Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The replicability of real-world evidence for the surgical treatment of proximal humeral fractures in older patients evaluated using sceptical p-value

Jeanette Köppe

jeanette.koeppe@ukmuenster.de

University of Münster

Karen Fischhuber

University of Münster

Josef Stolberg-Stolberg

University Hospital Münster

Ursula Marschall

BARMER Institute for Health System Research

Michael J. Raschke

University Hospital Münster

Andreas Faldum

University of Münster

J. Christoph Katthagen

University Hospital Münster

Article

Keywords: real world evidence, replication, sceptical p-value, health service research, proximal humeral fracture, reverse total shoulder arthroplasty

Posted Date: November 4th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6679075/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. JK, JSS and JCK report research funding from the German Society for Trauma Surgery sponsored by Stryker, out-side the submitted work. All the remaining authors declare no conflict of interest.

Version of Record: A version of this preprint was published at Scientific Reports on September 8th, 2026. See the published version at https://doi.org/10.1038/s41598-026-70157-6.

The replicability of real-world evidence for the surgical treatment of proximal humeral fractures in older patients evaluated using sceptical p-value

Jeanette Koeppe[1,2,*], Karen Fischhuber[1,2], Josef Stolberg-Stolberg[2,3], Ursula Marschall[4], Michael J. Raschke[3], Andreas Faldum[1], J. Christoph Katthagen[2,3]

[1]Institute of Biostatistics and Clinical Research, University of Münster, Schmeddingstr. 56, Münster, 48149, Germany.
[2]Research Group Mathematical Surgery, University of Münster, University Hospital Münster, Albert-Schweitzer-Campus 1, Building W1, Münster, 48149, Germany.
[3]Department of Trauma-, Hand- and Reconstructive Surgery, University Hospital Münster, Albert-Schweitzer-Campus 1, Building W1, Münster, 48149, Germany.
[4]BARMER Institute for Health System Research, BARMER, Lichtscheider Str. 89, Wuppertal, 42285, Germany.
[*]Corresponding author: Jeanette Koeppe: jeanette.koeppe@ukmuenster.de.

Abstract

Background: The replicability of study results plays a central role, particularly in the context of real-world evidence (RWE). However, the evaluation of the agreement between two study results is not straightforward and different metrics are available. As an example for RWE, the surgical treatment of proximal humeral fractures in older patients using locked plate fixation (LPF) or reverse total shoulder arthroplasty (RTSA) was analyzed by two different studies based on the data of the largest German health insurance company [1, 2], but the generalizability of the results should be further addressed.
Methods: Independent data of a different German insurance company were used to emulate two existing studies for the surgical treatment of proximal humeral fracture in older patients using RTSA vs LPF. Study protocol was a priori registered. The sceptical p-value was used to judge the agreement between known original effect estimate (OES) and the observed emulated treatment effect estimate (EES) and compared to other existing binary agreement metrics.
Results: Overall, 7 (64%) of the 11 evaluated endpoints were successfully replicated. The Pearson correlation coefficient ($\boldsymbol{r_p} = 0.97$) and the intra-class correlation coefficient (ICC3 = 0.94) indicated a good, positive correlation between original and emulation study. OES was successfully replicated for all short-term endpoints and surgical complications during follow-up (all $\boldsymbol{p_{\text{scep}} < 0.025}$).
Conclusion: All treatment effects of the original study for short-term endpoints and for surgical complications during follow-up were successfully replicated using an independent data base. The sceptical p value was a useful and well-interpretable metric to evaluate the agreement between original and emulation study.

Trial registration: A priori registered at ClinicalTrials.gov (Identifier: NCT06537024).
Level of Evidence: Level III, retrospective replication study.
Keywords: real world evidence, replication, sceptical p-value, health service research, proximal humeral fracture, reverse total shoulder arthroplasty.

1 Background

Randomized controlled clinical trials (RCTs) are the gold standard for proving the efficacy of a treatment, but also comprise several limitations based on the typically small and highly selected patient cohorts [3–6]. To proof the efficacy of a new treatment, a study environment is established that can be compared to a clean laboratory setting and is idealized to the highest degree for the cause of bias reduction. However, the high standardization of treatment and the strict limitation of the patient population imply that the study environment does not always reflect the real situation in the healthcare system. Thus, due to a higher variability, there is evidence that drugs (or treatment in general) perform worse in daily clinical care than in RCTs, which was known in the literature as efficacy-effectiveness gap [4, 7]. However, systematic analyzes failed to proof general significant differences between the effect estimate in observational studies and RCTs [8–12]. Even though it is necessary to proof the efficacy of a (new) therapy using RCTs, the potential lack of external validity of RCTs consequences that the findings are not necessarily generalizable to all patients in daily clinical healthcare, making post-market analysis of treatments in actual real-world care necessary to ensure patient safety [4, 13, 14]. Furthermore, studies based on so-called real-world data (RWD) play an increasingly important role in regulatory decisions [14–16] and complement the information provided by RCTs. RWD have a great potential to create real-world evidence (RWE) due to growing availability of new data sources, especially for research questions that are difficult to address otherwise, e.g. for rare diseases or patients with comorbidities, often defined as exclusion criteria [14, 17–19].

Current use cases evaluated the surgical treatment of proximal humeral fracture (PHF) in the older people using German statutory health insurance data [1, 2, 20–22]. Due to an aging society, managing osteoporosis-related fractures is a growing global challenge [23]. They were responsible for the loss of annually more than 37 billion EUR and more than one million quality-adjusted years in Europe [24]. With an incidence in older patients of approx. 290 – 354/100,000 population-years [25, 26], the PHF is the third-most common fracture in older patients. Along with a fracture fixation using locked plate fixation (LPF), the reverse total shoulder arthroplasty (RTSA) is one of the most common surgical treatment options for older patients and has been increasingly used over the last decade [1]. However, even among highly specialist surgeons, a consensus about the optimal treatment of PHF is still missing, also caused by the small number of RCTs with contradictory results [27–31]. Based on German health claims data of more than 50,000 patients, it was observed that RTSA compared to LPF was associated with higher complications during short-term [1], but had improved outcome on long-term [2]. These findings highlight the utility of RWD in assessing treatment safety and effectiveness and underscore the importance of tailoring treatment to individual risk profiles.
However, in the case of studies based on insurance data, the generalizability of the results to the entire population is sometimes questioned, especially when the insurance holders differ socio-economically between different companies. It is hence important to check the replicability of the results in order to strengthen their credibility, but remains a challenge due to unique data structures [17].

In general, there is a distinction between direct and conceptual replication. A direct replication (also named reproduction) is defined as an evaluation of an independent investigators to obtain the same results when applying the same design and operational choices at the same data source (e.g. same insurance company, but with new data collection) [32, 33]. As an example, Wang et al. [34] directly replicated 150 RWD studies using the same database and the methods described in the original paper to emulate the design. Even though the majority of the results were replicated, there are subgroups in which the results could not be reproduced, which was explained by incomplete or non-transparent reporting and updated data [34]. The analysis showed impressively the importance of transparent reporting of the methodology in the context of RWE. In contrast, a conceptual replication is defined as an evaluation of an independent investigators to obtain the similar findings

when applying the same design and operational choices at different/ independent data sources (e.g. different insurance company) [32]. As an example for conceptual replications, target trial emulations have become increasingly important in recent years. As in the original formulation, a target trial emulation is an emulation of a "hypothetical" RCT, using a two-step process [35, 36]. In the first step, the causal questions are defined in a protocol for the hypothetical RCT, which has to specify the key elements to define the causal estimands (i.e. eligibility criteria, treatment strategies, treatment assignment, the start and end of follow-up, outcomes, causal contrasts) and the data analysis plan [35]. The RCT described in the protocol is defined as the target study for causal inference, which will be emulated using observational data in the second step [35, 36]. Emulations of RCTs using RWD are used for regulatory decision making, but are also used to evaluate, why an emulation of RCTs in daily healthcare settings leads to different results [14–16, 37–39]. As one example, the RCT DUPLICATE initiative (Randomized, Controlled Trials Duplicated Using Prospective Longitudinal Insurance Claims: Applying Techniques of Epidemiology) conducted the emulation of 32 RCTs (conceptual replications) with RWD to check, if and under which circumstances RWD studies explicitly designed to emulate RCTs are useful to draw the same causal conclusions [37, 38, 40].

Whether direct or conceptual, the evaluation of the agreement between two study results (i.e. two treatment effects) is not straightforward and different metrics are available, e.g. checking regulatory and estimated agreement or determining the sceptical p-value [37, 40–44]. However, even though different methods are available, an established standard method to evaluate the replicability is missing today [42], but would further improve the evidence of conceptual replications, especially in the case of target trials emulations of an existing RCT or RWE study.

In the study at hand, the findings for the surgical treatment of proximal humeral fractures in older patients, as observed in [1] and [2], were conceptually replicated using an independent database from a different insurance company within the same healthcare system. The agreement of the study results was assessed using sceptical p-value and compared to various binary metrics. The aim of the study is to evaluate the replicability of these findings and to investigate how the sceptical p-value performs in judging the agreement in the context of RWE.

2 Methods

2.1 Emulation study

Data cohort of the emulation study

Study design and statistical methods were pre-defined and a priori registered at ClinicalTrials.gov (Identifier: NCT06537024; see study protocol) before the first effect size was calculated.

The original studies by [1] and [2] were based on data from the Federal Association of Local Health Insurance Funds (AOK). Retrospective health claims data (in- and outpatient data) of the German BARMER health insurance from 2005 to 2022 were available for the emulation study. Diagnoses were defined based on International Statistical Classification of Diseases, German Modification (ICD-10 GM) and procedures were coded using German procedure classification (*Operationen- und Prozedurenschlüssel* – OPS). The study cohort was defined in accordance to [1], to establish an emulation of the study design presented there. Thus, all older patients (age $\geq$ 65 years) with inpatient coded LPF (OPS 5-794.21, 5-794.k1) or RTSA (OPS 5-824.21) and with inpatient coded PHF (ICD-10 GM S42.2) from 01/2011 to 09/2022 were included to the study. The first hospitalization was defined as the index case. Exclusion criteria were defined in accordance to [1] and [2] (see CONSORT statement

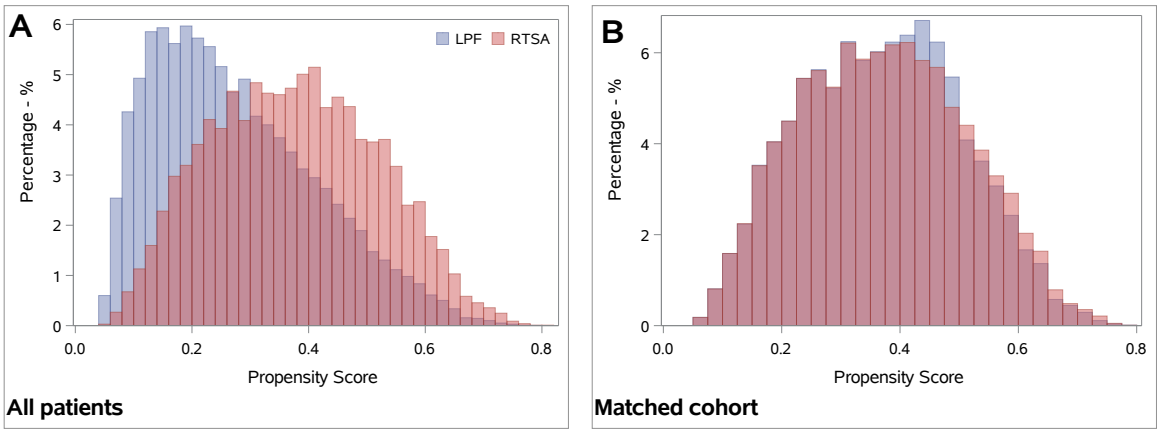


Fig. 1: Overlap of propensity score for entire cohort before matching (A) and 1:1 matched cohort. LPF – locked plate fixation; RTSA – revere total shoulder arthroplasty.

in Appendix figure A1). However, a COVID-19 infection during hospitalization was additionally defined as an exclusion criteria to avoid bias. All baseline parameters, such as comorbidities, prior surgery or prior medications, were analyzed within two years before the index hospitalization. The baseline characteristics of the emulation compared to the original study are presented in Table B1.

Statistical methods to determine effect sizes of the emulation study

Table 1: Summary of the analyzed primary and secondary endpoints.

Short-term [1]	**Long-term [2]**
primary endpoints	
In-hospital surgical complication (SC; binary endpoints)	Surgical complications after discharge (defined as time from discharge to surgical complication)
In-hospital Major adverse events (MAE; binary endpoints)	MAE after discharge (defined as time from discharge to first MAE)
In-hospital thrombo-embolic events (TE; binary endpoint)	TE or death after discharge (defined as time from discharge to TE or death from any cause)
30-day mortality (binary endpoint)	Overall survival (defined as time from discharge to death from any cause)
secondary endpoints	
Implant-assoc. complications (IAC; binary endpoint)	
Non-implant assoc. complications (NIAC; binary endpoint)	
	Minor outpatient complications (MOC; defined as time from discharge to minor outpatient complication)

Definitions can be found in [1, 2].

Table 1 presents all primary and secondary endpoints, which were checked for replicability with the treatment effect observed in the original studies of [1] and [2].

To determine the effect sizes of the emulation, a 1:1 propensity score matching (PSM) including age, sex and entire patient's risk profile was used. Balance between treatment groups were assessed using Z differences [45, 46], risk differences with 95% CI and sum of squared Z differences ($SSQZ_{diff}$), where a threshold $SSQZ_{diff} <$ "*number of included variables to PSM*" was considered to be sufficiently balanced (see table 2). Moreover, the overlap of the propensity score is presented in Figure 1.

The effect sizes for the treatment effect of RTSA vs LPF were determined as odds ratios (ORs) for short-term endpoints and hazard ratios (HRs) for long-term endpoints. For all endpoints, multivariable logistic regression models (binary endpoints) and multivariable Cox regression models (time-to-event endpoints), including age, sex, year of surgery and patient's individual risk profile, were used. In the case of surgical as well as minor outpatient complications after discharge, death was considered as a competing risk event and sub-distributional HRs using Fine&Gray Cox model

Table 2: Balance of the 1:1 propensity score matched cohort.

