

Kaplan-Meier Survival Analysis Overestimates the Risk of Revision Arthroplasty

A Meta-analysis

Sarah Lacny MSc, Todd Wilson BSc, Fiona Clement PhD, Derek J. Roberts MD,
Peter D. Faris PhD, William A. Ghali MD, MPH, Deborah A. Marshall PhD

Published online: 25 March 2015

© The Association of Bone and Joint Surgeons® 2015

Abstract

Background Although Kaplan-Meier survival analysis is commonly used to estimate the cumulative incidence of revision after joint arthroplasty, it theoretically overestimates the risk of revision in the presence of competing risks (such as death). Because the magnitude of overestimation is

not well documented, the potential associated impact on clinical and policy decision-making remains unknown.

Questions/purposes We performed a meta-analysis to answer the following questions: (1) To what extent does the Kaplan-Meier method overestimate the cumulative incidence of revision after joint replacement compared with alternative competing-risks methods? (2) Is the extent of overestimation influenced by followup time or rate of competing risks?

Methods We searched Ovid MEDLINE, EMBASE, BIOSIS Previews, and Web of Science (1946, 1980, 1980, and 1899, respectively, to October 26, 2013) and included article bibliographies for studies comparing estimated cumulative incidence of revision after hip or knee arthroplasty obtained using both Kaplan-Meier and competing-risks methods. We excluded conference abstracts, unpublished studies, or studies using simulated data sets. Two reviewers independently extracted data and evaluated the quality of reporting of the included studies. Among 1160 abstracts identified, six studies were included in our meta-analysis. The principal reason for the steep attrition (1160 to six) was that the initial search was for studies in any clinical area that compared the cumulative incidence estimated using the Kaplan-Meier versus competing-risks

One of the authors (SL) is supported by the Canadian Institutes of Health Research Master's Award and Alberta Innovates-Health Solutions Graduate Studentship Award. One of the authors (DJR) is supported by an Alberta Innovates-Health Solutions Clinician Fellowship Award, a Knowledge Translation Canada Strategic Training in Health Research Fellowship, and funding from the Canadian Institutes of Health Research. One of the authors (DAM) is a Canada Research Chair in Health Systems and Services Research and Arthur J. E. Child Chair in Rheumatology.

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request.

Clinical Orthopaedics and Related Research® neither advocates nor endorses the use of any treatment, drug, or device. Readers are encouraged to always seek additional information, including FDA-approval status, of any drug or device prior to clinical use.

This work was performed at the University of Calgary, Calgary, Alberta, Canada.

S. Lacny, T. Wilson, F. Clement
Department of Community Health Sciences, Cumming School of
Medicine, University of Calgary, 3280 Hospital Drive NW,
Calgary, AB T2N 4Z6, Canada

F. Clement, W. A. Ghali, D. A. Marshall
O'Brien Institute for Public Health, University of Calgary, 3280
Hospital Drive NW, Calgary, AB T2N 4Z6, Canada

D. J. Roberts
Departments of Community Health Sciences and Surgery,
Cumming School of Medicine, University of Calgary, Foothills
Medical Centre, 3280 Hospital Drive NW, Calgary,
AB T2N 4Z6, Canada

P. D. Faris
Research Priorities and Implementation, Alberta Health
Services, Foothills Medical Centre, 1403-29 Street NW, Calgary,
AB T2N 2T9, Canada

P. D. Faris, D. A. Marshall
Alberta Bone and Joint Health Institute, 3280 Hospital Drive
NW, Calgary, AB T2N 4Z6, Canada

W. A. Ghali, D. A. Marshall
Departments of Medicine and Community Health Sciences,
Cumming School of Medicine, University of Calgary, 3280
Hospital Drive NW, Calgary, AB T2N 4Z6, Canada

methods for any event (not just the cumulative incidence of hip or knee revision); we did this to minimize the likelihood of missing any relevant studies. We calculated risk ratios (RRs) comparing the cumulative incidence estimated using the Kaplan-Meier method with the competing-risks method for each study and used DerSimonian and Laird random effects models to pool these RRs. Heterogeneity was explored using stratified meta-analyses and metaregression.

