty (posterolateral approach vs. straight lateral approach), where allocation of the intervention was assumed to be mostly independent of patient characteristics. The second case study compared fixation methods (cemented vs. uncemented), where choice of fixation method was expected to depend on patient characteristics. The potential for confounding was quantified by differences between intervention groups and the impact of confounding adjustment.
Results: The study of posterolateral approach versus straight lateral approach included 73,750 and 16,557 patients, respectively, and showed no meaningful differences in patient characteristics between treatment groups (standardized mean differences <0.1) and also no relevant impact of confounding adjustment (Z-scores <1). The study of cemented versus uncemented total hip arthroplasty (THA) included 29,579 and 79,360 patients, respectively. Several meaningful imbalances were observed in patient characteristic between the two treatment groups (standardized mean differences >0.1), as well as a relevant impact of confounding adjustment (Z-scores >2).
Conclusions: This study provides insight in the reasoning behind the credibility of observational studies of surgical interventions using routinely collected data and when confounding is expected to have a major impact and thus additional precautions to limit confounding are needed.
© 2022 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

Introduction

Randomized controlled trials (RCTs) are generally considered to provide the highest level of evidence of treatment effects [1,2]. Randomization prevents confounding due to selective allocation of treatment to patients. Blinding, of patients and treating physicians, prevents selective changes in health care behavior or crossover, and efforts can be made to ensure that assessors of the outcome are blinded for the received treatment. Nevertheless, RCTs might not always be ethical, feasible, or necessary to address a specific research question. This is especially apparent in the surgical field, where variation in surgical practice can lead to practical challenges in terms of patient recruitment, randomization, and blinding [3,4]. Moreover, the patient populations encountered in daily clinical practice can differ from the often highly selected patient populations enrolled in RCTs [5]. Also, RCTs might not always have sufficient follow-up or sample size to assess rare outcomes or long-term treatment effects [6,7]. Consequently, often the results of RCTs are not implemented in surgical practice [8,9].

☆ None. The authors, their immediate families, and any research foundation with which they are affiliated have not received any financial payments or other benefits from any commercial entity related to the subject of this article. Individual Conflict of Interest forms are provided.

* Corresponding author: Department of Clinical Epidemiology, Leiden University Medical Center, Albinusdreef 2, 2333 ZA, Leiden, the Netherlands.
E-mail address: r.h.h.groenwold@lumc.nl (R.H. Groenwold).

https://doi.org/10.1016/j.annepidem.2022.10.004
1047-2797/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

Observational, non-randomized, studies could provide an alternative, or at least complementary, source of evidence, provided the observational studies are of sufficient quality [10]. A key aspect to consider is whether the treatment groups that are being compared are inherently different (confounding), or whether there might be situations in which comparability can be achieved. Particularly studies of acute operative treatments might be less sensitive to confounding when the treatment option depends on a surgeon's preference but not on individual patient characteristics [3]. In such cases, one can speculate that groups of patients who underwent different surgical treatments might be rather similar (except for the investigated treatments) [3]. This has been observed in different meta-analyses of various surgical treatments in orthopedic trauma surgery, in which the treatment arms appeared comparable in terms of patient characteristics [3,11–13]. However, whether this also holds for elective surgery has not been investigated.

The aim of this study was to assess the potential for confounding in observational studies of elective surgical interventions using routinely collected data. The first example study aimed to compare surgical approaches for total hip arthroplasty (specifically posterolateral approach (PLA) versus straight lateral approach (SLA)), where allocation of the intervention was assumed to be independent of patient characteristics. In secondary analyses, these approaches were also compared against the anterior approach (AA). The second example study compared fixation methods (cemented vs. uncemented), where choice of fixation method was expected to depend on patient characteristics notably age. For both case studies, we investigated comparability between intervention groups and the impact of confounding adjustment.