Variable	LPF	RTSA	risk difference (95%CI)	Z difference
Sample size – n (%)	9,896 (50.0%)	9,896 (50.0%)	n.a.	n.a.
Mean age (± SD) – years	79.9 (± 7.4)	80.0 (± 6.9)	-0.1 (-0.3; 0.1)	1.048
Mean index year (± SD) – years	2018.2 (±3.0)	2018.2 (±2.9)	-0.03 (-0.1; 0.1)	0.895
Female sex – n (%)	8,651 (87.4%)	8,637 (87.3%)	0.1% (-0.8%; 1.1%)	-0.299
Comorbidities – n (%)				
Alcohol abuse	547 (5.5%)	549 (5.5%)	-0.0% (-0.7%; 0.6%)	0.062
Atrial fibrillation/ flutter	2,055 (20.8%)	2,135 (21.6%)	-0.8% (-2.0%; 0.3%)	1.392
Atherosclerosis	1,752 (17.7%)	1,748 (17.7%)	0.0% (-1.0%; 1.1%)	-0.075
Cancer	2,732 (27.6%)	2,769 (28.0%)	-0.4% (-1.6%; 0.9%)	0.587
Congestive heart failure	2,553 (25.8%)	2,669 (27.0%)	-1.2% (-2.4%; 0.1%)	1.871
Coronary heart disease	2,497 (25.2%)	2,557 (25.8%)	-0.6% (-1.8%; 0.6%)	0.978
Chronic kidney disease	2,852 (28.8%)	2,924 (29.6%)	-0.7% (-2.0%; 0.5%)	1.126
Chronic polyarthritis	676 (6.8%)	693 (7.0%)	-0.2% (-0.9%; 0.5%)	0.476
Dementia	1,098 (11.1%)	1,126 (11.4%)	-0.2% (-1.2%; 0.6%)	0.630
Diabetes Mellitus	3,255 (32.9%)	3,320 (33.6%)	-0.7% (-2.0%; 0.7%)	0.981
Frozen shoulder	433 (4.4%)	428 (4.3%)	0.1% (-0.1%; 0.6%)	-0.174
Hypertension	8,669 (87.6%)	8,654 (87.5%)	0.2% (-0.8%; 1.1%)	-0.323
Nicotine abuse	704 (7.1%)	695 (7.0%)	0.1% (-0.6%; 0.8%)	-0.250
Obesity	2,527 (25.5%)	2,520 (25.5%)	0.1% (-1.1%; 1.3%)	-0.114
Omarthrosis	380 (3.8%)	381 (3.9%)	0.0% (-0.5%; 0.5%)	0.037
Osteoporosis	4,053 (41.0%)	4,053 (41.0%)	0.0% (-1.4%; 1.4%)	0.000
Parkinson disease	369 (3.7%)	394 (4.0%)	-0.3% (-0.8%; 0.3%)	0.923
Stroke and/or CVD before PHF	2,651 (26.8%)	2,637 (26.7%)	0.1% (-1.1%; 1.4%)	-0.225
Rotator cuff rupture	495 (5.0%)	503 (5.1%)	-0.1% (-0.7%; 0.5%)	0.260
Any anticoagulant	3,196 (32.3%)	3,275 (33.1%)	-0.8% (-2.1%; 0.5%)	1.197
Analgesics	1,501 (15.2%)	1,584 (16.0%)	-0.8%(-1.9%; 0.2%)	1.627
Antibiotics	172 (1.7%)	190 (1.9%)	-0.2% (-0.6%; 0.2%)	0.955
Any anti-osteoporotic drugs	1,296 (13.1%)	1,319 (13.3%)	-0.2% (-1.1%; 0.7%)	0.483
			Caliper width	0.1
			Number of variables	26
			Mean of $\|Z\|$	**0.650**
			Sum of squared Z-differences	**17.8**

Balance between treatment groups were assessed using risk differences with 95% confidence interval, Z differences, mean of absolute Z differences and sum of squared Z-differences. CI – confidence interval; – CVD – cerebrovascular disease; LPF – locked plate fixation; RTSA – reverse total shoulder arthroplasty; SD – standard deviation.

were presented. The dependence of the matching pairs were considered using stratified analysis for all endpoints. Full regression results for all endpoints are presented in table D3 – D13. In addition, all endpoints were presented descriptively, using proportions with 95% Clopper-Pearson confidence interval (CI) (binary variables), median and inter-quartile range (IQR, given by 25% quartile [Q1], and 75% quartile [Q3]; continuous variables) and event rates given by Aalen-Johansen estimates for the cumulative incidence functions (see table C2). Statistical analyzes were performed using the SAS software V9.4 (SAS Institute Inc., Cary, NC, USA) and R version 4.2.0, R foundation, Vienna, Austria.

Effect of COVID-19 pandemic

The original study was conducted prior to the SARS-CoV-2 pandemic. However, during the follow-up period of the study at hand, patients were influenced by the pandemic due to the infection itself on the one hand, but also influenced by lockdown periods and contact restrictions on the other hand. In addition, RTSA patients may had been affected more frequently due to an increasing proportion over study period. Thus, to address possible effects, two time-dependent variables were added to the Cox regression models. The first one reflects the beginning of the pandemic situation in Germany, and the second represent a diagnosed infection. As a sensitivity analysis, all regressions were repeated with COVID-19 infection were considered as a competing risk event and within a subset analysis including only patients with index surgery before 2020 ($N = 10,879$, with end of follow-up defined as 31.12.2019; see table E14 – E18).

2.2 Statistical agreement metrics to evaluate the replicability

After determining all emulation effect sizes (EESs), the sceptical p-value was used to check the agreement between EESs and original effect sizes (OESs).

The two-trials rule (TTR) is a standardized request of the FDA in the context of pivotal trials. It is required that at least two independent studies confirm the efficacy of a new drug [47, 48]. It is mostly implemented by two studies, which has to show significant results at a (one-sided) significance level of $\alpha = 0.025$ [47]. The combined p-value of the two trials rule is then given by the maximum of the (one-sided) p-values of the two studies, i.e. $p_{\text{TTR}} = \max(p_{\text{O}}, p_{\text{E}})$. $p_{\text{TTR}} \leq 0.025$ were interpreted as replication success. As an extension of the TTR, Held et al [42] recently suggested a novel method that combines a reverse-Bayes approach with a prior-predictive assessment to detect conflicts [44, 49]. The sceptical p-value p_{scep} is influenced by the p-values of each study and the ratio of variances between OEs and EES. Unlike treating the original and replication studies as interchangeable, this method imposes a penalty when the EES shrinks relative to the original one[43, 44]. Moreover, p_{scep} is an indirect measure of the degree of replication success [42]. It was shown that p_{spec} can be re-calibrated in such a way that an exact Type-I error control is fulfilled at level α^2, if the effect is null in both studies [44]. The sceptical p-value is a one-side p-value, to guarantee that replication is only successful, if both effects have the same sign. Further properties can be found in [42–44, 47, 49]. The controlled p_{spec} was determined using R package *ReplicationSuccess* version 1.3.2, $p_{\text{scep}} \leq 0.025$ were interpreted as replication success.

The following agreement metrics were additionally determined and compared to the conclusions from the sceptical p-value:

1. p-value from the TTR.
2. Regulatory agreement (RA): A binary agreement metric, which is 1, if the treatment effect estimates θ_O and θ_E (i.e. $\log(OR)$ or $\log(HR)$) are pointing in the same direction and both studies report the same statistical significance on the same significance level, or 0 otherwise. In the case of an original null effect, regulatory agreement is fulfilled, when both studies report no significance at the same significance level [37, 38].
3. Estimated agreement (EA): EES is enclosed to the 95% CI of OES [37, 38].
4. Standardized difference agreement (SDA) [37, 38]:

$$|Z| \leq 1.96, \text{ with } Z := \frac{\hat{\theta}_O - \hat{\theta}_E}{\sqrt{\hat{\sigma}_O^2 + \hat{\sigma}_E^2}}, \tag{1}$$

   where $\hat{\theta}_O$, $\hat{\theta}_E$ are the treatment effect estimates and $\hat{\sigma}_O^2$, $\hat{\sigma}_E^2$ are the associated variance of the original trial and emulation study, respectively [37, 38].
5. Fixed effect meta-analysis to combine the original effects and replication study: For positive OES, replication is fulfilled, when the meta-analysis provides a significant combined effect [41]. In the case of an original null effect, replication is fulfilled, when the combined effect size using meta-analysis indicates also null [41].

Moreover, Pearson's correlation and intra-class correlation (ICC3 - two-way mixed) between EES and OES were calculated for a general overview.

3 Results

Before PSM, N=32,952 patients were included to the emulation study. As presented in table B1, there are differences between the cohort of the original (AOK cohort) and emulation study

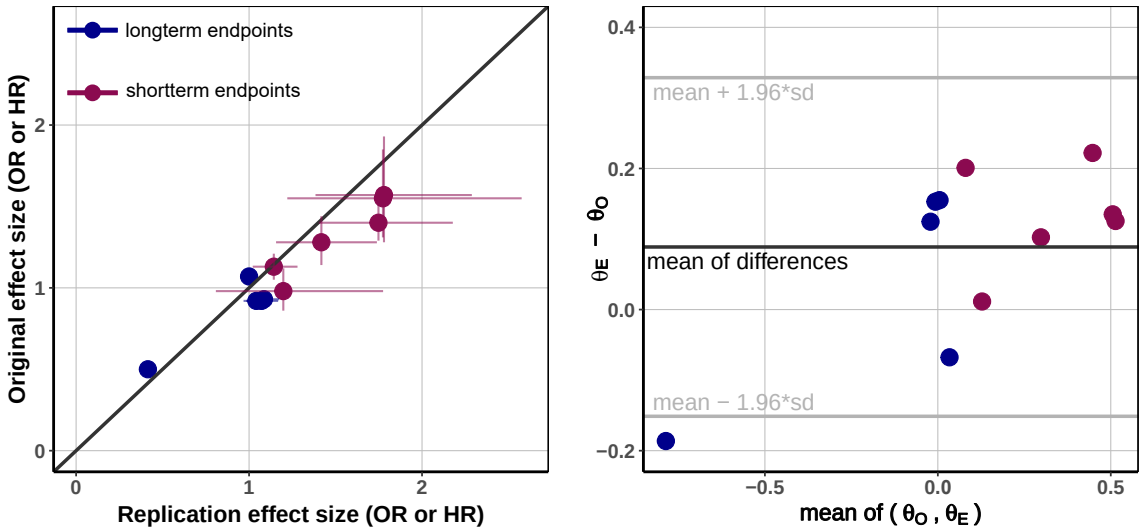


Fig. 2: Left: Calibration plot. Original effect estimates (odds ratio – OR or hazard ratio – HR) are plotted against the estimates of the emulation. **Right:** Bland Altman plot. The difference between the original and the emulation study model coefficients ($\theta_O - \theta_E$) are plotted against the averaged value for each pair.

(BARMER cohort). On the one hand, patients of the original study had more often a co-diagnosis of diabetes mellitus (AOK 40% vs BARMER 30%), of congestive heart failure (AOK 31% vs BARMER 23%) and of coronary heart disease (AOK 30% vs BARMER 24%) compared to the cohort of the emulation study. On the other hand, the emulation study had higher rates of patients with dementia (AOK 6% vs BARMER 10%) and higher rates of cancer (AOK 22% vs BARMER 27%). After 1:1 PSM, N=19,792 patients were included to the analysis set for further analysis.

Overall, 7 (64%) of the 11 evaluated endpoints were successfully replicated (s. table 3 and 4). The Pearson correlation coefficient of ($r_p = 0.97$; 95% CI 0.88 – 0.99) and the intra-class correlation coefficient (single-fixed raters; ICC3 = 0.94; 95% CI 0.78 – 0.98) indicated a good, positive correlation between original and emulation study. The mean differences of effect estimates ($\theta_O - \theta_E$) was 0.089 (95% CI 0.006 – 0.171), with a slight increasing trend in the Bland-Altman plot (see figure 2). However, there were distinct differences for short- and long-term endpoints.

Short-term endpoints

The effect estimates of the original study, the emulation study and the pooled effect from fixed-effect meta-analysis are presented in figure 3. For all analyzed endpoints, the pooled effect confirmed the results of the original study. RTSA was associated with higher risk for surgical complications (non-implant as well as implant-associated complications), higher risk for thrombo-embolic events and a higher risk for major adverse events during hospital stay (all $p < 0.05$). In both studies, no differences in 30-day mortality after surgery between RTSA and LPF were observed. Moreover, almost all analyzed agreement metrics are fulfilled, except of 30-day mortality (see table 3). Due to the original null effect for the 30-day mortality, the TTR resulted non-significant and the sceptical p-value cannot be determined. Nevertheless, the replication can also be considered successful, as various other criteria are met (see table 3).

long-term endpoints

In total, N=15.518 were included for the analysis of the follow-up outcome with exclusion criteria presented in figure A1. The effect estimates of the original study, the emulation study and the pooled effect from fixed-effect meta-analysis are presented in figure 4. For surgical complications after discharge, the original treatment effect was successfully replicated, because $p_{\text{scep}} < 0.001$, $p_{\text{TTR}} <$

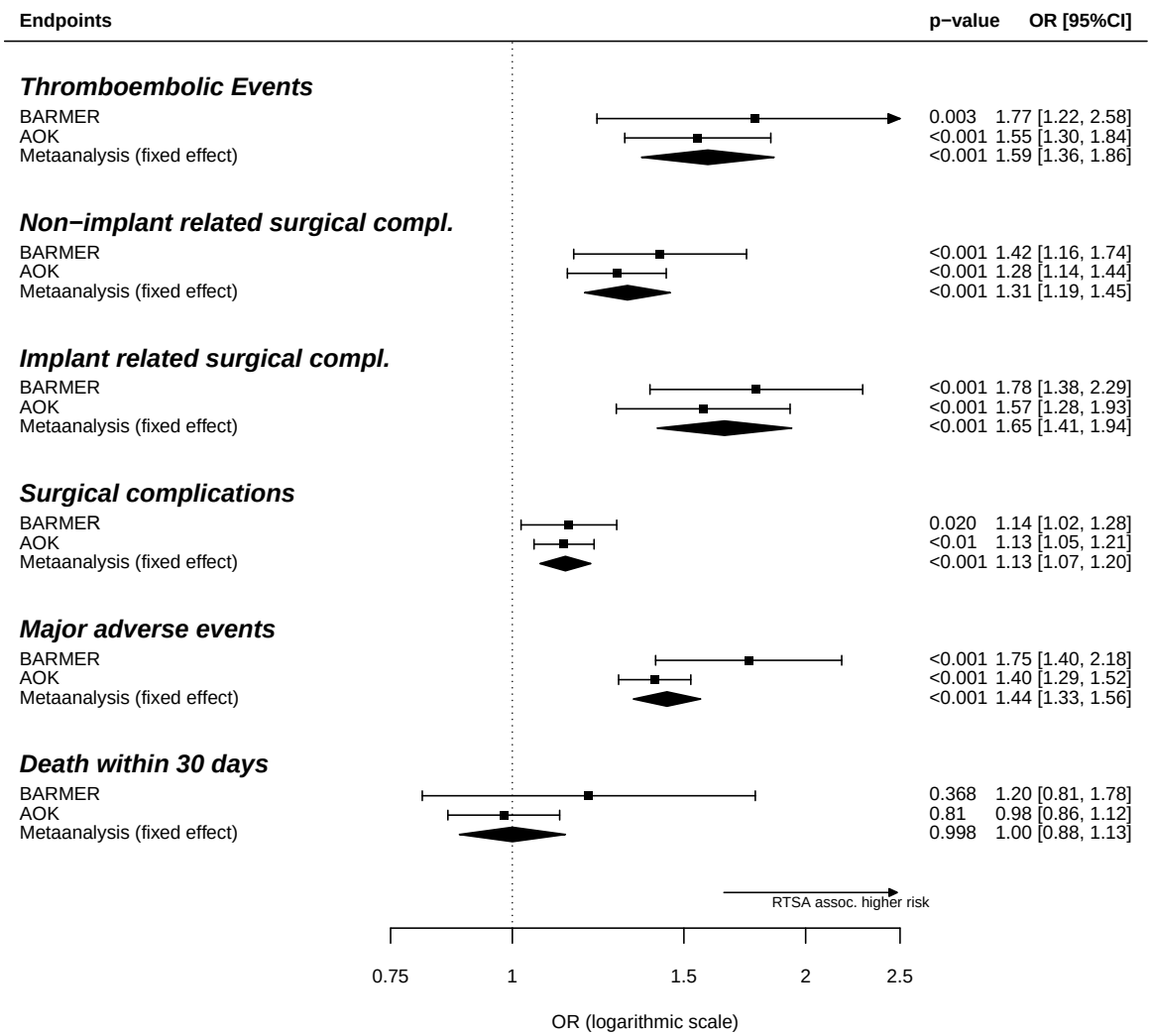


Fig. 3: Short-term outcomes of original study (AOK, [1]) and emulation study (BARMER). Combined effects determined using fixed-effect meta-analysis were included (diamond).