Results The pooled cumulative incidence of revision after hip or knee arthroplasty obtained using the Kaplan-Meier method was 1.55 times higher (95% confidence interval, 1.43–1.68; $p < 0.001$) than that obtained using the competing-risks method. Longer followup times and higher proportions of competing risks were not associated with increases in the amount of overestimation of revision risk by the Kaplan-Meier method (all $p > 0.10$). This may be due to the small number of studies that met the inclusion criteria and conservative variance approximation.

Conclusions The Kaplan-Meier method overestimates risk of revision after hip or knee arthroplasty in populations where competing risks (such as death) might preclude the occurrence of the event of interest (revision). Competing-risks methods should be used to more accurately estimate the cumulative incidence of revision when the goal is to plan healthcare services and resource allocation for revisions.

Introduction

Time to revision after joint arthroplasty is an important factor for assessing the quality of joint replacements, monitoring implant performance, and informing health policy planning decisions. The measure will play an increasingly important role in coming years given the growing demand for primary and revision hip and knee replacements [26, 28], particularly in younger, more physically active patients, who are likely to outlive their implants and undergo revision surgery [27].

Monitoring the incidence of revisions over time requires survival analysis because for some patients, time to revision is unknown because they are lost to followup, die before receiving a revision, or are alive and unrevised at the end of the observation period. Kaplan-Meier survival analysis [23] is often used, as seen in the orthopaedic literature and among joint replacement registries, to estimate the cumulative incidence of

revision after joint arthroplasty. However, because the method was designed to estimate the time to a single event that will eventually occur for everyone (such as death), it does not consider other “competing risks” that may preclude and alter the probability of the event of interest [18]; for example, a patient who has died cannot subsequently undergo revision surgery, and using the Kaplan-Meier estimator in this setting violates one of its principal assumptions regarding the independence of events. Stated otherwise, when estimating time to revision, death represents an important competing risk. By treating deaths as censored observations, the Kaplan-Meier method assumes the risk of revision is independent of the risk of death. Consequently, the Kaplan-Meier method theoretically overestimates the cumulative incidence of an event in the presence of competing risks [7, 36]. This bias is particularly problematic for older arthroplasty populations with high mortality rates and in studies involving longer followup durations [38], in which a larger number of patients are followed until death rather than censoring.

Alternative statistical methods have been developed to estimate cumulative incidence of an event in competing-risks settings. By acknowledging that patients can no longer be revised after death, competing-risks methods provide an estimate of the number of revisions expected to occur at a specific time point. Thus, the competing-risks method may provide more accurate estimates that can be used to inform healthcare planning and policy decisions [12, 25, 38]. In contrast, the Kaplan-Meier method estimates the probability of a revision at a certain time point assuming patients cannot die and may be useful for informing individual patients of their risk of revision under the assumption they will live a certain number of years after their primary surgery [16, 25, 31, 38]. Given that these methods differ in their treatment of patients who experience a competing event before the event of interest, in situations in which no patients die before revision throughout the duration of followup, the Kaplan-Meier method and competing-risks method will produce the same estimate.

The application of competing-risks methods is now feasible using a variety of statistical software programs. However, recent studies have noted the Kaplan-Meier method continues to be used in the presence of competing risks [24, 40]. The purpose of our systematic review and meta-analysis was therefore to provide empiric evidence of the magnitude of overestimation of the Kaplan-Meier compared with competing-risks method when estimating the cumulative incidence of revision. We also sought to examine whether the extent of overestimation is influenced by duration of followup or the rate of competing events relative to events of interest.