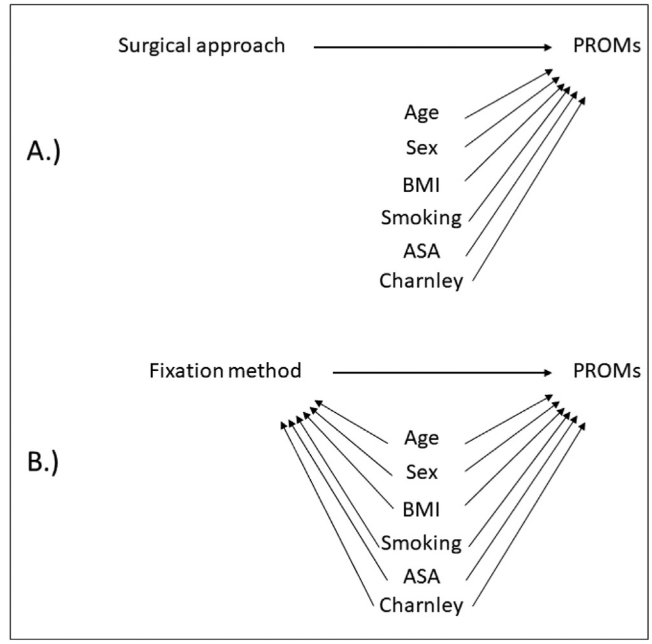

Fig. 1. Graphical representation of potential for confounding in observational studies of total hip arthroplasty. Panel A is causal diagram of possible relations between variables in an observational study of the effect of surgical approach (posterolateral approach vs. straight lateral approach) on patient-reported outcome measures (PROMs). Factors that influence PROMs, including age, sex, body mass index (BMI), smoking status (smoking), American Society of Anaesthesiologists (ASA) classification, and Charnley classification (Charnley), are not expected to influence the choice of surgical approach. Consequently, there are no arrows from these factors to approach. Panel B is causal diagram of possible relations between variables in an observational study of the effect of fixation method (cemented vs. uncemented) on PROMs. Factors that influence PROMs, including age, sex, BMI, smoking, ASA, and Charnley. These factors are expected to also influence the choice of fixation method. Therefore, arrows from these factors to cementation are included.

Methods

Example 1: Surgical approach of THA

THA is considered to be one of the most successful orthopedic procedures for patients with osteoarthritis, resulting in relief of pain, improved hip function, and substantial improvement in quality of life. However, there is no consensus regarding the optimal surgical approach [14–18]. Currently, the PLA and the SLA are the most frequently used techniques worldwide. Another approach, which has become more popular in recent years, is the AA [19]. The difference in outcomes seems small and each of the approaches have their own set of complications, benefits, and learning curves [14–18]. Therefore, the decision for the surgical approach is predominantly determined by surgeon preference and experience, as well as local hospital standards [20]. We hypothesize that groups of patients who are operated using either of the three approaches are similar in terms of prognostic relevant characteristics, that is, that surgical approach is independent of patient characteristics (Fig. 1*A*).

We compared the three groups of patients who were treated with primary total hip arthroplasty (THA) using the PLA, SLA, and AA. The primary comparison was made between the two traditional approaches, PLA versus SLA. Secondary comparisons were made between the more recent AA approach and each of the two traditional approaches; PLA versus AA, and SLA versus AA. Inclusion criteria were: (1) age 18 years or older, and (2) primary diagnosis osteoarthritis. Exclusion criteria were: (1) revision arthroplasties and (2) metal on metal arthroplasties. The three groups were compared in terms of preoperative patient characteristic, surgical variables, and patient-reported outcome measures (PROMs).

Example 2. fixation of THA

The success of THA and the worldwide acceptance is largely due to the development of the durable cemented low-friction arthroplasty with high survival rates. Although the initial components were cemented, the use of uncemented components has gained popularity over the years [21–24]. Both the cemented and uncemented techniques result in satisfactory fixation, but may differ regarding complications [21–24]. The cemented and uncemented fixation methods are used for heterogeneous groups, with different factors that can affect revision and survival rates such as geometry, material shape, surface finish, and bearings [21,22]. In the last decade, THA has changed from mainly cementation to mainly uncemented fixation and this trend is still continuing, particularly in younger patients [25]. We hypothesized that the choice for the cemented or uncemented method for THA is – to a large extent – based on patient characteristics and therefor groups of patients who are operated using either the cemented or uncemented fixation method differ in terms of their characteristics (Fig. 1*B*).

We compared the two groups of patients who were treated with primary THA using the cemented versus the uncemented fixation method. Inclusion criteria were: (1) age 18 years or older, and (2) primary diagnosis osteoarthritis. Exclusion criteria were: (1) revision arthroplasties, (2) metal on metal arthroplasties, or (3) arthroplasties with a hybrid fixation. The two groups were compared in terms of preoperative patient characteristic, surgical variables, and PROMs.

Data source

Routinely collected data from the nationwide Dutch Arthroplasty Register (LROI) were extracted for this study [26]. Prospective data collection started in 2007. The collection of PROMs of patients who underwent THA started in 2014. In 2016, data on primary THAs were provided by up to 99 hospitals and clinics (100% coverage of Dutch hospitals). The completeness of the data is checked against the hospital information systems and currently exceeds 99% for patient and surgery characteristics for primary THAs. Data on PROMs are provided by up to 80 centers [26–28]. For the current study, we used only available information regarding the patient characteristics, surgical procedures, and PROMS described below (i.e., complete records).