Table 3: Replication results for short-term outcomes

Agreement measure	30 day mortality	TE	MAE	SC	IAC	NIAC
Meta-analysis criterion fulfilled?	yes	yes	yes	yes	yes	yes
Regulatory agreement	yes	yes	yes	yes	yes	yes
Estimated agreement	no	yes	no	yes	yes	yes
SDA: $\lvert Z\rvert \leq 1.96$?	yes	yes	yes	yes	yes	yes
Standardized differences	0.948	0.643	1.85	0.168	0.758	0.853
p_{TTR}	0.610	0.001	<0.001	0.011	<0.001	<0.001
$p_{\textbf{scep}}$	n.a.†	<0.001	<0.001	0.006	<0.001	<0.001

† not applicable for null effects. Standardized differences defined as $Z := \hat{\theta}_O - \hat{\theta}_E / \sqrt{\hat{\sigma}_O^2 + \hat{\sigma}_E^2}$, with $\theta = \log(OR)$. IAC – implant-associated complication; MAE – major adverse event; NIAC – non-implant associated complication; SC – surgical complication; TE – thrombo-embolic event

0.001 and the fixed-effect meta-analysis confirmed the original treatment effect (see table 4). Since the treatment effect on surgical complications after discharge was even more pronounced in the emulation study (RTSA vs LPF: HR 0.41, 95%CI 0.38 – 0.46, p<0.001), the 95%CIs of the OES and EES did not overlap and EA and SDA were thus not fulfilled. For all other evaluated endpoints during follow-up, the replication was not successful, even if the pooled effects from meta-analysis were in accordance with the results of the original findings (see figure 4 and table 4).

Table 4: Replication results for long-term outcomes

Agreement measure	OS	TE	MAE	SC	MOC
Meta-analysis criterion fulfilled?	yes	yes	yes	yes	yes
Regulatory agreement	no	no	no	yes	no
Estimated agreement	no	no	no	no	no
SDA: $\lvert Z\rvert \leq 1.96$?	no	no	no	no	no
Standardized differences	3.177	3.592	3.024	-2.744	-2.037
p_{TTR}	0.942	0.980	0.864	<0.001	0.998
$p_{\mathbf{scep}}$	0.961	0.990	0.888	<0.001	0.512

Standardized differences defined as $Z := \hat{\theta}_O - \hat{\theta}_E / \sqrt{\hat{\sigma}_O^2 + \hat{\sigma}_E^2}$, with $\theta = \log(HR)$. MAE – major adverse event; MOC – minor outpatient complication; OS – overall survival; SC – surgical complication; TE – thrombo-embolic event

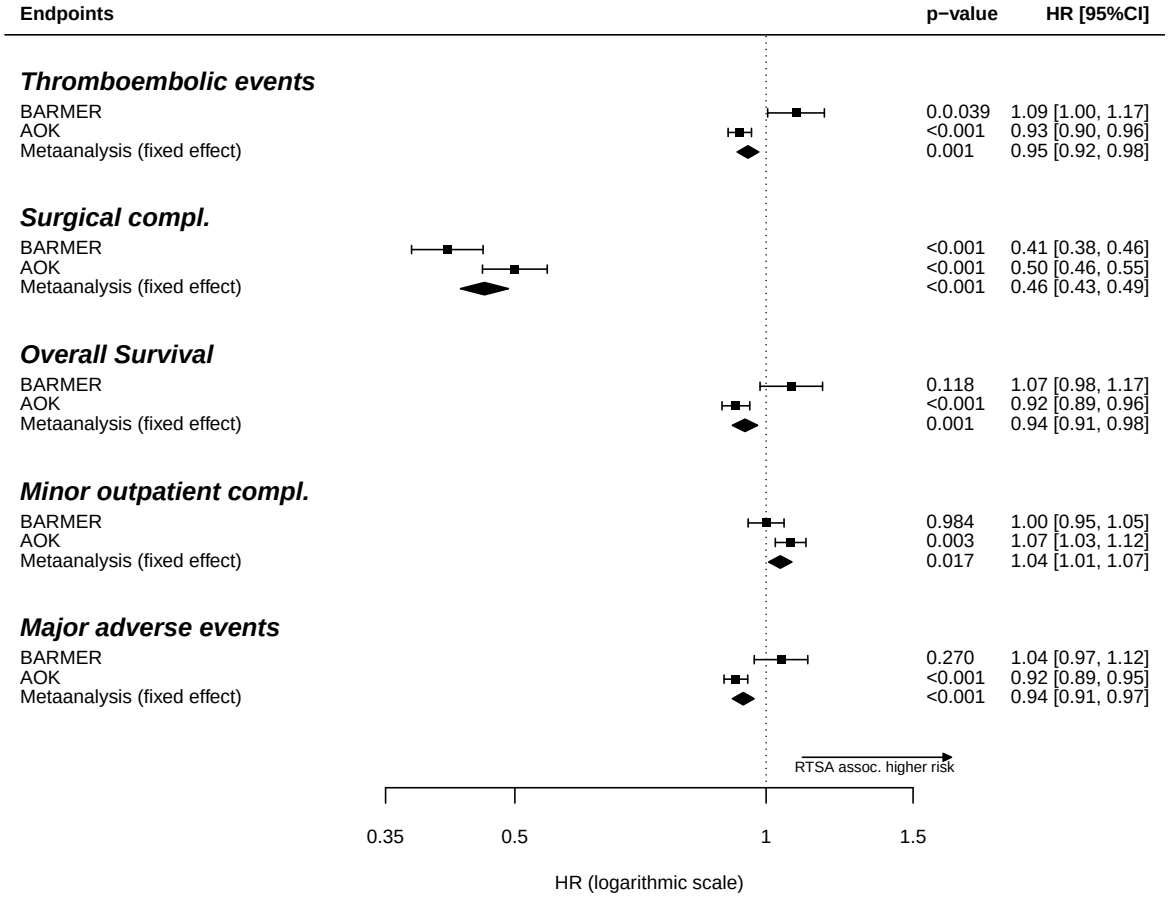


Fig. 4: Long-term outcomes of original study (AOK, [2]) and emulation study (BARMER). Combined effects determined using fixed-effect meta-analysis were included (diamond).

4 Discussion

For the implemented emulation study, all studied short-term outcomes and the treatment effect on surgical complication in the long-term after surgical treatment of PHF using LPF or RTSA were successfully replicated. The sceptical p-value was a well interpretable metric to judge the agreement of different studies also in the context of RWE.

The different agreement metrics have different advantages and disadvantages, which can lead to different judgments. For all analyzed enpoints, the TTR and the sceptical p-value came to the same conclusion. However, purely assessing the significance of the original and replication studies (TTR) is often used as a criterion, but it has the significant disadvantage of not considering the effect sizes of both, potentially leading to incorrect conclusions [44]. Additionally, the TTR has other major limitations, such as low project power or the inability to achieve successful replicability if the p-value of the original study is even slightly above the 5% mark (or 2.5% for one-sided hypotheses)

[44]. However, replicability of results that show a trend towards significance (i.e., barely missing the mark) is common practice, which can be also addressed using the sceptical p-value [44].
In the case of positive treatment effects, both, regulatory agreement as well as the TTR, are equivalent to each other. But, for null effects, the TTR is per definition not fulfilled, even though the second (emulation) study concluded also a null effect, which leads to discrepancies between the binary metric and the combined p-value. Furthermore, the sceptical p-value cannot be used for an original null effect and only binary agreement metrics (i.e. RA, EA or SDA) or a meta-analytic approach are available. Micheloud et al [50] presented a formalism for equivalence studies. However, when the original study was designed for superiority and a non-significant effect was observed (null-effect), it should not be confused with an equivalence test, as "*the absence of evidence is not evidence for absence*" [51]. However, in the case of an original null effect, it would be more appropriate to initiate a new study with the aim of proving the equivalence between treatments (with an equivalence test as the primary analysis) instead of starting a (further) replication study with the "'wrong design"'.

Meta-analytic approaches, could be useful to determined pooled effect estimates over different studies or different cohorts (i.e. different insurance companies). However, to evaluate the replication of a known effect, this approach has some distinct disadvantages. One the one hand, it is assumed that the studies are independent and exchangeable, which is not fulfilled in the case of a setting of original and emulation study [42, 47]. On the other hand, when the sample size of the original study is much larger than the one of the emulation study, it is possible that the meta-analytic approach confirm the original effect, even if the emulation study concluded qualitatively different effect estimates. This was the case for the studied long-term endpoints OS, MAE, TE and MOC where the 95%CI of the original effect estimates where smaller (for MOC greater) than one (all $p<0.001$), but the emulation failed to show any differences for RTSA compared to LPF.

Especially in the case of emulating RCTs using RWD, the regulatory, estimated and standardized difference agreement were common and often used [37, 38, 52–57]. However, the sceptical p-value has some advantages over the binary metrics, which were in detail described in [58]. In the presented study, EA and SDA were not fulfilled for surgical complications during long-term, as the 95% CIs of the original and emulated treatment effect did not overlap. Based on SDA and EA, one might conclude that the replication failed, even though the original effect was confirmed even more clearly. In contrast, the sceptical p-value leads to the correct conclusion.

As observed by Koeppe et al. [1], RTSA was associated with higher risk for different complications during hospital stay, but no differences were found for 30-day mortality. Moreover, the findings for surgical complications after discharge observed by Stolberg-Stolberg et al.[2] could also be replicated, i.e. RTSA compared to LPF was associated with a lower risk for surgical complications after discharge. Heyard et al [39] used meta-analytic methods to identify reasons for RWE results not being consistent with original RCT effects. They concluded that the most heterogeneity between RCT and related RWE emulation are caused by delayed effect of treatment, discontinuation of treatment during run-in period and treatment started in hospital. Wang et al [34] concluded that failure of direct replication was mostly explained by incomplete or in transparent reporting and updated data. However, none of the possible explanations reported in Heyard et al. [39] or Wang et al. [34] were present in the original studies, suggesting that a different source for the discrepancies must be assumed. As a potential explanation, the failure of replication success for the long-term endpoints overall survival, major adverse events and thrombo-embolic events could be caused by missing causal relationship between the treatment of PHF and events in the much later course of the follow-up. This is also underlined by the fact that, even years later, the occurrence of secondary shoulder surgery is very likely to be directly related to the past fracture and treatment and that the results for this endpoint (SC) of the original study were successfully replicated and thus confirmed.

Moreover, in the emulation, RTSA was even more clearly associated with less SCs than in the original study. However, even if there are statistically noticeable differences for OS, MAE and TE between original and emulation study, the observed effect is only of limited clinical relevance and, therefore, of secondary interest in the interpretation of content. In contrast, during hospital stay, all endpoints were successfully replicated, where the causal relationship between surgery and all endpoints seemed not unlikely. This may also indicate that replication is only successful, if a direct causal relationship between the treatment effect (to be tested) and the endpoint can be assumed. The observed discrepancies for the long-term endpoints OS, MAE and TE may then be due to the general differences between the insurance companies.

In general, replication of previous observed treatment effects is essential to increase the evidence of findings, especially in the context of real-world data studies. The sceptical p-value, originally developed for pivotal trials, also provides the opportunity for a comprehensive assessment of replication in the context of real-world evidence, which was demonstrated on the presented data an by Koeppe et al. [58]. Moreover, it is essential once again that the information on the design, definition of the cohort (including inclusion and exclusion criteria) and variables, as well as the statistical analysis, is transparent and fully available. The presented study illustrated how an evaluation of the agreement can further improve the evidence and validity of results from real-world data.

There are some limitations that have to be considered and are caused by the retrospective character of the study. First, as in the original study, the data are insurance data primarily collected for financial purposes. The database does not provide information about the underlying reason for the respective treatment with RTSA or LPF. The sceptical p-value is not applicable to replicate an observed null effect.

5 Conclusions

All treatment effects of the original study for short-term endpoints and for surgical complications during follow-up were successfully replicated using an independent data base. The sceptical p value was a very useful tool to evaluate the agreement between original and emulation study.

List of abbreviations

Abbreviation	Explanation
CI	Confidence interval
CONSORT	Consolidated standards of reporting trials
COVID-19	Corona virus Disease 2019
EES	Emulation effect estimate
HR	Hazard ratio
IAC	Implant-associated complications
ICD-10 GM	International Statistical Classification of Diseases, German Modification
IQR	Inter quartile range
LPF	Locked plate fixation
MAE	Major adverse event
MOC	Minor outpatient complications
NIAC	Non-implant associated complications
OES	Original effect size
OPS	*Operationen- und Prozeduren-Schlüssel* (German procedure classification)
OR	Odds ratio
PHF	Proximal humeral fracture
PI	Prediction interval
RCT	Randomized controlled trial
RWD	Real-world data
RWE	Real-world evidence
RTSA	Reverse total shoulder arthroplasty
SC	Surgical complications
TE	Thromboembolic event

Declarations

Funding. This work was supported by funding program „Innovative Medizinische Forschung“ (grant number I-KÖ122113), which is a faculty research support by the University of Münster, Germany.

Conflict of interest/Competing interests. JK, JSS and JCK report research funding from the German Society for Trauma Surgery sponsored by Stryker, out-side the submitted work. All the remaining authors declare no conflict of interest.

Ethics approval and consent to participate. The study was conducted in accordance with the Declaration of Helsinki and was approved by Ethics Committee Westfalen-Lippe (no. 2022-300-f-S). Due to the retrospective nature of the study, the need to obtain the informed consent was waived by the Ethics Committee Westfalen-Lippe.

Consent for publication. As for anonymity of insurance data, no prior written informed consent for publication of the data had to be obtained.

Data availability. The datasets generated and/or analysed during the current study are not publicly available due to German data protection laws (*'Bundesdatenschutzgesetz'*, BDSG). They are stored on a server of the BARMER Institute for Health System Research, to facilitate replication of the results. In general, access to data of statutory health insurance funds for research purposes is possible only under the conditions defined in German Social Law (SGB V § 287) and further inquiries can be directed to the corresponding author on reasonable request.

Materials availability. No.

Code availability. No.

Author contribution. **Conceptualization** JK, JCK, JSS; **methodology** JK; **software** JK, KF; **validation** KF; **formal analysis** JK; **investigation** JK; **resources** UM, MJR, AF; **data curation** UM, JK; **writing – original draft** JK; **writing – review & editing** JK, KF, JSS, UM, MJR, AF, JCK; **visualization** JK; **supervision** JK, JCK; **project administration** JK, JCK, **funding acquisition** JK, JCK, JSS. All authors have read and agreed to the published version of the manuscript.

Acknowledgments. None.