D. A. Marshall (✉)
Health Research Innovation Centre, Room 3C56, 3280 Hospital
Drive NW, Calgary, AB T2N 4Z6, Canada
e-mail: damarsha@ucalgary.ca

D. A. Marshall
McCaig Institute for Bone and Joint Health, Calgary, AB, Canada

Materials and Methods

Our search strategy was developed in consultation with a medical librarian/information scientist. We searched Ovid MEDLINE, EMBASE, BIOSIS Previews, and Web of Science from the first available date of each database (1946, 1980, 1980, and 1899, respectively) to October 26, 2013, without publication date, language, or other restrictions using Medical Subject Heading (MeSH) terms and keywords to cover the themes Kaplan-Meier and competing risks. For the Kaplan-Meier theme, we combined the MeSH term “Kaplan Meier Estimate” (Emtree term “Kaplan Meier method”) with a title and abstract keyword search for Kaplan Meier* or Kaplan-Meier* or Kaplan-meier* or Kaplan*Meyer* or censor*. For the competing-risks theme, we used a title and abstract search using the terms competing or cumulative incidence function* or cause*specific hazard*or sub*distribution*. The two themes were subsequently combined using the Boolean operator “AND”. To identify additional articles, we also used the PubMed “related articles” feature and hand-searched bibliographies of included studies and other potentially relevant citations identified during the search process.

Two independent reviewers (SL, TW) screened all identified titles and abstracts. Abstracts deemed potentially relevant by either reviewer were subsequently read in full. Full-text articles were included if: (1) both Kaplan-Meier and competing-risks methods, as defined subsequently, were applied to estimate the cumulative incidence of revision after joint arthroplasty; (2) cumulative incidence estimates were provided for both methods (either as point estimates or graphically); and (3) studies involved humans. Conference abstracts, unpublished studies, and studies using simulated data sets were excluded. In situations where multiple studies analyzed the same data or data subsets, we included the study that reported the most detailed information with respect to requirements for our meta-analysis (eg, count of events, number at risk) or the study with the earliest publication date. Agreement between reviewers was quantified using the κ statistic [30]. Disagreements were resolved by consensus.

The Kaplan-Meier method was defined as the Kaplan-Meier failure function (complement of the Kaplan-Meier survival function), which estimates the probability of an event of interest occurring at a specific time point among those who had not already experienced that event. Patients who die are excluded from the at-risk population at the time of their deaths and, similar to those lost to followup, are assumed to have the same probability of revision as those remaining in the risk set [38]. The competing-risks method was defined as the cumulative incidence function using the approach of Kalbfleisch and Prentice [22], which

estimates the probability of the event of interest occurring at a specific time point given that neither the event of interest nor the competing event has yet occurred. Thus, the competing-risks method depends on the risk of the event of interest and the competing event, whereas the Kaplan-Meier estimate considers only the event of interest.

Among 1162 unique citations identified by our search strategy, 101 full-text articles were assessed for eligibility. Seven cohort studies compared the Kaplan-Meier and competing-risks methods when estimating the time to revision after joint arthroplasty and were included in our systematic review (κ statistic = 1) [6, 7, 16, 17, 24, 38, 41], of which six included enough data to be included in our meta-analysis [6, 7, 16, 17, 24, 41] (Fig. 1). Publication years ranged from 2001 to 2012 (Table 1). Five studies assessed time to revision after partial or total hip arthroplasty (THA) [6, 16, 17, 38, 41]; one assessed time to revision after acetabular revision [24] and one assessed time to revision after total knee arthroplasty (TKA) with a megaprosthesis after bone tumor resection [7]. Death was identified as a competing risk in all studies. One study also considered amputation as a competing risk [7].

The same two reviewers (SL, TW) independently extracted data using a predesigned and pilot-tested data extraction tool. We extracted data regarding author, year of publication, study design, sample size, age of the population, followup time, type and number of events of interest, competing events, and the statistical software package used.