Data collection

Data were obtained from all adult patients who were treated with primary THA between 2014 and 2018. Information about the following preoperative patient characteristics was collected from the LROI database; age, sex, body mass index (BMI), smoking status (yes/no), American Society of Anesthesiologists (ASA) classification (I, II, III–IV), Charnley classification (A, B1, B2, C), and previous surgical procedures on the involved hip (yes/no). In addition, information was collected about surgical approach (PLA, SLA, AA) and fixation method (cemented, hybrid, uncemented). PROMs were collected preoperative, postoperative at 3 months, and 12 months, and consisted of the three-level version EuroQol-5 Dimensions (EQ-5D), Numeric Rating Scale (NRS) for pain (during activity and at rest), Hip disability and Osteoarthritis Outcome Score, and Oxford Hip Score [29–32]. The three-level EQ-5D was converted to a continuous utility measure using Dutch population tariff [33]. The EQ-5D index score measured 12 months after surgery was considered the outcome of primary interest.

Statistical analysis

In both studies, the following analyses were performed. First, a comparison was made between the intervention groups regarding preoperative patient characteristics, surgical variables, and preoperative PROMs. Differences between groups were quantified per variable by means of the standardized mean difference (SMD), where a SMD of >0.1 was considered a meaningful imbalance between intervention groups [34]. The relation between the interventions and post-treatment (3 months and 12 months) PROMs were assessed using linear regression analysis, with and without adjustment for baseline information. Results are presented as mean difference with 95% confidence interval (CI). Adjustment was done for preoperative patient characteristic (age, sex, ASA, previous surgery, BMI, Charnley, smoking status), surgical variables (fixation technique in study one and surgical approach in study two), and preoperative PROMs (the same PROM measure as the outcome measure). These variables were selected as potential confounders, because of their possible relations with patient reported outcome measures. The magnitude and direction of the difference between the crude and adjusted mean difference was quantified by means of a Z-score, which in this case provides a standardized measure of the change in effect estimate when adjustment for potential confounders is made. Z-score values >2 indicate a relevant change [35]. The comparisons have a descriptive nature, focusing on comparability of treatment groups.

All analyses were performed in R version 3.6.1 (R Development Core Team, Released 2013, Vienna, Austria: R Foundation for Statistical Computing) [36].

Results

Example 1: Surgical approach of THA

Patient characteristics

In total, 120,902 patients met the inclusion criteria for example 1. The baseline characteristics are shown in Table 1. The PLA group included 73,750 patients (61%), the SLA group 16,557 patients (14%), and the AA group 30,595 patients (25%). There were no meaningful differences in preoperative patient characteristics between the PLA and the SLA groups (all SMD <0.1). However, the PLA and AA groups differed regarding various preoperative patient characteristics, for example age (SMD 0.109), ASA classification (SMD 0.172), and BMI (SMD 0.178). Also, the SLA and AA groups differed regarding various preoperative patient characteristics, for example age (SMD 0.141), ASA classification (SMD 0.131), and BMI (SMD 0.188).

Outcomes

The results of the crude and adjusted regression analyses are shown in Table 2. The mean EQ-5D index score at 12 months was 0.859 (SD 0.188) in the PLA group, compared to 0.826 (SD 0.200) in the SLA group; crude mean difference -0.033 (95% CI -0.040 to -0.026). The adjusted mean difference in EQ-5D index score at 12 months was -0.036 (95% CI -0.044 to -0.029). The corresponding Z-score for the EQ-5D index score at 12 months between the crude and adjusted differences was 0.613, indicating no relevant change in treatment effect estimate after adjustment for observed potential confounders. Also, for the other outcomes, the change in effect estimate was relatively small, with Z-scores <1.

For the other comparisons (PLA vs. AA, SLA vs. AA), larger Z-scores were observed, owing to the observed baseline incomparability (Tables A1 and A2 in the Appendix). For example, the comparison PLA versus AA, the corresponding Z-score for the EQ-5D index score at 12 months between the crude and adjusted differences was 5.984.

Example 2. fixation of THA

Patient characteristics

In total, 108,939 patients were included in example 2. The characteristics are shown in Table 3. The cemented group included 29,579 patients (27%) and the uncemented group 79,360 patients (73%). There were meaningful imbalances of preoperative patient characteristic in the comparison of the cemented versus uncemented regarding age (SMD 0.913), sex (SMD 0.258), ASA classification (SMD 0.384), Charnley classification (SMD 0.126), and smoking (SMD 0.102).

Outcomes

The results of the crude and adjusted regression analyses are shown in Table 4. The mean EQ-5D index score at 12 months was 0.824 (SD 0.206) in the cemented group, compared to 0.877 (SD 0.176) in the uncemented group; crude mean difference -0.053 (95% CI -0.058 to -0.048). The adjusted mean difference in EQ-5D index score at 12 months was -0.022 (95% CI -0.028 to -0.016). The corresponding Z-score for the EQ-5D index score at 12 months between the crude and adjusted differences was -8.646, indicating a relevant change in treatment effect estimate after adjustment for observed potential confounders.