References

[1] Köppe, J., Stolberg-Stolberg, J., Rischen, R., Faldum, A., Raschke, M.J., Katthagen, J.C.: In-hospital Complications Are More Likely to Occur After Reverse Shoulder Arthroplasty Than After Locked Plating for Proximal Humeral Fractures. Clinical Orthopaedics and Related Research **479**(10), 2284–2292 (2021) https://doi.org/10.1097/CORR.0000000000001776

[2] Stolberg-Stolberg, J., Köppe, J., Rischen, R., Freistühler, M., Faldum, A., Katthagen, J.C., Raschke, M.J.: The Surgical Treatment of Proximal Humeral Fractures in Elderly Patients. Dtsch Arztebl International **118**(48), 817–823 (2021) https://doi.org/10.3238/arztebl.m2021.0326

[3] Black, N.: Why we need observational studies to evaluate the effectiveness of health care. BMJ **312**(7040), 1215–1218 (1996) https://doi.org/10.1136/bmj.312.7040.1215

[4] Eichler, H.G., Abadie, E., Breckenridge, A., Flamion, B., Gustafsson, L.L., Leufkens, H., Rowland, M., Schneider, C.K., Bloechl-Daum, B.: Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response. Nat Rev Drug Discov **10**(7), 495–506 (2011) https://doi.org/10.1038/nrd3501

[5] Corrigan-Curay, J., Sacks, L., Woodcock, J.: Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness. JAMA **320**(9), 867–868 (2018) https://doi.org/10.1001/jama.2018.10136

[6] Wang, S.V., Schneeweiss, S., Gagne, J.J., Evers, T., Gerlinger, C., Desai, R., Najafzadeh, M.: Using real-world data to extrapolate evidence from randomized controlled trials. Clinical Pharmacology & Therapeutics **105**(5), 1156–1163 (2019) https://doi.org/10.1002/cpt.1210

[7] Bhandari, M., Tornetta, P., Ellis, T., Audigé, L., Sprague, S., Kuo, J., Swiontkowski, M.: Hierarchy of evidence: Differences in results between non-randomized studies and randomized trials in patients with femoral neck fractures. Archives of orthopaedic and trauma surgery **124**, 10–6 (2004) https://doi.org/10.1007/s00402-003-0559-z

[8] Concato, J., Shah, N., Horwitz, R.I.: Randomized, controlled trials, observational studies, and the hierarchy of research designs. New England Journal of Medicine **342**(25), 1887–1892 (2000) https://doi.org/10.1056/NEJM200006223422507

[9] Golder, S., Loke, Y.K., Bland, M.: Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: Methodological overview. PLOS Medicine **8**(5), 1–13 (2011) https://doi.org/10.1371/journal.pmed.1001026

[10] Kuss, O., Legler, T., Börgermann, J.: Treatments effects from randomized trials and propensity score analyses were similar in similar populations in an example from cardiac surgery. Journal of Clinical Epidemiology **64**(10), 1076–1084 (2011) https://doi.org/10.1016/j.jclinepi.2011.01.005

[11] Lonjon, G., Boutron, I., Trinquart, L., Ahmad, N., Aim, F., Nizard, R., Ravaud, P.: Comparison of treatment effect estimates from prospective nonrandomized studies with propensity score analysis and randomized controlled trials of surgical procedures. The Milbank Quarterly **259**(1), 18–25 (2014) https://doi.org/10.1097/SLA.0000000000000256

[12] Anglemyer, A., Horvath, H.T., Bero, L.: Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database of Systematic Reviews **14**(4) (2014) https://doi.org/10.1002/14651858.MR000034.pub2

[13] Schneeweiss, S., Avorn, J.: A review of uses of health care utilization databases for epidemiologic

research on therapeutics. Journal of Clinical Epidemiology **58**(4), 323–337 (2005) https://doi.org/10.1016/j.jclinepi.2004.10.012

[14] Franklin, J.M., Glynn, R.J., Martin, D., Schneeweiss, S.: Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. Clinical Pharmacology & Therapeutics **105**(4), 867–877 (2019) https://doi.org/10.1002/cpt.1351

[15] US Food and Drug Administration, e.a.: Real-World Evidence. https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence

[16] Orsini, L.S., Berger, M., Crown, W., Daniel, G., Eichler, H.-G., Goettsch, W., Graff, J., Guerino, J., Jonsson, P., Lederer, N.M., Monz, B., Mullins, C.D., Schneeweiss, S., Brunt, D.V., Wang, S.V., Willke, R.J.: Improving transparency to build trust in real-world secondary data studies for hypothesis testing—why, what, and how: Recommendations and a road map from the real-world evidence transparency initiative. Value in Health **23**(9), 1128–1136 (2020) https://doi.org/10.1016/j.jval.2020.04.002

[17] Liu, F., Panagiotakos, D.: Real-world data: a brief review of the methods, applications, challenges and opportunities. BMC Medical Research Methodology **22**(1), 287 (2022) https://doi.org/10.1186/s12874-022-01768-6

[18] Franklin, J.M., Schneeweiss, S.: When and how can real world data analyses substitute for randomized controlled trials? Clinical Pharmacology & Therapeutics **102**(6), 924–933 (2017) https://doi.org/10.1002/cpt.857

[19] Booth, C.M., Karim, S., Mackillop, W.J.: Real-world data: towards achieving the achievable in cancer care. Nat Rev Clin Oncol **16**, 312–325 (2019) https://doi.org/10.1038/s41571-019-0167-7

[20] Koeppe, J., Katthagen, J.C., Rischen, R., Freistuehler, M., Faldum, A., Raschke, M.J., Stolberg-Stolberg, J.: Male sex is associated with higher mortality and increased risk for complications after surgical treatment of proximal humeral fractures. Journal of Clinical Medicine **10**(11) (2021) https://doi.org/10.3390/jcm10112500

[21] Koeppe, J., Stolberg-Stolberg, J., Rischen, R., Freistuehler, M., Faldum, A., Raschke, M.J., Katthagen, J.C.: Increased complication rates of salvage reverse total shoulder arthroplasty (rtsa) after failed locked plate fixation compared with primary rtsa in the treatment of proximal humeral fractures in elderly patients. Journal of Shoulder and Elbow Surgery **32**(8), 1574–1583 (2023) https://doi.org/10.1016/j.jse.2022.12.020

[22] Katthagen, J.C., Raschke, M.J., Fischhuber, K., Iking, J., Marschall, U., Sußiek, J., Faldum, A., Stolberg-Stolberg, J., Köppe, J.: Conservative Versus Operative Treatment of Proximal Humerus Fractures in Older Individuals. Dtsch Arztebl International **121**(14), 454–460 (2024) https://doi.org/10.3238/arztebl.m2024.0059

[23] Wu, A.-M., Bisignano, C., James, S.L., Abady, G.G., Abedi, A., Abu-Gharbieh, E., Alhassan, R.K., Alipour, V., Arabloo, J., Asaad, M., Asmare, W.N., Awedew, A.F., Banach, M., Banerjee, S.K., Bijani, A., Birhanu, T.T.M., Bolla, S.R., Cámera, L.A., Chang, J.-C., Cho, D.Y., Chung, M.T., Couto, R.A.S., Dai, X., Dandona, L., Dandona, R., Farzadfar, F., Filip, I., Fischer, F., Fomenkov, A.A., Gill, T.K., Gupta, B., Haagsma, J.A., Haj-Mirzaian, A., Hamidi, S., Hay, S.I., Ilic, I.M., Ilic, M.D., Ivers, R.Q., Jürisson, M., Kalhor, R., Kanchan, T., Kavetskyy, T., Khalilov, R., Khan, E.A., Khan, M., Kneib, C.J., Krishnamoorthy, V., Kumar, G.A., Kumar, N., Lalloo, R., Lasrado, S., Lim, S.S., Liu, Z., Manafi, A., Manafi, N., Menezes, R.G., Meretoja, T.J., Miazgowski, B., Miller, T.R., Mohammad, Y., Mohammadian-Hafshejani, A., Mokdad, A.H., Murray, C.J.L., Naderi, M., Naimzada, M.D., Nayak, V.C., Nguyen, C.T., Nikbakhsh,

R., Olagunju, A.T., Otstavnov, N., Otstavnov, S.S., Padubidri, J.R., Pereira, J., Pham, H.Q., Pinheiro, M., Polinder, S., Pourchamani, H., Rabiee, N., Radfar, A., Rahman, M.H.U., Rawaf, D.L., Rawaf, S., Saeb, M.R., Samy, A.M., Sanchez Riera, L., Schwebel, D.C., Shahabi, S., Shaikh, M.A., Soheili, A., Tabarés-Seisdedos, R., Tovani-Palone, M.R., Tran, B.X., Travillian, R.S., Valdez, P.R., Vasankari, T.J., Velazquez, D.Z., Venketasubramanian, N., Vu, G.T., Zhang, Z.-J., Vos, T.: Global, regional, and national burden of bone fractures in 204 countries and territories, 1990–2019: a systematic analysis from the global burden of disease study 2019. The Lancet Healthy Longevity **2**(9), 580–592 (2021) https://doi.org/10.1016/S2666-7568(21)00172-0

[24] Borgström, F., Karlsson, L., Ortsäter, G., Norton, N., Halbout, P., Cooper, C., Lorentzon, M., McCloskey, E.V., Harvey, N.C., Javaid, M.K., Kanis for the International Osteoporosis Foundation, J.A.: Fragility fractures in Europe: burden, management and opportunities. Archives of Osteoporosis **15**(59) (2020) https://doi.org/10.1007/s11657-020-0706-y

[25] Rupp, M., Walter, N., Pfeifer, C., Lang, S., Kerschbaum, M., Krutsch, W., Baumann, F., Alt, V.: The Incidence of Fractures Among the Adult Population of Germany. Dtsch Arztebl International **118**(40), 665–669 (2021) https://doi.org/10.3238/arztebl.m2021.0238

[26] Koeppe, J., Stolberg-Stolberg, J., Fischhuber, K., Iking, J., Marschall, U., Raschke, M.J., Katthagen, J.C.: The Incidence of Proximal Humerus Fracture. Dtsch Arztebl International **120**(33-34), 555–556 (2023) https://doi.org/10.3238/arztebl.m2023.0132

[27] Fraser, A., Bjørdal, J., Wagle, T.M., Karlberg, A.C., Lien, O.A., Eilertsen, L., Mader, K., Apold, H., Larsen, L.B., Madsen, J.E., Fjalestad, T.: Reverse Shoulder Arthroplasty Is Superior to Plate Fixation at 2 Years for Displaced Proximal Humeral Fractures in the Elderly: A Multicenter Randomized Controlled Trial. The Journal of Bone and Joint Surgery **102**(2), 477–485 (2020) https://doi.org/10.2106/JBJS.19.01071

[28] Klug, A., Harth, J., Hoffmann, R., Gramlich, Y.: Surgical treatment of complex proximal humeral fractures in elderly patients: a matched-pair analysis of angular-stable plating vs. reverse shoulder arthroplasty. Journal of Shoulder and Elbow Surgery **29**(9), 1796–1803 (2020) https://doi.org/10.1016/j.jse.2020.01.073

[29] Giardella, A., Ascione, F., Mocchi, M., Berlusconi, M., Romano, A., Oliva, F., Maradei, L.: Reverse total shoulder versus angular stable plate treatment for proximal humeral fractures in over 65 years old patients. Muscles, Ligaments and Tendons Journal **7**(2), 271–278 (2017) https://doi.org/10.11138/mltj/2017.7.2.271

[30] Porschke, F., Bockmeyer, J., Nolte, P.-C., Studier-Fischer, S., Guehring, T., Schnetzke, M.: More adverse events after osteosyntheses compared to arthroplasty in geriatric proximal humeral fractures involving anatomical neck. Journal of Clinical Medicine **10**(5) (2021) https://doi.org/10.3390/jcm10050979

[31] Handoll, H., Brealey, S., Rangan, A., Keding, A., Corbacho, B., Jefferson, L., Chuang, L.-H., Goodchild, L., Hewitt, C., Torgerson., D.: The ProFHER (PROximal Fracture of the Humerus: Evaluation by Randomisation) trial – a pragmatic multicentre randomised controlled trial evaluating the clinical effectiveness and cost-effectiveness of surgical compared with non-surgical treatment for proximal fracture of the humerus in adults. Health Technology Assessment **19**(24) (2019) https://doi.org/10.3310/hta19240

[32] Wang, S.V., Verpillat, P., Rassen, J.A., Patrick, A., Garry, E.M., Bartels, D.B.: Transparency and reproducibility of observational cohort studies using large healthcare databases. Clinical Pharmacology & Therapeutics **99**(3), 325–332 (2016) https://doi.org/10.1002/cpt.329

[33] Wang, S.V., Schneeweiss, S., Berger, M.L., Brown, J., Vries, F., Douglas, I., Gagne, J.J., Gini, R., Klungel, O., Mullins, C.D., Nguyen, M.D., Rassen, J.A., Smeeth, L., Sturkenboom, M., Health Care Decision Making: Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies v1.0. Pharmacoepidemiology and Drug Safety **26**(9), 1018–1032 (2017) https://doi.org/10.1002/pds.4295

[34] Wang, S.V., Sreedhara, S.K., Schneeweiss, S., Franklin, J.M., Initiative, R.: Reproducibility of real-world evidence studies using clinical practice data to inform regulatory and coverage decisions. Nature Communications **13**(1), 5126 (2022) https://doi.org/10.1038/s41467-022-32310-3

[35] Hernán, M.A., Robins, J.M.: Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. American Journal of Epidemiology **183**(8), 758–764 (2016) https://doi.org/10.1093/aje/kwv254

[36] Hernán, M.A., Wang, W., Leaf, D.E.: Target Trial Emulation: A Framework for Causal Inference From Observational Data. JAMA **328**(24), 2446–2447 (2022) https://doi.org/10.1001/jama.2022.21383

[37] Franklin, J.M., Patorno, E., Desai, R.J., Glynn, R.J., Martin, D., Quinto, K., Pawar, A., Bessette, L.G., Lee, H., Garry, E.M., Gautam, N., Schneeweiss, S.: Emulating randomized clinical trials with nonrandomized real-world evidence studies. Circulation **143**(10), 1002–1013 (2021) https://doi.org/10.1161/CIRCULATIONAHA.120.051718

[38] Wang, S.V., Schneeweiss, S., Initiative, R.-D.: Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses: Results of 32 Clinical Trials. JAMA **329**(16), 1376–1385 (2023) https://doi.org/10.1001/jama.2023.4221

[39] Heyard, R., Held, L., Schneeweiss, S., Wang, S.V.: Design differences and variation in results between randomised trials and non-randomised emulations: meta-analysis of rct-duplicate data. BMJ Medicine **3**(1), 000709 (2024) https://doi.org/10.1136/bmjmed-2023-000709

[40] Franklin, J.M., Pawar, A., Martin, D., Glynn, R.J., Levenson, M., Temple, R., Schneeweiss, S.: Nonrandomized Real-World Evidence to Support Regulatory Decision Making: Process for a Randomized Trial Replication Project. Clinical Pharmacology & Therapeutics **107**(4), 817–826 (2020) https://doi.org/10.1002/cpt.1633

[41] Errington, T.M., Mathur, M., Soderberg, C.K., Denis, A., Perfito, N., Iorns, E., Nosek, B.A.: Investigating the replicability of preclinical cancer biology. eLife **10**, 71601 (2021) https://doi.org/10.7554/eLife.71601

[42] Held, L.: A New Standard for the Analysis and Design of Replication Studies. Journal of the Royal Statistical Society Series A: Statistics in Society **183**(2), 431–448 (2020) https://doi.org/10.1111/rssa.12493

[43] Held, L., Micheloud, C., Pawel, S.: The assessment of replication success based on relative effect size. The Annals of Applied Statistics **16**(2), 706–720 (2022) https://doi.org/10.1214/21-AOAS1502