The primary data elements extracted from each study were the cumulative incidence estimates obtained for the Kaplan-Meier and competing-risks methods. These outcomes are often reported at multiple time points throughout a followup period; therefore, we extracted estimates and 95% confidence intervals (CIs), when reported, across all reported time points for each study. For studies reporting multiple stratified analyses, we extracted data for each stratum. To ensure only mutually exclusive strata from each study were included, we conducted two separate analyses for strata containing: (1) the largest number of events of interest (ie, revisions); and (2) the highest rate of competing risks relative to events of interest (ie, number of competing risks observed/number of events of interest observed). For example, Gillam et al. [17] analyzed three subsets of data. The subset with the largest number of events of interest compared the cumulative incidence of revision for patients receiving THA for osteoarthritis who were younger than 70 years old with patients who were aged 70 years and older. The subset with the highest proportion of competing risks compared the cumulative incidence of revision after THA for two types of monoblock prostheses.

Because no validated tool is available to assist in examining the quality of reporting specifically for survival

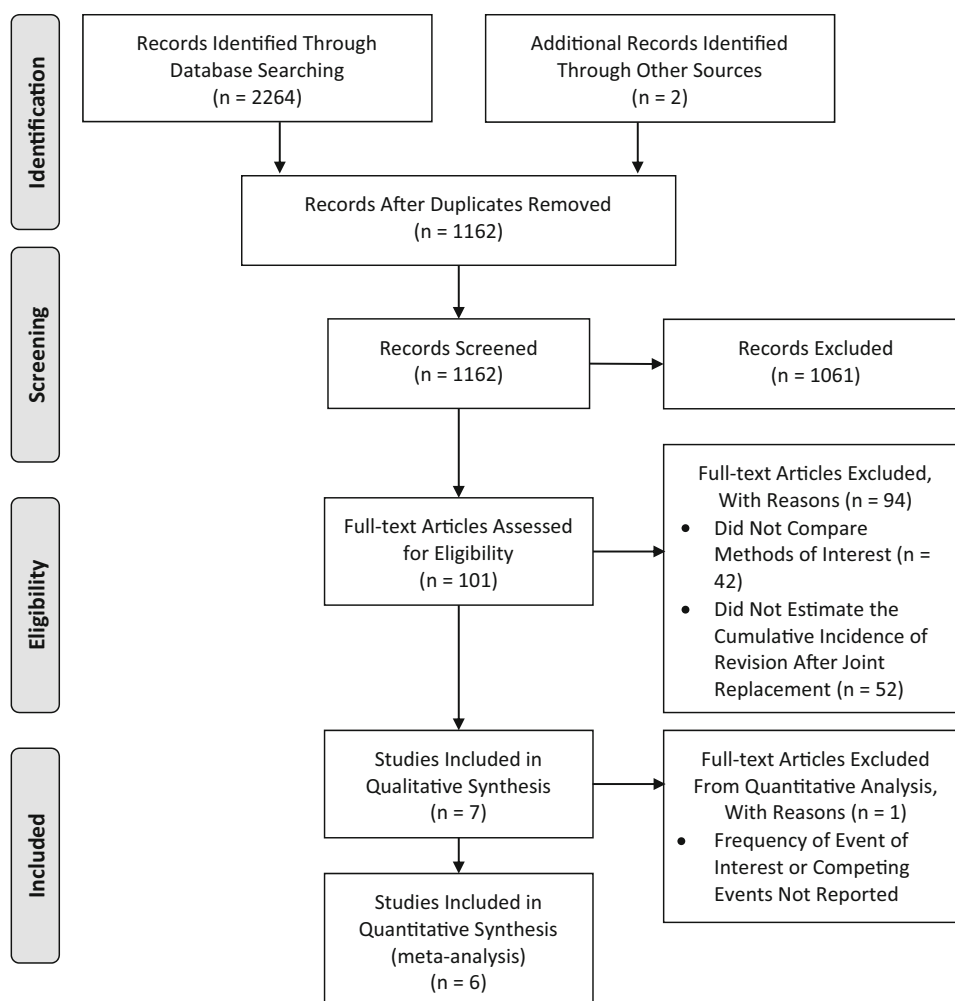


Fig. 1 The flow of articles through the systematic review process is illustrated using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram.