Discussion

The aim of this study was to assess the potential for confounding in observational studies of elective surgical interventions using routinely collected data. In the first example on surgical approach (PLA, SLA, and AA) of THA, the primary comparison between the two traditional approaches, PLA versus SLA, showed no meaningful differences in preoperative patient characteristics (SMD < 0.1) and also no relevant impact of adjustment for baseline information (Z-scores < 1). The PLA and AA groups and the SLA and AA groups differed slightly regarding various patient characteristics. The AA is a relative new approach and is thought to include a steep learning

Table 1
Baseline characteristics of patients undergoing total hip arthroplasty, stratified by surgical approach.

	Posterolateral	Straight lateral	Anterior	SMD	SMD	SMD
N	73,750	16,557	30,595	PLA versus SLA	PLA versus AA	SLA versus AA
Age	69.83 (9.75)	70.14 (9.76)	68.78 (9.51)	0.032	0.109	0.141
Sex (%)						
Male	25,554 (34.7)	5600 (33.8)	10,246 (33.5)	0.018	0.025	0.007
Female	48,101 (65.3)	10,950 (66.2)	20,341 (66.5)			
ASA classification (%)						
ASA I	12,250 (16.6)	2857 (17.3)	6537 (21.4)	0.047	0.172	0.131
ASA II	48,203 (65.4)	11,001 (66.5)	20,137 (65.9)			
ASA III-IV	13,215 (17.9)	2678 (16.2)	3897 (12.7)			
Previous operation (%)						
Yes	1557 (2.1)	424 (2.6)	269 (0.9)	0.032	0.103	0.133
No	70,958 (97.9)	15,670 (97.4)	29,844 (99.1)			
BMI (%)						
Underweight (<18.5)	448 (0.6)	97 (0.6)	198 (0.6)	0.011	0.178	0.188
Normal weight (18.5–25)	21,793 (30.2)	4761 (29.9)	11,049 (36.3)			
Overweight (25–30)	31,051 (43.0)	6819 (42.8)	13,147 (43.2)			
Obese (30–40)	17,944 (24.8)	4034 (25.3)	5880 (19.3)			
Class 3 Obese (>40)	1022 (1.4)	227 (1.4)	189 (0.6)			
Charnley classification (%)						
A	31,432 (43.6)	7274 (44.7)	13,920 (45.8)	0.038	0.097	0.074
B1	21,749 (30.2)	4979 (30.6)	9617 (31.7)			
B2	16,987 (23.6)	3577 (22.0)	6285 (20.7)			
C	1953 (2.7)	447 (2.7)	544 (1.8)			
Smoking (%)						
Yes	7425 (10.7)	1713 (11.5)	2963 (9.8)	0.025	0.029	0.054
No	61,815 (89.3)	13,166 (88.5)	27,135 (90.2)			
Fixation method (%)						
Cemented	22,078 (30.0)	4280 (25.9)	3221 (10.5)	0.094	0.517	0.439‘
Hybrid	7384 (10.0)	1881 (11.4)	2607 (8.5)			
Uncemented	44,229 (60.0)	10,375 (62.7)	24,756 (80.9)			
EQ-5D index score	0.54 (0.28)	0.55 (0.28)	0.59 (0.26)	0.009	0.188	0.179
NRS pain score during activity	7.27 (2.07)	7.28 (2.02)	6.88 (2.18)	0.003	0.188	0.192
NRS pain score at rest	5.34 (2.55)	5.14 (2.63)	4.98 (2.60)	0.075	0.139	0.063
HOOS-PS score	50.06 (18.02)	48.92 (18.22)	46.81 (17.66)	0.063	0.182	0.118
OHS score	22.45 (8.67)	22.33 (8.69)	23.78 (8.48)	0.015	0.155	0.170

EQ-5D = EuroQol-5 dimensions; HOOS-PS = hip disability and osteoarthritis outcome score; NRS = numeric rating scale for pain; OHS = oxford hip score; SMD = standardized mean difference.
Continuous variables presented as mean (SD).

Table 2
Patient-reported outcome measures of patients receiving total hip arthroplasty, stratified by surgical approach.