[44] Micheloud, C., Balabdaoui, F., Held, L.: Assessing replicability with the sceptical p-value: Type-i error control and sample size planning. Statistica Neerlandica **77**(4), 573–591 (2023) https://doi.org/10.1111/stan.12312

[45] Kuss, O.: The z-difference can be used to measure covariate balance in matched propensity score analyses. Journal of Clinical Epidemiology **66**(11), 1302–1307 (2013) https://doi.org/10.1016/j.

jclinepi.2013.06.001

[46] Kuss, O., Blettner, M., Börgermann, J.: Propensity Score: an Alternative Method of Analyzing Treatment Effects. Dtsch Arztebl International **113**(35-36), 597–603 (2016) https://doi.org/10.3238/arztebl.2016.0597

[47] Held, L.: Beyond the two-trials rule. Statistics in Medicine **43**(26), 5023–5042 (2024) https://doi.org/10.1002/sim.10055

[48] US Food and Drug Administration, e.a.: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/providing-clinical-evidence-effectiveness-human-drug-and-biological-products

[49] Held, L., Matthews, R., Ott, M., Pawel, S.: Reverse-bayes methods for evidence assessment and research synthesis. RESEARCH SYNTHESIS METHODS **13**(3), 295–314 (2022) https://doi.org/10.1002/jrsm.1538

[50] Micheloud, C., Held, L.: The replication of equivalence studies. Biometrical Journal **66**(8), 202300232 (2024) https://doi.org/10.1002/bimj.202300232

[51] Altman, D.G., Bland, J.M.: Statistics notes: Absence of evidence is not evidence of absence. BMJ **311**(7003), 485 (1995) https://doi.org/10.1136/bmj.311.7003.485 https://www.bmj.com/content/311/7003/485.full.pdf

[52] Jin, X., Ding, C., Hunter, D.J., Gallego, B.: Effectiveness of vitamin d supplementation on knee osteoarthritis - a target trial emulation study using data from the osteoarthritis initiative cohort. Osteoarthritis and Cartilage **30**(11), 1495–1505 (2022) https://doi.org/10.1016/j.joca.2022.06.005

[53] Jang, H.Y., Kim, I.-W., Oh, J.M.: Using real-world data for supporting regulatory decision making: Comparison of cardiovascular and safety outcomes of an empagliflozin randomized clinical trial versus real-world data. Frontiers in Pharmacology **13** (2022) https://doi.org/10.3389/fphar.2022.928121

[54] Merola, D., Campbell, U., Gautam, N., Rubens, A., Schneeweiss, S., Wang, S.V., Carrigan, G., Chia, V., Ovbiosa, O.E., Pinheiro, S., Bruno, A., Jiao, X., Stewart, M., Hendricks-Sturrup, R., Rodriguez-Watson, C., Khosla, S., Zhang, Y., Rimawi, M., Huang, J., Taylor, A., Becnel, L., McRoy, L., Eckert, J., Taylor, B.: The aetion coalition to advance real-world evidence through randomized controlled trial emulation initiative: Oncology. Clinical Pharmacology & Therapeutics **113**(6), 1217–1222 (2023) https://doi.org/10.1002/cpt.2800

[55] Antoine, A., Pérol, D., Robain, M., Bachelot, T., Choquet, R., Jacot, W., Ben Hadj Yahia, B., Grinda, T., Delaloge, S., Lasset, C., Drouet, Y.: Assessing the real-world effectiveness of 8 major metastatic breast cancer drugs using target trial emulation. European Journal of Cancer **213**, 115072 (2024) https://doi.org/10.1016/j.ejca.2024.115072

[56] Signori, A., Ponzano, M., Kalincik, T., Ozakbas, S., Horakova, D., Kubala Havrdova, E., Alroughani, R., Patti, F., Kuhle, J., Izquierdo, G., Eichau, S., Yamout, B., Khoury, S.J., Karabudak, R., Grammond, P., Duquette, P., Roos, I., Butzkueven, H., Walt, A., Sormani, M.P.: Emulating randomised clinical trials in relapsing-remitting multiple sclerosis with non-randomised real-world evidence: an application using data from the msbase registry. Journal of Neurology, Neurosurgery & Psychiatry **95**(7), 620–625 (2024) https://doi.org/10.1136/jnnp-2023-332603

[57] D'Andrea, E., Schneeweiss, S., Franklin, J.M., Kim, S.C., Glynn, R.J., Lee, S.B., Wang, S.V.: Efficacy versus effectiveness: The horizon-pivotal fracture trial and its emulation in claims data. Arthritis & Rheumatology **n/a**(n/a) (2024) https://doi.org/10.1002/art.42968

[58] Köppe, J., Micheloud, C., Erdmann, S., Heyard, R., Held, L.: Assessing the replicability of RCTs in RWE emulations (2025). https://arxiv.org/abs/2412.09334

[59] Stolberg-Stolberg, J., Köppe, J., Rischen, R., Freistühler, M., Faldum, A., Katthagen, J.C., Raschke, M.J.: Influence of complications and comorbidities on length of hospital stay and costs for surgical treatment of proximal humeral fractures. Chirurg **92**(10), 907–915 (2021) https://doi.org/10.1007/s00104-021-01491-w

Supplementary information

Appendix A CONSORT flow-chart of the emulation study

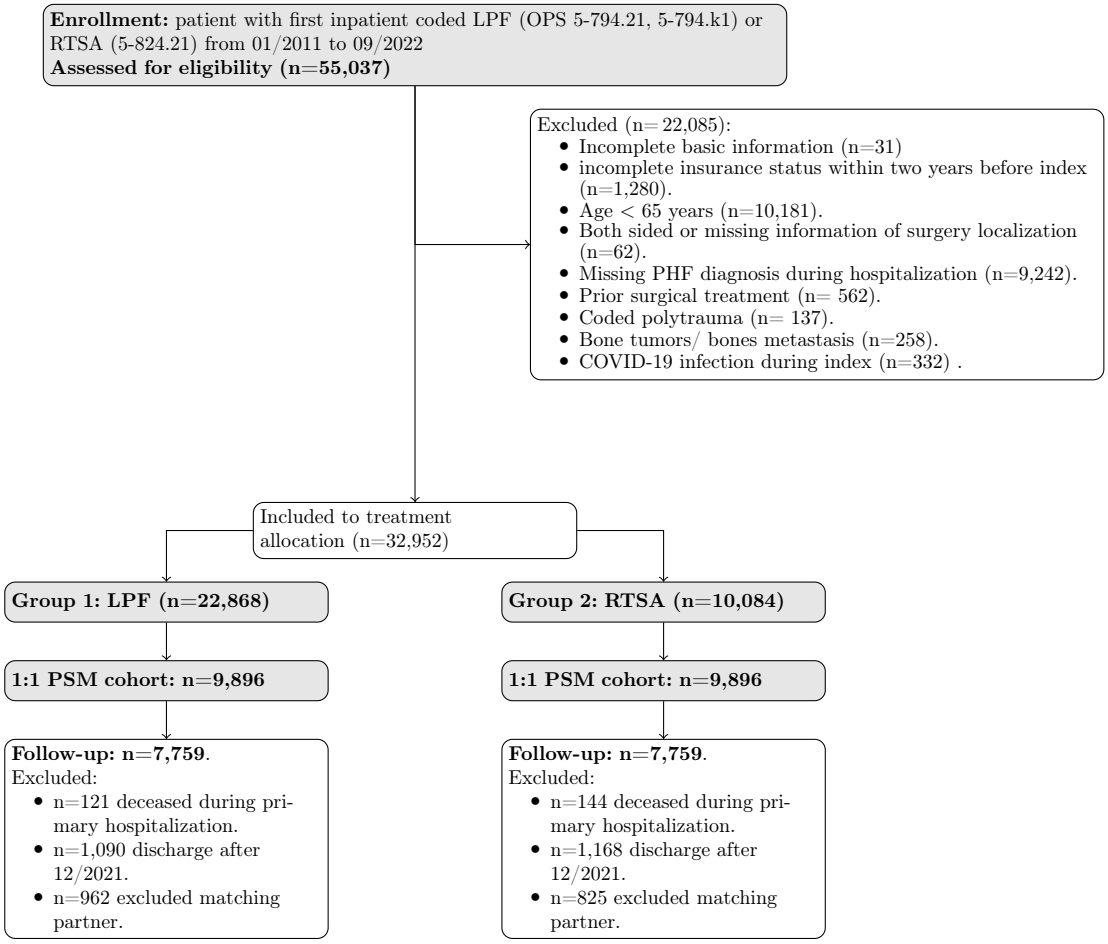


Fig. A1: CONSORT flow chart of the emulation study. LPF – locked plate fixation; PHF – proximal humeral fracture; PSM – propensity score matching; RTSA – revere total shoulder arthroplasty.

Appendix B Patient characteristics replication study before matching

.

Table B1: Patient characteristics of the replication study for PHF compared to the original study cohort presented in [1]

	Emulation study cohort		**Original study cohort: [1]**	
	LPF	RTSA	LPF	RTSA
Frequency – n (%)	**22,868 (69%)**	**10,084 (31%)**	**41,216 (74.8%)**	**13,854 (25.2%)**
Median age – years (IQR)	77 (12)	80 (10)	78 (11)	81 (9)
Age $\geq$ 80 years – % (n)	37% (8,556)	54% (5,475)	43% (17,857)	56% (7,785)
Female sex – % (n)	86% (19,621)	87% (8,809)	84% (34,476)	87% (12,012)
Comorbidities – % (n):				
Osteoporosis	36% (8,163)	42% (4,185)	35% (14,466)	41% (5,639)
Cancer	27% (6,107)	28% (2,835)	22% (8,996)	23% (3,192)
Diabetes Mellitus	29% (6,602)	34% (3,406)	39% (16,191)	43% (5,965)
Dementia	10% (2,262)	11% (1,157)	5% (2,199)	6% (835)
Chronic polyarthritis	6% (1,452)	7% (708)	6% (2,505)	6%(873)
Omarthrosis	3% (654)	4% (394)	2% (873)	3% (392)
Obesity	22% (4,978)	26% (2,590)	27% (11,073)	32% (4,481)
Nicotine abuse	7% (1,632)	7% (700)	6% (2,500)	5% (667)
Parkinson disease	4% (805)	4% (401)	4% (1,668)	4% (568)
Rotator cuff rupture	4% (816)	5% (519)	6% (2,621)	10% (1,369)
Alcohol abuse	5% (1,216)	6% (560)	6% (2,294)	5% (723)
Stroke and/or CVD before PHF	25% (5,631)	27% (2,699)	25% (10,262)	27% (3,692)
Frozen shoulder	4% (990)	4% (436)	4% (1,838)	5% (624)
Atrial fibrillation/ flutter	16% (3,758)	22% (2,197)	18% (7,512)	23% (3,248)
Congestive heart failure	21% (4,832)	27% (2,763)	29% (12,070)	35% (4,885)
Coronary heart disease	23% (5,328)	26% (2,626)	30% (12,471)	33% (4,615)
Hypertension	82% (18,742)	88% (8,834)	87% (35,830)	91% (12,644)
Atherosclerosis	16% (3,579)	18% (1,780)	17% (6,815)	19% (2,624)
Chronic kidney disease	22% (5,095)	30% (3,026)	26% (10,857)	35% (4,824)
Prev. drugs – % (n):				
Any anticoagulant	28% (6,309)	33% (3,368)	27% (11,299)	33% (4,519)
Analgesics[1]	14% (3,205)	16% (1,615)	8% (3,340)	8% (1,109)
Antibiotics[1]	2% (490)	2% (192)	1% (517)	1% (135)
Vitamin D or Calcium	7% (1,537)	8% (830)	4% (1,668)	6% (763)
Bisphosphonates	7% (1,585)	7% (715)	7% (2,837)	7% (1,026)

Definitions of all parameters can be found in [1, 2].

[1]Only within 14 days before admission.

CVD – cerebrovascular disease; IQR – inter-quartile range; LPF – locked plate fixation; PHF – proximal humeral fracture; RTSA – reverse total shoulder arthroplasty.

Appendix C Univariate outcome of the replication study

Table C2: Univariate outcome of the replication study for PHF compared to the original study cohort presented in [1, 2, 59].

	Emulation study cohort		**Original study cohort**	
	LPF	RTSA	LPF	RTSA
	Inhospital outcomes			
Mean LOS – day (± SD)	11.5 (±8.2)	15.2 (±8.8)	14.6 (±11.4)	20.0 (±13.5)
Mean charges – EUR (± SD)	6895.0 (±5139.0)	11,200.8 (±4846.0)	7030.1 (±5532.0)	11,165.7 (±5884.4)
MAEs – % (95%CI)	3.2% (3.0%; 3.4%)	6.2% (5.7%; 6.7%)	4%	7%
TEs – % (95%CI)	0.6% (0.5%; 0.7%)	1.0% (0.8%; 1.2%)	1%	2%
SC – %(95%CI)	6.5% (6.2%; 6.8%)	7.4% (6.9%; 7.9%)	8%	9%
Implant-assoc. compl. – %(95%CI)	1.0% (0.9%; 1.2%)	2.3% (2.0%; 2.6%)	1%	1%
Non-implant assoc. compl. – %(95%CI)	2.0% (1.9%; 2.2%)	2.7% (2.4%; 3.0%)	3%	3%
Death (any cause) – % (95%CI)	0.9% (0.8%; 1.1%)	1.5% (1.3%; 1.7%)	2%	2%
30-day mortality – % (95%CI)	1.5% (1.3%; 1.6%)	2.2% (2.0%; 2.5%)	2%	3%
	Long-term outcomes			
OS after discharge				
1 year	6.4% (6.1%; 6.7%)	8.7% (8.1%; 9.3%)	8.4% (8.1%; 8.6%)	9.1% (8.7%; 9.7%)
5 years	28.0% (27.3%; 28.7%)	36.8% (35.5%; 38.0%)	34.5% (34.0%; 35.1%)	39.4% (38.2%; 40.7%)
MAEs after discharge				
1 year	10.4% (10.0%; 10.8%)	13.6% (12.9%; 14.4%)	12.4% (12.1%; 12.8%)	14.2% (13.6%; 14.8%)
5 years	37.6% (36.8%; 38.3%)	47.5% (46.2%; 48.8%)	44.7% (41.4%; 45.3%)	50.0% (48.7%; 51.2%)
TEs after discharge				
1 year	8.7% (8.3%; 9.1%)	11.6% (10.9%; 12.3%)	11.0% (10.7%; 11.3%)	12.2% (11.6%; 12.8%)
5 years	33.5% (32.8%; 34.2%)	42.5% (41.2%; 43.8%)	40.9% (40.3%; 41.5%)	45.9% (44.7%; 47.2%)
SC after discharge				
1 year	11.6% (11.2%; 12.1%)	4.0% (3.6%; 4.4%)	9.5% (9.3%; 9.8%)	4.0% (3.6%; 4.3%)
5 years	15.8% (15.3%; 16.3%)	8.0% (7.3%; 8.6%)	12.0% (11.7%; 12.3%)	6.2% (5.7%; 6.7%)
MOC after discharge				
1 year	24.1% (23.5%; 24.7%)	23.3% (22.4%; 24.2%)	20.8% (20.4%; 21.2%)	21.9% (21.2%; 22.6%)
5 years	34.2% (33.5%; 34.9%)	33.6% (32.6%; 34.7%)	30.6% (30.1%; 31.1%)	31.7% (30.8%; 32.6%)

Definitions of all parameters can be found in [1, 2]. For binary endpoints (short-term endpoints) of the replication cohort, ratios with 95% Clopper-Pearson confidence interval (CI) are presented. For time-to-event endpoints, event rates were given by the cumulative incidence function determined using Aalen-Johansen estimate with 95%-CI.