analysis studies, we developed criteria based on recommendations and guidelines for reporting these types of analyses [1, 3, 9, 12, 34, 35, 38]. Nine criteria were assessed independently by the same two reviewers (SL, TW), who asked: (1) Was the number of patients at risk presented at each followup time? (2) Was the observed number of events of interest and competing events provided? (3) Were losses to followup clearly described? (4) Was the handling of losses to followup described (eg, treated as censored at the time of loss to followup)? (5) Was a description of censoring provided? (6) Were graphic representations of cumulative incidence of the event of interest and competing event(s) provided for the Kaplan-Meier method? (7) Were graphic representations of cumulative incidence of the event of interest and competing event(s) provided for the competing-risks method? (8) Were estimates of precision of the cumulative incidence provided (ie, SEs or CIs)? (9) Was the name of the statistical software provided? Questions answered “yes”

received one point and those answered “no” received zero points. We calculated the percentage of studies that received points for each criterion to assess the overall quality of reporting for the body of our study literature and identified inconsistencies in reporting. Of the seven studies included, three (43%) provided the number at risk at each followup time or the name of the statistical software used (Table 2). The number of events was provided by six studies (86%). Five studies (71%) reported the number of losses to followup, four of which described how losses to followup were accounted for in their analysis. All seven studies described the censoring mechanisms used, although only three studies (43%) reported the number of censored observations. Cumulative incidence curves were provided in all seven studies. Only three studies (43%) provided CIs for both Kaplan-Meier and competing-risks methods. We calculated risk ratios (RRs) to compare the cumulative incidence estimated using the Kaplan-Meier method with the competing-risks method for each study, where:

Table 1. Characteristics of studies measuring time to revision following hip or knee arthroplasty included in systematic review (n = 7) and meta-analysis (n = 6)

Study, author, publication year, country	Study characteristic—study design, population size, age of population	Event of interest—verification of events [†]	Competing event(s)—verification of events	Followup [‡]	Competing-risks method	Software
Biau and Hamadouche [6], 2011, France	Cohort study; 118 THA in 106 patients between 1979 and 1980; mean patient age: 62.2 years (range, 32–89 years)	Revision THA - Data obtained from patient contact or family contact of deceased patients	Death - Data obtained from patient contact or family contact of deceased patients	Maximum 20 years	Cumulative incidence function	Unknown
Biau et al. [7], 2007, France	Cohort study; 53 men and 38 women patients underwent resection of malignant knee tumor followed by reconstruction with custom-made megaprosthesis (from May 1972 to April 1994); median patient age: 27 years (range, 12–78 years)	Revision of a total knee megaprosthesis not related to malignant knee tumor - Data retrieved retrospectively from health records	Death or amputation for reasons unrelated to the implant - Data retrieved retrospectively from health records	Maximum 15 years Median 62 months (range, 0.5–343 months)	Cumulative incidence estimator	R 1.9.1 (R Foundation for Statistical Computing, Vienna, Austria) S-Plus 2000 (Mathsoft, Seattle, WA, USA)
Fenemma and Lubsen [16], 2010, The Netherlands	Cohort study; 405 cemented THAs operated consecutively between January 1993 and May 1994; mean age not reported	Revision of a total hip prosthesis - Verification of events not indicated	Death - Verification of events not indicated	Maximum 12 years	Cumulative incidence of competing risks	Excel 2003 (Microsoft Inc, Redmond, WA, USA)
Gillam et al. [17], 2010, Australia	Cohort (registry) study; 91,795 patients who received partial or total arthroplasty for fractured neck of femur (patients aged 75–84 years) and of patients who received THA for osteoarthritis (patients younger than 70 years versus patients 70 years or older) from January 1, 2002, to December 31, 2008; mean age not reported	First revision of a total hip prosthesis - Data from the Australian Orthopaedic Association National Joint Replacement Registry	Death - Data from the National Death Index, maintained by the Australian Institute of Health and Welfare	Maximum 6 years	Cumulative incidence function	Unknown
Keurentjes et al. [24], 2012, The Netherlands	Cohort study; 62 acetabular revisions in 58 patients between January 1989 and March 1986 at the Radboud University Medical Center in Nijmegen, The Netherlands; mean patient age: 59.2 years (range, 23–82 years)	Revision of an acetabular revision - Verification of events not indicated	Death - Verification of events not indicated	Mean 23 years	Cumulative incidence function	R (R Foundation for Statistical Computing)
Ranstam et al. [38], 2011,* Norway, Denmark, and Sweden	Cohort (registry) study; 84,843 hip replacements recorded by the Danish Hip Arthroplasty Register between 1995 and 2008; mean patient age not reported	Implant failure after THA - Data from the Danish Hip Arthroplasty Register	Death - Verification of events not indicated	Maximum 10 years	Cumulative incidence function	Unknown