	Posterolateral		Straight lateral		Crude			Adjusted			
	Mean	SD	Mean	SD	Mean difference	95% CI		Mean difference	95% CI		Z-score
EQ-5D score											
3 mo	0.821	0.188	0.794	0.189	-0.027	-0.034	-0.02	-0.027	-0.034	-0.019	-0.075
12 mo	0.859	0.188	0.826	0.200	-0.033	-0.04	-0.026	-0.036	-0.044	-0.029	0.613
NRS activity											
3 mo	2.103	2.277	2.458	2.389	0.355	0.271	0.438	0.352	0.262	0.443	0.034
12 mo	1.470	2.184	1.961	2.508	0.491	0.403	0.579	0.507	0.410	0.603	-0.252
NRS at rest											
3 mo	1.190	1.864	1.315	1.979	0.125	0.056	0.194	0.144	0.07	0.218	-0.379
12 mo	0.873	1.745	1.177	2.070	0.305	0.234	0.375	0.342	0.265	0.42	-0.744
HOOS-PS score											
3 mo	18.276	14.458	21.707	14.68	3.43	2.871	3.99	3.519	2.923	4.115	-0.219
12 mo	13.800	14.824	17.784	16.753	3.984	3.376	4.593	4.142	3.485	4.799	-0.360
OHS score											
3 mo	39.003	7.791	37.819	7.339	-1.184	-1.481	-0.887	-1.100	-1.412	-0.788	-0.395
12 mo	41.873	7.418	40.173	7.909	-1.699	-1.998	-1.401	-1.630	-1.944	-1.315	-0.324

EQ-5D = EuroQol-5 dimensions; HOOS-PS = hip disability and osteoarthritis outcome score; NRS = numeric rating scale for pain; OHS = oxford hip score.
Z-score magnitude and direction of the change between the crude and adjusted mean difference.

curve, which might explain these differences [37]. In the second example on fixation method (cemented vs. uncemented) of THA, there were several meaningful imbalances in patient characteristic between treatment groups (SMD > 0.1), and there was a relevant impact of adjustment for baseline characteristics (Z-scores > 2).

Our aim was to evaluate comparability of patients receiving different elective surgical orthopedic interventions. We did not look into the differences in effect estimates between observational studies and RCTs, which have been investigated in previous research. Ioannidis et al. [38] and Hemkens et al. [39] evaluated the results of randomized and nonrandomized studies for a variety of topics and found that observational studies overestimate treatment effects compared to RCTs. In contrast, Benson et al. [40] and Concato et al. [41] found little evidence for systematic differences between results of observational studies and RCTs. Focusing on surgical interventions, Abraham et al. [42] found that results of high-quality observational studies were similar to those of RCTs. Clearly, based on these studies, one cannot conclude that results of obser-

Table 3
Baseline characteristics of patients undergoing total hip arthroplasty, stratified by fixation method.

	Cemented	Uncemented	SMD
N	29,579	79,360	
Age	75.37 (8.14)	67.39 (9.32)	0.913
Sex (%)			
Male	7685 (26.0)	30,071 (37.9)	0.258
Female	21,867 (74.0)	49,212 (62.1)	
ASA classification (%)			
ASA I	2924 (9.9)	16,612 (21.0)	0.384
ASA II	19,483 (65.9)	52,104 (65.7)	
ASA III–IV	7142 (24.2)	10,559 (13.3)	
Previous operation (%)			
Yes	637 (2.2)	1314 (1.7)	0.035
No	28,695 (97.8)	76,271 (98.3)	
BMI (%)			
Underweight (<18.5)	267 (0.9)	412 (0.5)	0.072
Normal weight (18.5–25)	9632 (33.2)	24,251 (31.1)	
Overweight (25–30)	12,211 (42.1)	33,914 (43.5)	
Obese (30–40)	6504 (22.4)	18,511 (23.7)	
Obese (30–40)	6504 (22.4)	18,511 (23.7)	
Class 3 Obese (>40)	395 (1.4)	871 (1.1)	
Charnley classification (%)			
A	11,794 (40.4)	35,275 (45.3)	0.126
B1	9042 (31.0)	24,072 (30.9)	
B2	7383 (25.3)	16,762 (21.5)	
C	943 (3.2)	1683 (2.2)	
Smoking (%)			
Yes	2367 (8.4)	8511 (11.4)	0.102
No	25,905 (91.6)	66,097 (88.6)	
Surgical approach (%)			
Posterolateral	22,078 (74.6)	44,229 (55.7)	0.519
Straight lateral	4280 (14.5)	10,375 (13.1)	
Anterior	3221 (10.9)	24,756 (31.2)	
EQ-5D index score	0.51 (0.29)	0.58 (0.27)	0.229
NRS pain score during activity	7.29 (2.11)	7.12 (2.09)	0.08
NRS pain score at rest	5.31 (2.65)	5.14 (2.56)	0.067
HOOS-PS score	50.97 (18.65)	48.08 (17.69)	0.159
OHS score	21.34 (9.01)	23.42 (8.44)	0.238

EQ-5D = EuroQol-5 dimensions; HOOS-PS = hip disability and osteoarthritis outcome score; NRS = numeric rating scale for pain; OHS = oxford hip score; SMD = standardized mean difference.
Continuous variables presented as mean (SD).

vational studies are always different from those of RCTs, nor that they always concur. It probably largely depends on the type of interventions being compared, the context in which the comparison is made, and the quality of the observational study including the data being used [10]. Our findings support the viewpoint that, when treatment decisions are largely independent of patient characteristics, one can indeed speculate that groups of patients who undergo different orthopedic surgical interventions will be comparable with respect to patient characteristics, and therefore results of such observational studies would be a valuable addition to evidence based on RCTs [3]. Availability of large regional and national registries of prospectively collected information show the feasibility of performing observational studies of medical interventions.