LOS – length of hospital stay, LPF – locked plate fixation, MAE – major adverse event, MOC – minor outpatient complications, TE– thrombo-embolic events, SC – surgical complications, SD– standard deviation, RTSA – reverse total shoulder arthroplasty.

Appendix D Full regression results for all presented models

Table D3: Results from logistic regression model – major adverse events.

Variable	OR	95% CI	p-value (two-sided)
Treatment group (RTSA vs LPF)	1.75	1.40 – 2.19	<0.001
Age at index surgery (years)	0.95	0.67 – 1.35	0.759
Male sex	2.08	0.70 – 6.18	0.186
Year of index surgery (years)	0.72	0.23 – 2.30	0.584
Any anticoagulant	1.04	0.78 – 1.38	0.790
Any anti-osteoporotic drug	0.98	0.58 – 1.66	0.934
Cancer	1.15	0.85 – 1.56	0.356
Diabetes mellitus	0.78	0.27 – 2.21	0.637
Dementia	1.92	1.03 – 3.59	0.040
Chronic polyarthritis	1.22	0.64 – 2.30	0.550
Obesity	0.45	0.08 – 2.73	0.388
Nicotine abuse	1.51	0.91 – 2.50	0.113
Parkinson disease	1.02	0.45 – 2.32	0.964
Rotator cuff rupture	0.31	0.02 – 5.38	0.417
Alcohol abuse	0.91	0.13 – 6.60	0.927
Atrial fibrillation/ flutter	1.39	0.78 – 2.47	0.260
Hypertension	1.02	0.32 – 3.26	0.970
Congestive heart failure	2.11	1.12 – 3.97	0.021
Coronary heart disease	1.61	1.08 – 2.39	0.020
Atherosclerosis	1.08	0.76 – 1.53	0.682
Chronic kidney disease	1.86	1.23 – 2.83	0.004
Stroke and/or CVD before PHF	1.13	0.71 – 1.80	0.608
Omarthrosis	0.31	0.08 – 1.29	0.108
Frozen shoulder	0.74	0.33 – 1.62	0.446
Analgesics*	0.62	0.45 – 0.87	0.005
Antibiotics*	1.68	0.55 – 5.10	0.363

*only within 14 days before admission. CVD – cerebrovascular disease, LPF – locked plate fixation, OR – odds ratio, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty.

Table D4: Results from logistic regression model – thrombo-embolic events.

Variable	OR	95% CI	p-value (two-sided)
Treatment group (RTSA vs LPF)	1.77	1.22 – 2.58	0.003
Male sex	0.97	0.42 – 2.22	0.939
Age at index surgery (years)	0.91	0.85 – 0.97	0.005
Year of index surgery (years)	0.71	0.66 – 0.77	<0.001
Any anticoagulant	0.88	0.46 – 1.68	0.697
Any anti-osteoporotic drug	1.16	0.49 – 2.71	0.736
Cancer	1.29	0.73 – 2.29	0.386
Diabetes mellitus	0.71	0.40 – 1.27	0.245
Dementia	2.56	1.16 – 5.66	0.020
Chronic polyarthritis	1.90	0.59 – 6.14	0.284
Obesity	0.50	0.24 – 1.00	0.051
Nicotine abuse	0.88	0.31 – 2.45	0.800
Parkinson disease	0.82	0.14 – 4.91	0.824
Rotator cuff rupture	0.27	0.08 – 0.86	0.027
Alcohol abuse	0.17	0.04 – 0.86	0.032
Atrial fibrillation/ flutter	1.52	0.72 – 3.20	0.271
Hypertension	0.50	0.16 – 1.51	0.217
Congestive heart failure	1.21	0.64 – 2.31	0.563
Coronary heart disease	1.97	1.01 – 3.84	0.046
Atherosclerosis	1.18	0.55 – 2.50	0.676
Chronic kidney disease	0.70	0.37 – 1.32	0.268
Stroke and/or CVD before PHF	1.59	0.90 – 2.81	0.107
Omarthrosis	0.35	0.08 – 1.47	0.153
Frozen shoulder	0.74	0.20 – 2.67	0.644
Analgesics*	1.17	0.56 – 2.44	0.684
Antibiotics*	0.57	0.10 – 3.46	0.544

*only within 14 days before admission. CVD – cerebrovascular disease, LPF – locked plate fixation, OR – odds ratio, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty.

Table D5: Results from logistic regression model – surgical complications.

Variable	OR	95% CI	p-value (two-sided)
Treatment group (RTSA vs LPF)	1.14	1.02 – 1.28	0.021
Male sex	2.67	2.07 – 3.45	<0.001
Age at index surgery (years)	0.91	0.89 – 0.92	<0.001
Year of index surgery (years)	0.71	0.70 – 0.73	<0.001
Any anticoagulant	1.27	1.03 – 1.57	0.023
Any anti-osteoporotic drug	1.23	0.96 – 1.57	0.104
Cancer	1.18	0.98 – 1.42	0.083
Diabetes mellitus	0.74	0.62 – 0.89	0.001
Dementia	1.56	1.21 – 2.01	<0.001
Chronic polyarthritis	0.98	0.71 – 1.35	0.909
Obesity	0.61	0.50 – 0.76	<0.001
Nicotine abuse	0.98	0.71 – 1.34	0.889
Parkinson disease	1.17	0.80 – 1.72	0.420
Rotator cuff rupture	0.70	0.51 – 0.96	0.025
Alcohol abuse	0.78	0.54 – 1.11	0.171
Atrial fibrillation/ flutter	0.91	0.72 – 1.14	0.393
Hypertension	0.73	0.55 – 0.96	0.025
Congestive heart failure	0.93	0.76 – 1.15	0.513
Coronary heart disease	0.91	0.75 – 1.12	0.387
Atherosclerosis	1.03	0.84 – 1.28	0.760
Chronic kidney disease	1.12	0.92 – 1.35	0.259
Stroke and/or CVD before PHF	1.14	0.94 – 1.37	0.194
Omarthrosis	0.88	0.59 – 1.31	0.517
Frozen shoulder	0.93	0.63 – 1.38	0.719
Analgesics*	0.95	0.76 – 1.18	0.615
Antibiotics*	1.21	0.63 – 2.29	0.568

*only within 14 days before admission. CVD – cerebrovascular disease, LPF – locked plate fixation, OR – odds ratio, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty.

Table D6: Results from logistic regression model – implant-associated complications.

Variable	OR	95% CI	p-value (two-sided)
Treatment group (RTSA vs LPF)	1.78	1.38 – 2.29	<0.001
Male sex	3.97	2.26 – 6.96	<0.001
Age at index surgery (years)	0.85	0.82 – 0.89	<0.001
Year of index surgery (years)	0.71	0.69 – 0.74	<0.001
Any anticoagulant	0.97	0.61 – 1.54	0.895
Any anti-osteoporotic drug	1.17	0.66 – 2.10	0.590
Cancer	1.26	0.83 – 1.92	0.286
Diabetes mellitus	0.57	0.38 – 0.85	0.005
Dementia	2.13	1.24 – 3.67	0.007
Chronic polyarthritis	1.20	0.60 – 2.38	0.612
Obesity	0.66	0.42 – 1.04	0.075
Nicotine abuse	0.88	0.46 – 1.66	0.684
Parkinson disease	1.60	0.69 – 3.72	0.279
Rotator cuff rupture	1.02	0.51 – 2.07	0.947
Alcohol abuse	0.75	0.33 – 1.71	0.489
Atrial fibrillation/ flutter	0.86	0.52 – 1.45	0.577
Hypertension	0.41	0.21 – 0.81	0.010
Congestive heart failure	1.01	0.65 – 1.58	0.958
Coronary heart disease	0.88	0.57 – 1.34	0.545
Atherosclerosis	1.32	0.84 – 2.08	0.230
Chronic kidney disease	1.42	0.94 – 2.14	0.098
Stroke and/or CVD before PHF	1.30	0.86 – 1.97	0.220
Omarthrosis	0.93	0.43 – 2.02	0.863
Frozen shoulder	0.45	0.20 – 1.00	0.049
Analgesics*	1.27	0.78 – 2.06	0.341
Antibiotics*	1.13	0.31 – 4.09	0.854

*only within 14 days before admission. CVD – cerebrovascular disease, LPF – locked plate fixation, OR – odds ratio, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty.

Table D7: Results from logistic regression model – non-implant associated complications.

Variable	OR	95% CI	p-value (two-sided)
Treatment group (RTSA vs LPF)	1.42	1.16 – 1.74	<0.001
Male sex	2.53	1.57 – 4.07	<0.001
Age at index surgery (years)	0.94	0.91 – 0.97	<0.001
Year of index surgery (years)	0.71	0.69 – 0.73	<0.001
Any anticoagulant	1.37	0.95 – 1.97	0.093
Any anti-osteoporotic drug	1.45	0.90 – 2.35	0.128
Cancer	1.27	0.90 – 1.79	0.171
Diabetes mellitus	0.66	0.48 – 0.92	0.015
Dementia	1.69	1.07 – 2.65	0.023
Chronic polyarthritis	0.95	0.51 – 1.78	0.869
Obesity	0.39	0.26 – 0.59	<0.001
Nicotine abuse	1.85	0.99 – 3.45	0.053
Parkinson disease	0.88	0.42 – 1.82	0.723
Rotator cuff rupture	0.58	0.31 – 1.08	0.086
Alcohol abuse	0.91	0.49 – 1.68	0.752
Atrial fibrillation/ flutter	1.32	0.89 – 1.96	0.171
Hypertension	0.80	0.48 – 1.35	0.406
Congestive heart failure	0.77	0.53 – 1.12	0.169
Coronary heart disease	0.94	0.65 – 1.36	0.734
Atherosclerosis	1.30	0.89 – 1.91	0.176
Chronic kidney disease	1.11	0.79 – 1.57	0.551
Stroke and/or CVD before PHF	1.18	0.83 – 1.67	0.369
Omarthrosis	1.52	0.73 – 3.19	0.265
Frozen shoulder	2.15	0.98 – 4.70	0.055
Analgesics*	0.98	0.65 – 1.47	0.912
Antibiotics*	1.96	0.44 – 8.73	0.379

*only within 14 days before admission. CVD – cerebrovascular disease, LPF – locked plate fixation, OR – odds ratio, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty.

Table D8: Results from logistic regression model – 30-day mortality.

Variable	OR	95% CI	p-value (two-sided)
Treatment group (RTSA vs LPF)	1.20	0.81 – 1.78	0.368
Male sex	1.18	0.19 – 7.34	0.859
Age at index surgery (years)	1.46	0.79 – 2.69	0.224
Year of index surgery (years)	2.35	0.32 – 17.24	0.400
Any anticoagulant	1.24	0.72 – 2.13	0.436
Any anti-osteoporotic drug	0.94	0.37 – 2.37	0.898
Cancer	0.91	0.54 – 1.55	0.735
Diabetes mellitus	3.02	0.52 – 17.61	0.220
Dementia	1.09	0.37 – 3.18	0.880
Chronic polyarthritis	1.42	0.42 – 4.78	0.568
Obesity	3.10	0.14 – 68.80	0.474
Nicotine abuse	2.51	1.02 – 6.17	0.046
Parkinson disease	1.77	0.43 – 7.33	0.431
Rotator cuff rupture	3.93	0.03 – 561.0	0.589
Alcohol abuse	5.18	0.17 – 160.4	0.348
Atrial fibrillation/ flutter	3.29	1.28 – 8.47	0.014
Hypertension	1.66	0.23 – 11.79	0.613
Congestive heart failure	4.33	1.41 – 13.25	0.010
Coronary heart disease	1.09	0.54 – 2.23	0.808
Atherosclerosis	0.87	0.47 – 1.62	0.656
Chronic kidney disease	1.14	0.54 – 2.38	0.732
Stroke and/or CVD before PHF	0.86	0.39 – 1.89	0.712
Omarthrosis	1.67	0.14 – 19.88	0.685
Frozen shoulder	0.82	0.23 – 2.95	0.755
Analgesics*	0.93	0.52 – 1.65	0.802
Antibiotics*	0.46	0.07 – 2.89	0.406

*only within 14 days before admission. CVD – cerebrovascular disease, LPF – locked plate fixation, OR – odds ratio, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty.

Table D9: Results from Cox regression model – overall survival.

Variable	**HR**	**95% CI**	**p-value (two-sided)**
Treatment group (RTSA vs LPF)	1.07	0.98 – 1.17	0.118
LOS during index (days)	1.02	1.01 – 1.03	<0.001
Male sex	1.56	1.00 – 2.45	0.052
Age at index surgery (years)	1.13	0.98 – 1.30	0.083
Year of index surgery (years)	1.10	0.70 – 1.73	0.679
Any anticoagulant	1.13	0.98 – 1.31	0.095
Any anti-osteoporotic drug	0.91	0.71 – 1.15	0.411
Osteoporosis	n.a.	n.a.	n.a.
Cancer	1.07	0.92 – 1.24	0.373
Diabetes Mellitus	1.51	0.99 – 2.28	0.053
Dementia	1.85	1.40 – 2.46	<0.001
Chronic polyarthritis	1.17	0.89 – 1.53	0.253
Obesity	0.96	0.47 – 1.96	0.905
Nicotine abuse	1.80	1.39 – 2.35	<0.001
Parkinson disease	1.52	1.05 – 2.19	0.026
Rotator cuff rupture	1.39	0.45 – 4.33	0.573
Alcohol abuse	2.18	1.00 – 4.77	0.051
Atrial fibrillation/ flutter	1.17	0.91 – 1.51	0.215
Hypertension	1.03	0.65 – 1.65	0.892
Congestive heart failure	1.45	1.11 – 1.89	0.007
Coronary heart disease	0.91	0.76 – 1.09	0.313
Atherosclerosis	0.94	0.79 – 1.11	0.444
Chronic kidney disease	1.40	1.15 – 1.70	<0.001
Stroke and/or CVD before PHF	1.00	0.82 – 1.23	0.981
Omarthrosis	1.18	0.64 – 2.17	0.593
Frozen shoulder	0.83	0.60 – 1.15	0.268
General compl. during IH	1.67	1.39 – 2.02	<0.001
Surgical compl. during IH	1.05	0.85 – 1.31	0.652
Start of COVID-19 pandemic	0.96	0.78 – 1.17	0.678
COVID-19 infection	2.23	1.76 – 2.84	<0.001

Start of COVID-19 pandemic and COVID-19 infection during follow-up were considered using time-dependent variables. CVD – cerebrovascular disease, HR – hazard ratio, IH – index hospitalization, LOS – length of hospital stay, LPF – locked plate fixation, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty.

Table D10: Results from Cox regression model – major adverse events.