Table 1. continued

Study, author, publication year, country	Study characteristic—study design, population size, age of population	Event of interest—verification of events [†]	Competing event(s)—verification of events	Followup [‡]	Competing-risks method	Software
Schwarzer et al. [41], 2001, Germany and Switzerland	Cohort study; 239 total hip prostheses made of a titanium alloy (Titan GS; Landos, Inc, Malvern, PA, USA) implanted between July 1987 and November 1993 (followed until March 1997) in a specialized hospital in Liestal, Switzerland; 68% of patients aged > 65 years	Revision of a total hip prosthesis - Verification of events not indicated	Death - Verification of events not indicated	Median 6.0 years 1368.1 person-years	Cumulative incidence using a competing-risks model	Unknown

* Excluded from meta-analysis because frequencies of events (ie, revisions and deaths) not reported; † data regarding number and time of event may have been obtained using administrative data, registry data, medical records, etc; ‡ mean, median, or maximum followup time or total person-years.

$$RR = \frac{\text{Cumulative Incidence}_{\text{Kaplan-Meier}}}{\text{Cumulative Incidence}_{\text{competing-risks}}}$$

Because we did not have individual patient data required to calculate the variance around the RRs, we used an approximation that has been proposed to estimate the variance (var) of a hazard ratio (HR) using summary data [43], where:

$$\text{Var}(\log(\text{HR})) \approx \frac{1}{\text{observed number of events of interest} - \frac{1}{\text{number at risk}}}$$

Because we could not find an approximation for the variance of the ratio of cumulative incidences, we used this approximation for the log HR given that both the RR and HR compare the measure of occurrence of events over time, while accounting for censoring, in the form of a ratio. We also performed a sensitivity analysis using an alternative approximation [43], where:

$$\text{Var}(\log(\text{HR})) \approx \frac{4}{\text{observed number of events of interest}}$$

It is important to note that these variances were primarily used for the purposes of weighting each individual study for our meta-analysis. Therefore, the CIs estimated using this variance approximation must be interpreted carefully.

A DerSimonian and Laird [11] random-effects model was used to pool RR estimates across studies. RR estimates were log-transformed before being entered into the model. As we anticipated, the time points at which estimates were reported varied across studies, so we included estimates reported at the longest followup time point for each study. To assess inter-study heterogeneity, we inspected forest plots stratified by followup time (< 10 years, ≥ 10 years) and the rate of competing risks relative to events of interest (< 1, 1–10, > 10). We did not observe differences in the magnitude of overestimation of the Kaplan-Meier method when assessing these forest plots (data not shown). We used univariate metaregression to examine the effect of the covariates on the estimated pooled RR with p values < 0.10 considered significant given the low power of these tests [14]. All analyses were performed using Stata/SE Version 12.0 (StataCorp, College Station, TX, USA).