We hypothesized that the choice for PLA and SLA is largely independent of patient characteristics and the results of our study do not provide evidence against this hypothesis. Instead, we argued that this choice is mainly based on physician (or surgeon) preference. Physician preference has been used before in observational comparative effectiveness studies, in which case it is sometimes referred to as an instrumental variable [43]. In studies of surgical interventions, key assumptions of instrumental variable analysis are that there is substantial variation in preference between surgeons, that the preference is indeed independent of patient characteristics (i.e., different physicians operate similar patients), and that differences in outcomes between physicians are only due to the differences in the surgical technique that is investigated. The latter implies, for example, that learning effects do not differ between physicians. In the study comparing PLA and SLA, we assumed that learning effects were negligible for these established surgical approaches.

Both studies that we conducted assessed the comparability of groups of patients who all received a surgical intervention, be it that the approach or type of intervention differed. The potential for confounding may be larger in studies in which pharmacological treatments for surgical patients are compared. Also, studies comparing a surgical intervention to a conservative treatment may be more prone to confounding [10].

In addition to the potential for confounding, other sources of bias in observational studied should be considered. Electronic health record data may be affected by for instance errors in data linkage, misclassification, and missing values, all of which could also impact the quality of observational research using these data [44]. The data used in this study were extracted from the LROI, a prospective longitudinal cohort containing high-quality data. The completeness of the LROI data (100% coverage of Dutch hospitals) is checked against the hospital information systems and currently exceeds 99% for primary THAs [26–28]. Hence, the phenomena observed in this study are not necessarily to be expected in other observational studies. Blinding of the outcome assessor is typically not implemented in observational studies, nor is it in the LROI. However, in case of patient reported outcome measures, like the

Table 4
Patient-reported outcome measures of patients receiving total hip arthroplasty, stratified by fixation method.

	Cemented Mean	SD	Uncemented Mean	SD	Crude Mean difference	95% CI		Adjusted Mean difference	95% CI		Z-score
EQ-5D score											
3 mo	0.797	0.199	0.841	0.179	-0.044	-0.048	-0.039	-0.022	-0.027	-0.016	-6.686
12 mo	0.824	0.206	0.877	0.176	-0.053	-0.058	-0.048	-0.022	-0.028	-0.016	-8.646
NRS activity											
3 mo	2.244	2.387	1.965	2.228	0.279	0.224	0.335	0.200	0.134	0.266	1.975
12 mo	1.654	2.315	1.364	2.134	0.230	0.168	0.292	0.180	0.110	0.251	2.637
NRS at rest											
3 mo	1.356	2.031	1.100	1.799	0.256	0.210	0.301	0.153	0.100	0.207	3.117
12 mo	0.991	1.876	0.809	1.695	0.182	0.136	0.227	0.055	-0.002	0.111	3.837
HOOS-PS score											
3 mo	20.301	15.223	16.709	14.178	3.593	3.215	3.971	1.602	1.156	2.047	7.298
12 mo	16.434	16.091	12.257	14.214	4.177	3.767	4.587	1.330	0.839	1.821	9.631
OHS score											
3 mo	37.625	8.359	40.029	7.277	-2.405	-2.601	-2.208	-1.549	-1.779	-1.318	-6.033
12 mo	40.228	8.015	42.702	7.065	−2.474	-2.678	-2.271	-1.120	-1.361	-0.879	-9.219

EQ-5D = EuroQol-5 dimensions; HOOS-PS = hip disability and osteoarthritis outcome score; NRS = numeric rating scale for pain; OHS = oxford hip score.
Z-score magnitude and direction of the change between the crude and adjusted mean difference.

ones used in this study, such blinding is unfeasibly. Regarding the primary outcome of our study, that is, EQ-5D, we note that a three-level version was used. Nowadays, five-level versions of the EQ-5D may be more commonly used in research, yet this information was unavailable for the study period considered (2014–2018).

The aim of this study was not to provide evidence on the relative benefits of the different discussed surgical techniques, nor do these studies provide evidence for all studies of elective surgical treatment options. It does, however, provide support that there are cases in which observational studies of surgical treatment options are viable and provide valuable information. Particularly studies of surgical treatments might be less sensitive to confounding if treatment preference is not subject to patient characteristics and "allocation to" a particular intervention is close to a random process. It is up to the researchers of such studies to provide the arguments to substantiate the claim that treatment groups are expected to be comparable and why a particular research question could be answered using an observational study design.