Variable	**HR**	**95% CI**	**p-value (two-sided)**
Treatment group (RTSA vs LPF)	1.04	0.97 – 1.12	0.270
LOS during index (days)	1.02	1.01 – 1.03	<0.001
Male sex	1.58	1.06 – 2.36	0.024
Age at index surgery (years)	1.06	0.93 – 1.20	0.383
Year of index surgery (years)	0.97	0.64 – 1.45	0.865
Any anticoagulant	1.01	0.89 – 1.15	0.856
Any anti-osteoporotic drug	0.94	0.76 – 1.16	0.552
Osteoporosis	n.a.	n.a.	n.a.
Cancer	1.06	0.94 – 1.21	0.331
Diabetes Mellitus	1.24	0.86 – 1.79	0.254
Dementia	1.71	1.33 – 2.20	<0.001
Chronic polyarthritis	1.25	0.99 – 1.57	0.061
Obesity	0.89	0.47 – 1.75	0.719
Nicotine abuse	1.41	1.13 – 2.35	0.002
Parkinson disease	1.51	1.09 – 2.08	0.014
Rotator cuff rupture	1.04	0.38 – 2.84	0.946
Alcohol abuse	1.72	0.86 – 3.45	0.127
Atrial fibrillation/ flutter	1.31	1.05 – 1.64	0.016
Hypertension	1.00	0.66 – 1.52	0.987
Congestive heart failure	1.23	0.97 – 1.56	0.087
Coronary heart disease	1.06	0.90 – 1.23	0.499
Atherosclerosis	0.96	0.83 – 1.11	0.571
Chronic kidney disease	1.23	1.17 – 1.64	<0.001
Stroke and/or CVD before PHF	1.11	0.93 – 1.33	0.238
Omarthrosis	0.94	0.55 – 2.17	0.821
Frozen shoulder	0.84	0.60 – 1.60	0.209
General compl. during IH	1.42	1.21 – 1.67	<0.001
Surgical compl. during IH	1.17	0.97 – 1.41	0.111
Start of COVID-19 pandemic	0.87	0.74 – 1.03	0.106
COVID-19 infection	2.16	1.72 – 2.73	<0.001

Start of COVID-19 pandemic and COVID-19 infection during follow-up were considered using time-dependent variables. CVD – cerebrovascular disease, HR – hazard ratio, IH – index hospitalization, LOS – length of hospital stay, LPF – locked plate fixation, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty.

Table D11: Results from Cox regression model – thrombo-embolic events.

Variable	**HR**	**95% CI**	**p-value (two-sided)**
Treatment group (RTSA vs LPF)	1.09	1.00 – 1.17	0.039
LOS during index (days)	1.02	1.01 – 1.03	<0.001
Male sex	1.82	1.20 – 2.76	0.005
Age at index surgery (years)	1.05	0.93 – 1.20	0.424
Year of index surgery (years)	0.91	0.60 – 1.39	0.675
Any anticoagulant	1.00	0.87 – 1.14	0.975
Any anti-osteoporotic drug	0.95	0.76 – 1.18	0.643
Osteoporosis	n.a.	n.a.	n.a.
Cancer	1.12	0.98 – 1.28	0.095
Diabetes Mellitus	1.15	0.78 – 1.68	0.477
Dementia	1.82	1.40 – 2.37	<0.001
Chronic polyarthritis	1.06	0.83 – 1.36	0.625
Obesity	0.78	0.40 – 1.51	0.457
Nicotine abuse	1.48	1.17 – 1.87	0.001
Parkinson disease	1.45	1.03 – 2.04	0.035
Rotator cuff rupture	0.89	0.31 – 2.56	0.833
Alcohol abuse	1.50	0.72 – 3.09	0.277
Atrial fibrillation/ flutter	1.09	0.87 – 1.38	0.459
Hypertension	0.91	0.59 – 1.41	0.681
Congestive heart failure	1.22	0.95 – 1.56	0.120
Coronary heart disease	1.01	0.86 – 1.20	0.867
Atherosclerosis	0.97	0.83 – 1.13	0.655
Chronic kidney disease	1.31	1.10 – 1.57	0.002
Stroke and/or CVD before PHF	1.13	0.94 – 1.37	0.192
Omarthrosis	0.89	0.51 – 1.55	0.689
Frozen shoulder	0.87	0.64 – 1.17	0.343
General compl. during IH	1.60	1.34 – 1.90	<0.001
Surgical compl. during IH	1.16	0.95 – 1.41	0.143
Start of COVID-19 pandemic	0.92	0.77 – 1.10	0.400
COVID-19 infection	2.00	1.59 – 2.53	<0.001

Start of COVID-19 pandemic and COVID-19 infection during follow-up were considered using time-dependent variables. CVD – cerebrovascular disease, HR – hazard ratio, IH – index hospitalization, LOS – length of hospital stay, LPF – locked plate fixation, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty.

Table D12: Results from Cox regression model – surgical complications.

Variable	HR	95% CI	p-value (two-sided)
Treatment group (RTSA vs LPF)	0.42	0.38 – 0.46	0<0.001
LOS during index (days)	1.01	1.01 – 1.02	0.001
Male sex	1.82	1.04 – 3.18	0.037
Age at index surgery (years)	0.95	0.80 – 1.13	0.570
Year of index surgery (years)	0.91	0.52 – 1.60	0.736
Any anticoagulant	1.08	0.92 – 1.27	0.333
Any anti-osteoporotic drug	0.98	0.73 – 1.30	0.875
Osteoporosis	n.a.	n.a.	n.a.
Cancer	1.05	0.90 – 1.24	0.519
Diabetes Mellitus	1.01	0.62 – 1.67	0.961
Dementia	1.21	0.86 – 1.70	0.276
Chronic polyarthritis	1.23	0.91 – 1.66	0.185
Obesity	1.14	0.47 – 2.78	0.776
Nicotine abuse	0.90	0.69 – 1.16	0.397
Parkinson disease	1.22	0.78 – 1.91	0.386
Rotator cuff rupture	0.90	0.22 – 3.67	0.885
Alcohol abuse	0.95	0.36 – 2.54	0.925
Atrial fibrillation/ flutter	0.97	0.71 – 1.32	0.837
Hypertension	1.14	0.64 – 2.02	0.654
Congestive heart failure	1.07	0.78 – 1.45	0.691
Coronary heart disease	0.99	0.81 – 1.22	0.937
Atherosclerosis	1.17	0.97 – 1.42	0.109
Chronic kidney disease	1.15	0.92 – 1.44	0.224
Stroke and/or CVD before PHF	1.04	0.82 – 1.32	0.762
Omarthrosis	0.86	0.43 – 1.70	0.662
Frozen shoulder	1.50	1.05 – 2.15	0.026
General compl. during IH	0.86	0.70 – 1.06	0.156
Surgical compl. during IH	1.22	0.98 – 1.51	0.070
Start of COVID-19 pandemic	0.80	0.65 – 0.98	0.031
COVID-19 infection	1.37	0.85 – 2.22	0.194

Start of COVID-19 pandemic and COVID-19 infection during follow-up were considered using time-dependent variables. CVD – cerebrovascular disease, HR – hazard ratio, IH – index hospitalization, LOS – length of hospital stay, LPF – locked plate fixation, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty.

Table D13: Results from Cox regression model – minor outpatient complications.

Variable	HR	95% CI	p-value (two-sided)
Treatment group (RTSA vs LPF)	1.00	0.95 – 1.05	0.984
LOS during index (days)	1.00	0.99 – 1.00	0.310
Male Sex	1.07	0.94 – 1.21	0.327
Age at index surgery (years)	0.98	0.96 – 1.01	0.166
Year of index surgery (years)	1.02	0.93 – 1.11	0.720
Any anticoagulant	0.91	0.83 – 1.00	0.052
Any anti-osteoporotic drug	1.08	0.96 – 1.20	0.200
Osteoporosis	n.a.	n.a.	n.a.
Cancer	1.06	0.98 – 1.15	0.162
Diabetes Mellitus	0.89	0.80 – 0.99	0.031
Dementia	0.69	0.61 – 0.79	<0.001
Dementia	1.18	1.03 – 1.35	0.014
Obesity	1.23	1.05 – 1.43	0.010
Nicotine abuse	0.95	0.83 – 1.09	0.467
Rotator cuff rupture	1.87	1.43 – 2.43	<0.001
Atrial fibrillation/ flutter	1.05	0.94 – 1.17	0.435
Hypertension	1.18	1.02 – 1.36	0.028
Congestive heart failure	1.09	0.98 – 1.20	0.101
Coronary heart disease	1.03	0.94 – 1.13	0.518
Atherosclerosis	0.96	0.88 – 1.06	0.447
Chronic kidney disease	1.02	0.94 – 1.12	0.628
Stroke and/or CVD before PHF	1.05	0.96 – 1.14	0.271
Omarthrosis	1.88	1.54 – 2.30	<0.001
Frozen shoulder	4.15	3.43 – 5.03	<0.001
General compl. during IH	0.78	0.69 – 0.89	<0.001
Surgical compl. during IH	1.14	1.00 – 1.30	0.046
Start of COVID-19 pandemic	1.02	0.91 – 1.14	0.702
COVID-19 infection	1.08	0.83 – 1.40	0.557

Start of COVID-19 pandemic and COVID-19 infection during follow-up were considered using time-dependent variables. CVD – cerebrovascular disease, HR – hazard ratio, IH – index hospitalization, LOS – length of hospital stay, LPF – locked plate fixation, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty.

Appendix E Sensitivity analysis – Different approaches for COVID-19.

Table E14: Sensitivity analysis considering COVID-19 – overall survival.

Variable	HR	95% CI	p-value (two-sided)
Approach 1			
Treatment group (RTSA vs LPF)	1.04	0.98 – 1.10	0.235
LOS during index (days)	1.01	1.01 – 1.02	<0.001
Male sex	1.56	1.43 – 1.71	<0.001
Age at index surgery (years)	1.09	1.08 – 1.09	<0.001
Year of index surgery (years)	0.89	0.89 – 0.90	<0.001
Any anticoagulant	1.10	1.02 – 1.18	0.011
Any anti-osteoporotic drug	1.08	0.99 – 1.19	0.081
Osteoporosis	1.00	0.94 – 1.07	0.924
Cancer	1.06	0.99 – 1.13	0.087
Diabetes Mellitus	1.18	1.11 – 1.26	<0.001
Dementia	1.69	1.56 – 1.84	<0.001
Chronic polyarthritis	1.10	0.98 – 1.24	0.103
Obesity	0.87	0.81 – 0.94	<0.001
Nicotine abuse	1.36	1.18 – 1.56	<0.001
Parkinson disease	1.37	1.20 – 1.57	<0.001
Rotator cuff rupture	0.90	0.79 – 1.03	0.119
Alcohol abuse	1.51	1.31 – 1.74	<0.001
Atrial fibrillation/ flutter	1.19	1.10 – 1.29	<0.001
Hypertension	0.95	0.86 – 1.06	0.365
Congestive heart failure	1.25	1.17 – 1.34	<0.001
Coronary heart disease	0.98	0.92 – 1.05	0.636
Atherosclerosis	0.99	0.91 – 1.07	0.705
Chronic kidney disease	1.22	1.14 – 1.30	<0.001
Stroke and/or CVD before PHF	1.05	0.98 – 1.12	0.163
Omarthrosis	0.98	0.82 – 1.18	0.867
Frozen shoulder	0.90	0.78 – 1.04	0.150
General compl. during IH	1.50	1.37 – 1.64	<0.001
Surgical compl. during IH	1.05	0.94 – 1.18	0.407
Approach 2			
Treatment group (RTSA vs LPF)	1.00	0.92 – 1.08	0.968
LOS during index (days)	1.01	1.01 – 1.02	<0.001
Male sex	1.50	1.34 – 1.68	<0.001
Age at index surgery (years)	1.09	1.09 – 1.10	<0.001
Year of index surgery (years)	0.98	0.96 – 1.01	0.134
Any anticoagulant	1.12	1.01 – 1.23	0.025
Any anti-osteoporotic drug	1.12	0.99 – 1.27	0.066
Osteoporosis	1.07	0.98 – 1.16	0.161
Cancer	1.04	0.95 – 1.13	0.410
Diabetes Mellitus	1.19	1.09 – 1.30	<0.001
Dementia	1.91	1.73 – 2.11	<0.001
Chronic polyarthritis	1.05	0.89 – 1.23	0.554
Obesity	0.80	0.72 – 0.89	<0.001
Nicotine abuse	1.49	1.25 – 1.79	<0.001
Parkinson disease	1.48	1.26 – 1.73	<0.001
Rotator cuff rupture	0.93	0.78 – 1.11	0.401
Alcohol abuse	1.50	1.27 – 1.78	<0.001
Atrial fibrillation/ flutter	1.20	1.08 – 1.32	<0.001
Hypertension	1.01	0.87 – 1.18	0.858
Congestive heart failure	1.31	1.20 – 1.44	<0.001
Coronary heart disease	0.96	0.88 – 1.05	0.400
Atherosclerosis	1.03	0.93 – 1.14	0.578
Chronic kidney disease	1.27	1.16 – 1.39	<0.001
Stroke and/or CVD before PHF	0.97	0.89 – 1.07	0.566
Omarthrosis	1.06	0.84 – 1.33	0.619
Frozen shoulder	0.95	0.78 – 1.15	0.572
General compl. during IH	1.53	1.37 – 1.72	<0.001
Surgical compl. during IH	1.16	1.01 – 1.33	0.033

Approach 1: COVID-19 infection was considered as a competing risk event.

Approach 2: Subset analysis including only patients with index surgery before 2020 ($N = 10,879$) and with end of follow-up defined as 31.12.2019.

CVD – cerebrovascular disease, HR – hazard ratio, IH – index hospitalization, LOS – length of hospital stay, LPF – locked plate fixation, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty. Clustering of data (i.e. dependency of matching pairs) was not considered.

Table E15: Sensitivity analysis considering COVID-19 – major adverse events

Variable	HR	95% CI	p-value (two-sided)
Approach 1			
Treatment group (RTSA vs LPF)	1.04	0.99 – 1.10	0.138
LOS during index (days)	1.01	1.01 – 1.02	<0.001
Male sex	1.53	1.42 – 1.65	<0.001
Age at index surgery (years)	1.06	1.05 – 1.06	<0.001
Year of index surgery (years)	0.91	0.90 – 0.92	<0.001
Any anticoagulant	1.07	1.00 – 1.14	0.045
Any anti-osteoporotic drug	1.03	0.95 – 1.12	0.477
Osteoporosis	1.00	0.95 – 1.06	0.978
Cancer	1.05	0.99 – 1.11	0.102
Diabetes Mellitus	1.22	1.16 – 1.29	<0.001
Dementia	1.51	1.40 – 1.63	<0.001
Chronic polyarthritis	1.09	0.99 – 1.21	0.090
Obesity	0.93	0.87 – 0.99	0.020
Nicotine abuse	1.29	1.14 – 1.45	<0.001
Parkinson disease	1.32	1.16 – 1.49	<0.001
Rotator cuff rupture	0.97	0.87 – 1.08	0.578
Alcohol abuse	1.54	1.37 – 1.721	<0.001
Atrial fibrillation/ flutter	1.21	1.13 – 1.30	<0.001
Hypertension	1.06	0.97 – 1.16	0.229
Congestive heart failure	1.21	1.14 – 1.28	<0.001
Coronary heart disease	1.06	0.99 – 1.12	0.082
Atherosclerosis	1.00	0.94 – 1.07	0.910
Chronic kidney disease	1.29	1.22 – 1.37	<0.001
Stroke and/or CVD before PHF	1.09	1.03 – 1.16	0.004
Omarthrosis	0.99	0.86 – 1.15	0.92
Frozen shoulder	0.90	0.80 – 1.02	0.096
General compl. during IH	1.46	1.34 – 1.58	<0.001
Surgical compl. during IH	1.01	0.92 – 1.12	0.820
Approach 2			
Treatment group (RTSA vs LPF)	1.04	0.97 – 1.11	0.323
LOS during index (days)	1.01	1.01 – 1.02	<0.001
Male sex	1.48	1.34 – 1.63	<0.001
Age at index surgery (years)	1.06	1.06 – 1.07	<0.001
Year of index surgery (years)	1.00	0.98 – 1.01	0.623
Any anticoagulant	1.10	1.01 – 1.19	0.031
Any anti-osteoporotic drug	1.08	0.97 – 1.20	0.143
Osteoporosis	1.03	0.95 – 1.11	0.488
Cancer	1.02	0.95 – 1.10	0.605
Diabetes Mellitus	1.19	1.11 – 1.28	<0.001
Dementia	1.66	1.52 – 1.82	<0.001
Chronic polyarthritis	1.00	0.87 – 1.15	0.972
Obesity	0.86	0.79 – 0.94	<0.001
Nicotine abuse	1.42	1.23 – 1.65	<0.001
Parkinson disease	1.37	1.19 – 1.58	<0.001
Rotator cuff rupture	1.04	0.89 – 1.20	0.629
Alcohol abuse	1.52	1.31 – 1.75	<0.001
Atrial fibrillation/ flutter	1.24	1.14 – 1.36	<0.001
Hypertension	1.11	0.98 – 1.26	0.111
Congestive heart failure	1.26	1.17 – 1.36	<0.001
Coronary heart disease	1.03	0.95 – 1.12	0.469
Atherosclerosis	1.07	0.98 – 1.17	0.119
Chronic kidney disease	1.36	1.26 – 1.47	<0.001
Stroke and/or CVD before PHF	1.04	0.96 – 1.12	0.312
Omarthrosis	1.07	0.88 – 1.30	0.521
Frozen shoulder	0.91	0.78 – 1.08	0.287
General compl. during IH	1.46	1.32 – 1.61	<0.001
Surgical compl. during IH	1.13	1.00 – 1.27	0.048

Approach 1: COVID-19 infection was considered as a competing risk event.