Conclusion

This study using data from the nationwide LROI provides insight in the potential for confounding in observational studies of THA and exemplifies what considerations need to be made when using routinely collected data for studies of elective surgical interventions. Particularly studies of surgical interventions might be less sensitive to confounding if allocation of the intervention does not materially depend on patient characteristics. The comparison between surgical approaches (PLA vs. SLA) for THA is an example of this.

Ethical review committee statement

Ethics approval and consent to participate was not applicable, as all data were received completely anonymous. Ethical approval was not required for this study. Routinely collected data from the nationwide Dutch Arthroplasty Register (LROI) were extracted for this study.

Contributors

RMH and RHHG conceived and designed the study. YO, RHHG, MGJG, and RMH acquired, analyzed, and interpreted the data. YO drafted the first version of the manuscript. YO, RHHG, MGJG, and RMH drafted the subsequent versions of the manuscript. RGHHN and LPHL gave clinical advice in interpreting the results. Finally, all authors contributed critically to subsequent revisions and approved the final manuscript.

References

[1] Vandenbroucke JP. Observational research, randomized trials, and two views of medical science. PLoS Med. 2008;5(3):e67.
[2] Black N. Why we need observational studies to evaluate the effectiveness of health care. Bmj 1996;312(7040):1215.
[3] Houwert RM, Beks RB, Dijkgraaf MGW, Roes KCB, Oner FC, Hietbrink F, et al. Study methodology in trauma care: towards question-based study designs. Eur J Trauma Emerg Surg 2019;47(2):479–84.
[4] McCulloch P, Altman DG, Campbell WB, Flum DR, Glasziou P, Marshall JC, et al. No surgical innovation without evaluation: the IDEAL recommendations. Lancet 2009;374(9695):1105.
[5] Van Spall HG, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. JAMA 2007;297(11):1233.
[6] Frieden TR. Evidence for health decision making - beyond randomized, controlled trials. N Eng J Med 2017;377:465.
[7] Arditi C, Burnand B, Peytremann-Bridevaux I. Adding non-randomised studies to a Cochrane review brings complementary information for healthcare stakeholders: an augmented systematic review and meta-analysis. BMC Health Serv Res 2016;16:598.
[8] Oberkofler CE, Hamming JF, Staiger RD, Brosi P, Biondo S, Farges O, et al. Procedural Surgical RCTs in Daily Practice: do surgeons adopt or is it just a waste of time? Ann. Surg. 2019;270(5):727.
[9] Chapman SJ, Shelton B, Mahmood H, Fitzgerald JE, Harrison EM, Bhangu A. Discontinuation and non-publication of surgical randomized controlled trials: observational study. Bmj 2014;349:g6870.
[10] Beks RB, Bhashyam AR, Houwert RM, van der Velde D, van Heijl M, Smeeing DPJ, et al. When observational studies are as helpful as randomized trials: examples from orthopedic trauma. J Trauma Acute Care Surg 2019;87(3):730.
[11] Smeeing DPJ, van der Ven DJC, Hietbrink F, Timmers TK, van Heijl M, Kruyt MC, et al. Surgical versus nonsurgical treatment for midshaft clavicle fractures in patients aged 16 years and older: a systematic review, meta-analysis, and comparison of randomized controlled trials and observational studies. Am J Sports Med 2017;45(8):1937.
[12] Beks RB, Ochen Y, Frima H, Smeeing DPJ, van der Meijden O, Timmers TK, et al. Operative versus nonoperative treatment of proximal humeral fractures: a systematic review, meta-analysis, and comparison of observational studies and randomized controlled trials. J Shoulder Elbow Surg 2018;27:1526.
[13] Ochen Y, Beks RB, van Heijl M, Hietbrink F, Leenen LPH, van der Velde D, et al. Operative treatment versus nonoperative treatment of Achilles tendon ruptures: systematic review and meta-analysis. BMJ 2019;364:k5120.
[14] Higgins BT, Barlow DR, Heagerty NE, Lin TJ. Anterior vs. posterior approach for total hip arthroplasty, a systematic review and meta-analysis. J Arthroplasty 2015;30(3):419.
[15] Putananon C, Tuchinda H, Arirachakaran A, Wongsak S, Narinsorasak T, Kongtharvonskul J. Comparison of direct anterior, lateral, posterior and posterior-2 approaches in total hip arthroplasty: network meta-analysis. Eur J Orthop Surg Traumatol 2018;28(2):255.
[16] Wang Z, Hou JZ, Wu CH, Zhou YJ, Gu XM, Wang HH, et al. A systematic review and meta-analysis of direct anterior approach versus posterior approach in total hip arthroplasty. J Orthop Surg Res 2018;13(1):229.
[17] Wang Z, Bao HW, Hou JZ. Direct anterior versus lateral approaches for clinical outcomes after total hip arthroplasty: a meta-analysis. J Orthop Surg Res 2019;14(1):63.
[18] Kucukdurmaz F, Sukeik M, Parvizi J. A meta-analysis comparing the direct anterior with other approaches in primary total hip arthroplasty. Surgeon 2019;17(5):291.
[19] den Daas A, Reitsma EA, Knobben BAS, Ten Have B, Somford MP. Patient satisfaction in different approaches for total hip arthroplasty. Orthop Traumatol Surg Res 2019;105(7):1277.
[20] Peters RM, van Beers L, van Steenbergen LN, Wolkenfelt J, Ettema HB, Ten Have B, et al. Similar superior patient-reported outcome measures for anterior and posterolateral approaches after total hip arthroplasty: postoperative patient-reported outcome measure improvement after 3 months in 12,774 primary total hip arthroplasties using the anterior, anterolateral, straight lateral, or posterolateral approach. J Arthroplasty 2018;33(6):1786.
[21] Morshed S, Bozic KJ, Ries MD, Malchau H, Colford JM. Comparison of cemented and uncemented fixation in total hip replacement: a meta-analysis. Acta Orthop 2007;78(3):315.
[22] Toossi N, Adeli B, Timperley AJ, Haddad FS, Maltenfort M, Parvizi J. Acetabular components in total hip arthroplasty: is there evidence that cementless fixation is better? J Bone Joint Surg Am 2013;95(2):168.
[23] Abdulkarim A, Ellanti P, Motterlini N, Fahey T, O'Byrne JM. Cemented versus uncemented fixation in total hip replacement: a systematic review and meta–analysis of randomized controlled trials. Orthop Rev (Pavia) 2013;5(1):e8.
[24] Yoon BH, Ha YC, Lee YK, Koo KH. Postoperative deep infection after cemented versus cementless total hip arthroplasty: a meta-analysis. J Arthroplasty 2015;30(10):1823.
[25] Kuijpers MFL, Hannink G, van Steenbergen LN, Schreurs BW. Total hip arthroplasty in young patients in the netherlands: trend analysis of >19,000 primary hip replacements in the Dutch arthroplasty register. J Arthroplasty 2018;33(12):3704.
[26] van Steenbergen LN, Denissen GA, Spooren A, van Rooden SM, van Oosterhout FJ, Morrenhof JW, et al. More than 95% completeness of reported procedures in the population-based Dutch Arthroplasty Register. Acta Orthop 2015;86(4):498.
[27] LROI. Completeness of registering hospitals and completeness of registered arthroplasties in the lroi based on the hospital information system in 2016. [accessed 03.04. 20]. https://www.lroi-report.nl/hip/proms/response/.
[28] Hesseling B, Mathijssen NMC, van Steenbergen LN, Melles M, Vehmeijer SBW, Porsius JT. Fast starters, slow starters, and late dippers: trajectories of patient-reported outcomes after total hip arthroplasty: results from a Dutch nationwide database. J Bone Joint Surg Am 2019;101(24):2175.
[29] Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001;33(5):337.
[30] Ferreira-Valente MA, Pais-Ribeiro JL, Jensen MP. Validity of four pain intensity rating scales. Pain 2011;152:2399.
[31] Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, Conaghan PG, et al. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008;16(5):551.
[32] Wylde V, Learmonth ID, Cavendish VJ. The Oxford hip score: the patient's perspective. Health Qual Life Outcomes 2005;3:66.
[33] Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach JJ. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ 2006;15(10):1121–32.

[34] Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med 2009;28(25):3083.
[35] Franklin JM, Pawar A, Martin D, Glynn RJ, Levenson M, Temple R, Schneeweiss S. Nonrandomized real-world evidence to support regulatory decision making: process for a randomized trial replication project. Clin. Pharmacol. Ther. 2020;107(4):817.
[36] R Core Team, R Foundation for Statistical Computing R: A Language and Environment for Statistical Computing. Vienna, Austria; 2019. https://www.R-project.org/
[37] Meermans G, Konan S, Das R, Volpin A, Haddad FS. The direct anterior approach in total hip arthroplasty: a systematic review of the literature. Bone Joint J 2017;99-b(6):732.
[38] Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001;286(7):821.
[39] Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey. BMJ 2016;352:i493.
[40] Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. N Eng J Med 2000;342:1878.
[41] Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Eng J Med 2000;342:1887.
[42] Abraham NS, Byrne CJ, Young JM, Solomon MJ. Meta-analysis of well-designed nonrandomized comparative studies of surgical procedures is as good as randomized controlled trials. J Clin Epidemiol 2010;63:238.
[43] Widding-Havneraas T, Chaulagain A, Lyhmann I, Zachrisson HD, Elwert F, Markussen S, et al. Preference-based instrumental variables in health research rely on important and underreported assumptions: a systematic review. J Clin Epidemiol 2021;139:269–78.
[44] Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Routinely collected data and comparative effectiveness evidence: promises and limitations. CMAJ 2016;188(8):E158.