Approach 2: Sensitivity analysis; Subset analysis including only patients with index surgery before 2020 ($N = 10,879$) and with end of follow-up defined as 31.12.2019.

CVD – cerebrovascular disease, HR – hazard ratio, IH – index hospitalization, LOS – length of hospital stay, LPF – locked plate fixation, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty. Clustering of data (i.e. dependency of matching pairs) was not considered.

Table E16: Sensitivity analysis considering COVID-19 – thrombo-embolic events

Variable	HR	95% CI	p-value (two-sided)
Approach 1			
Treatment group (RTSA vs LPF)	1.05	0.99 – 1.11	0.102
LOS during index (days)	1.01	1.01 – 1.02	<0.001
Male sex	1.51	1.39 – 1.64	<0.001
Age at index surgery (years)	1.07	1.07 – 1.08	<0.001
Year of index surgery (years)	0.90	0.89 – 0.91	<0.001
Any anticoagulant	1.06	0.99 – 1.13	0.105
Any anti-osteoporotic drug	1.06	0.97 – 1.15	0.202
Osteoporosis	1.02	0.96 – 1.09	0.452
Cancer	1.07	1.01 – 1.14	0.029
Diabetes Mellitus	1.17	1.10 – 1.24	<0.001
Dementia	1.59	1.47 – 1.72	<0.001
Chronic polyarthritis	1.06	0.95 – 1.19	0.273
Obesity	0.90	0.84 – 0.96	0.003
Nicotine abuse	1.27	1.12 – 1.44	<0.001
Parkinson disease	1.30	1.14 – 1.47	<0.001
Rotator cuff rupture	0.94	0.83 – 1.06	0.292
Alcohol abuse	1.44	1.28 – 1.63	<0.001
Atrial fibrillation/ flutter	1.16	1.08 – 1.25	<0.001
Hypertension	1.00	0.91 – 1.10	0.987
Congestive heart failure	1.18	1.11 – 1.26	<0.001
Coronary heart disease	1.00	0.93 – 1.06	0.873
Atherosclerosis	1.02	0.95 – 1.09	0.618
Chronic kidney disease	1.19	1.12 – 1.27	<0.001
Stroke and/or CVD before PHF	1.09	1.03 – 1.16	0.005
Omarthrosis	1.03	0.88 – 1.21	0.730
Frozen shoulder	0.95	0.83 – 1.08	0.443
General compl. during IH	1.46	1.34 – 1.59	<0.001
Surgical compl. during IH	1.04	0.93 – 1.15	0.507
Approach 2			
Treatment group (RTSA vs LPF)	1.05	0.97 – 1.13	0.245
LOS during index (days)	1.01	1.01 – 1.02	<0.001
Male sex	1.44	1.30 – 1.61	<0.001
Age at index surgery (years)	1.08	1.07 – 1.08	<0.001
Year of index surgery (years)	0.98	0.96 – 1.00	0.025
Any anticoagulant	1.06	0.97 – 1.16	0.177
Any anti-osteoporotic drug	1.11	0.99 – 1.24	0.081
Osteoporosis	1.05	0.97 – 1.14	0.225
Cancer	1.06	0.98 – 1.15	0.146
Diabetes Mellitus	1.15	1.06 – 1.24	<0.001
Dementia	1.72	1.57 – 1.89	<0.001
Chronic polyarthritis	1.04	0.90 – 1.21	0.563
Obesity	0.86	0.78 – 0.94	0.001
Nicotine abuse	1.39	1.18 – 1.64	<0.001
Parkinson disease	1.35	1.16 – 1.56	<0.001
Rotator cuff rupture	0.96	0.82 – 1.13	0.602
Alcohol abuse	1.45	1.24 – 1.70	<0.001
Atrial fibrillation/ flutter	1.17	1.06 – 1.28	0.001
Hypertension	1.05	0.92 – 1.20	0.483
Congestive heart failure	1.26	1.16 – 1.37	<0.001
Coronary heart disease	0.98	0.90 – 1.06	0.554
Atherosclerosis	1.06	0.97 – 1.17	0.218
Chronic kidney disease	1.24	1.14 – 1.34	<0.001
Stroke and/or CVD before PHF	1.05	0.97 – 1.14	0.229
Omarthrosis	1.07	0.87 – 1.32	0.541
Frozen shoulder	0.95	0.80 – 1.13	0.556
General compl. during IH	1.50	1.35 – 1.67	<0.001
Surgical compl. during IH	1.15	1.02 – 1.31	0.029

Approach 1: COVID-19 infection was considered as a competing risk event.

Approach 2: Subset analysis including only patients with index surgery before 2020 ($N = 10,879$) and with end of follow-up defined as 31.12.2019.

CVD – cerebrovascular disease, HR – hazard ratio, IH – index hospitalization, LOS – length of hospital stay, LPF – locked plate fixation, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty. Clustering of data (i.e. dependency of matching pairs) was not considered.

Table E17: Sensitivity analysis considering COVID-19 – surgical complications

Variable	HR	95% CI	p-value (two-sided)
Approach 1			
Treatment group (RTSA vs LPF)	0.47	0.42 – 0.52	<0.001
LOS during index (days)	1.01	1.00 – 1.01	0.054
Male sex	1.34	1.16 – 1.54	<0.001
Age at index surgery (years)	0.97	0.97 – 0.98	<0.001
Year of index surgery (years)	0.95	0.93 – 0.96	<0.001
Any anticoagulant	0.98	0.86 – 1.11	0.745
Any anti-osteoporotic drug	0.91	0.78 – 1.07	0.254
Osteoporosis	1.07	0.96 – 1.19	0.243
Cancer	1.04	0.93 – 1.16	0.472
Diabetes Mellitus	1.01	0.91 – 1.12	0.895
Dementia	1.02	0.86 – 1.20	0.859
Chronic polyarthritis	1.22	1.02 – 1.45	0.028
Obesity	1.23	1.09 – 1.37	<0.001
Nicotine abuse	1.08	0.90 – 1.30	0.415
Parkinson disease	1.28	1.01 – 1.61	0.039
Rotator cuff rupture	1.07	0.86 – 1.32	0.555
Alcohol abuse	1.23	1.02 – 1.49	0.030
Atrial fibrillation/ flutter	1.02	0.88 – 1.17	0.838
Hypertension	1.27	1.07 – 1.50	0.006
Congestive heart failure	1.02	0.90 – 1.56	0.736
Coronary heart disease	1.10	0.98 – 1.24	0.118
Atherosclerosis	1.06	0.94 – 1.21	0.343
Chronic kidney disease	1.13	1.00 – 1.26	0.043
Stroke and/or CVD before PHF	0.97	0.87 – 1.09	0.637
Omarthrosis	0.92	0.71 – 1.20	0.547
Frozen shoulder	1.18	0.96 – 1.46	0.122
General compl. during IH	1.08	0.90 – 1.29	0.412
Surgical compl. during IH	1.25	1.05 – 1.50	0.014
Approach 2			
Treatment group (RTSA vs LPF)	0.42	0.37 – 0.48	<0.001
LOS during index (days)	1.01	1.00 – 1.01	0.046
Male sex	1.38	1.16 – 1.65	<0.001
Age at index surgery (years)	0.97	0.96 – 0.98	<0.001
Year of index surgery (years)	0.98	0.95 – 1.00	0.088
Any anticoagulant	0.93	0.80 – 1.09	0.359
Any anti-osteoporotic drug	0.84	0.69 – 1.03	0.088
Osteoporosis	1.13	0.99 – 1.29	0.070
Cancer	1.04	0.91 – 1.19	0.601
Diabetes Mellitus	1.02	0.89 – 1.16	0.779
Dementia	1.14	0.94 – 1.38	0.196
Chronic polyarthritis	1.12	0.90 – 1.40	0.318
Obesity	1.19	1.03 – 1.37	0.016
Nicotine abuse	1.14	0.90 – 1.44	0.288
Parkinson disease	1.28	0.97 – 1.68	0.083
Rotator cuff rupture	1.22	0.96 – 1.56	0.111
Alcohol abuse	1.15	0.90 – 1.45	0.263
Atrial fibrillation/ flutter	1.05	0.89 – 1.25	0.572
Hypertension	1.26	1.02 – 1.56	0.031
Congestive heart failure	0.99	0.85 – 1.15	0.876
Coronary heart disease	1.13	0.98 – 1.31	0.100
Atherosclerosis	1.05	0.90 – 1.23	0.551
Chronic kidney disease	1.17	1.02 – 1.35	0.030
Stroke and/or CVD before PHF	1.07	0.93 – 1.24	0.320
Omarthrosis	0.77	0.54 – 1.11	0.160
Frozen shoulder	1.13	0.87 – 1.47	0.373
General compl. during IH	1.15	0.93 – 1.43	0.209
Surgical compl. during IH	1.12	0.88 – 1.39	0.385

Approach 1: COVID-19 infection was considered as a competing risk event.

Approach 2: Subset analysis including only patients with index surgery before 2020 ($N = 10,879$) and with end of follow-up defined as 31.12.2019.

CVD – cerebrovascular disease, HR – hazard ratio, IH – index hospitalization, LOS – length of hospital stay, LPF – locked plate fixation, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty. Clustering of data (i.e. dependency of matching pairs) was not considered.

Table E18: Sensitivity analysis considering COVID-19 – minor outpatient complications

Variable	HR	95% CI	p-value (two-sided)
Approach 1			
Treatment group (RTSA vs LPF)	1.02	0.96 – 1.08	0.470
LOS during index (days)	1.00	0.99 – 1.00	0.091
Male sex	0.92	0.84 – 1.02	0.097
Age at index surgery (years)	0.98	0.97 – 0.98	<0.001
Year of index surgery (years)	0.98	0.97 – 0.99	<0.001
Any anticoagulant	0.98	0.91 – 1.05	0.513
Any anti-osteoporotic drug	1.04	0.96 – 1.14	0.353
Osteoporosis	1.12	1.05 – 1.19	<0.001
Cancer	1.04	0.97 – 1.10	0.279
Diabetes Mellitus	0.91	0.86 – 0.97	0.005
Dementia	0.66	0.59 – 0.74	<0.001
Chronic polyarthritis	1.10	0.99 – 1.23	0.082
Obesity	1.19	1.12 – 1.28	<0.001
Nicotine abuse	0.99	0.88 – 1.11	0.802
Parkinson disease	0.82	0.68 – 0.97	0.022
Rotator cuff rupture	1.66	1.48 – 1.87	<0.001
Alcohol abuse	0.93	0.82 – 1.06	0.286
Atrial fibrillation/ flutter	0.96	0.88 – 1.05	0.336
Hypertension	1.10	1.00 – 1.20	0.042
Congestive heart failure	0.96	0.89 – 1.04	0.320
Coronary heart disease	1.04	0.97 – 1.12	0.235
Atherosclerosis	1.05	0.97 – 1.13	0.254
Chronic kidney disease	1.03	0.96 – 1.11	0.364
Stroke and/or CVD before PHF	1.06	0.99 – 1.14	0.087
Omarthrosis	1.79	1.58 – 2.03	<0.001
Frozen shoulder	3.41	3.08 – 3.79	<0.001
General compl. during IH	0.77	0.68 – 0.87	<0.001
Surgical compl. during IH	1.29	1.16 – 1.44	<0.001
Approach 2			
Treatment group (RTSA vs LPF)	1.04	0.97 – 1.12	0.246
LOS during index (days)	1.00	0.99 – 1.00	0.180
Male sex	0.95	0.84 – 1.07	0.371
Age at index surgery (years)	0.98	0.97 – 0.99	<0.001
Year of index surgery (years)	1.00	0.99 – 1.02	0.659
Any anticoagulant	1.01	0.92 – 1.10	0.918
Any anti-osteoporotic drug	1.00	0.89 – 1.12	0.936
Osteoporosis	1.11	1.02 – 1.20	0.012
Cancer	1.02	0.94 – 1.11	0.642
Diabetes Mellitus	0.92	0.85 – 1.00	0.055
Dementia	0.71	0.62 – 0.82	<0.001
Chronic polyarthritis	1.08	0.94 – 1.24	0.268
Obesity	1.22	1.12 – 1.33	<0.001
Nicotine abuse	1.01	0.87 – 1.17	0.896
Parkinson disease	0.79	0.64 – 0.99	0.036
Rotator cuff rupture	1.69	1.46 – 1.94	<0.001
Alcohol abuse	0.94	0.80 – 1.11	0.481
Atrial fibrillation/ flutter	0.94	0.84 – 1.05	0.261
Hypertension	1.06	0.94 – 1.19	0.325
Congestive heart failure	1.02	0.93 – 1.12	0.715
Coronary heart disease	1.02	0.93 – 1.12	0.684
Atherosclerosis	1.00	0.91 – 1.11	0.979
Chronic kidney disease	1.01	0.92 – 1.10	0.885
Stroke and/or CVD before PHF	1.04	0.95 – 1.13	0.432
Omarthrosis	1.83	1.57 – 2.14	<0.001
Frozen shoulder	3.59	3.17 – 4.06	<0.001
General compl. during IH	0.79	0.68 – 0.92	0.002
Surgical compl. during IH	1.25	1.09 – 1.44	0.001

Approach 1: COVID-19 infection was considered as a competing risk event.

Approach 2: Subset analysis including only patients with index surgery before 2020 ($N = 10,879$) and with end of follow-up defined as 31.12.2019.

CVD – cerebrovascular disease, HR – hazard ratio, IH – index hospitalization, LOS – length of hospital stay, LPF – locked plate fixation, PHF – proximal humeral fracture, RTSA – reverse total shoulder arthroplasty. Clustering of data (i.e. dependency of matching pairs) was not considered.