Results

To What Extent Does the Kaplan-Meier Method Overestimate the Cumulative Incidence of Revision?

Kaplan-Meier survivorship resulted in a larger estimate of the risk of revision than did the competing-risks estimator

Table 2. Quality of reporting assessment for seven studies included in the systematic review

Quality of reporting criterion	Biau et al. [6]	Biau et al. [7]	Fenemma and Lubsen [16]	Gillam et al. [17]	Keurentjes et al. [24]	Ranstam et al. [38]*	Schwarzer et al. [41]
Was the number at risk presented at each followup time? (yes; no)	No	Yes	No	Yes	No	No	Yes
Were the number of events of interest and competing events provided? (yes; no)	Yes	Yes	Yes	Yes	Yes	No	Yes
Was the number of losses to followup provided? (yes—count, proportion, or reason provided; no)	Yes,	count [†]	Yes, count and reason	Yes, count	NA [‡]	Yes	No
Yes							
Was the handling of losses to followup explicitly described? (yes; no)	No	Yes	Yes	NA [‡]	Yes	No	Yes
Was an adequate description of censoring provided? (yes—count provided; no)	Yes	Yes	Yes, count	Yes, count	Yes, count	Yes	Yes
Were cumulative incidence curves provided?							
KM method	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CR method	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Were estimates of precision around the cumulative incidence provided? (yes—described; no)	Yes, CIs	Yes, CI for KM method only	Yes, CIs	Yes, CIs	Yes, CI for KM method only	No	No
Was the name of the statistical software provided? (yes; no)	No	Yes	Yes	No	Yes	No	No

* Excluded from meta-analysis because frequencies of events (ie, revisions and deaths) were not reported.

[†] Provided in original article [21]; [‡]no losses to followup; CI = confidence interval; CR = competing-risks; KM = Kaplan-Meier; NA = not applicable.

when we considered the seven strata within the population of included studies that contained a high proportion of patients who had died during the followup period. The pooled RR was 1.55 (95% CI, 1.43–1.68; $p < 0.001$), indicating that the cumulative incidence of revision estimated using the Kaplan-Meier approach was 55% greater than that obtained using the competing-risks estimator (Fig. 2A). The RRs for these six studies, including seven mutually exclusive strata, ranged from 1.15 (95% CI, 0.82–1.62; $p = 0.429$), demonstrating no difference in RR between Kaplan-Meier and competing-risks estimators, to 1.79 (95% CI, 1.43–2.24; $p < 0.001$), demonstrating a significant difference in RR (Fig. 2A).

When we considered the seven strata that recorded the largest number of revisions, the Kaplan-Meier estimate of revision risk was once again greater than the competing-risks method. The pooled RR was 1.07 (95% CI, 1.00–1.14; $p = 0.049$), demonstrating that the cumulative incidence estimated using the Kaplan-Meier method was 1.07 times greater than the competing-risks method, corresponding to a relative increase in estimation of 7% (Fig. 2B). RRs for these studies ranged from 1.02 (95% CI, 0.96–1.08; $p = 0.540$) to 1.62 (95% CI, 1.00–2.63;

$p = 0.051$), both of which demonstrate no difference in RR between Kaplan-Meier and competing-risks estimators.

Is the Extent of Overestimation Influenced by Followup Time or Frequency of Competing Risks?

Increasing duration of followup was not associated with an increase in the amount of overestimation of revision risk by the Kaplan-Meier method. This may be due to the small number of studies that met the inclusion criteria and conservative variance approximation. Using metaregression, we found the RR comparing the Kaplan-Meier estimator with the competing-risks estimator for studies with followup times less than 10 years was not different than the RR obtained for studies with followup times greater than or equal to 10 years in either our analysis of strata containing the largest number of revisions ($p = 0.125$) or our analysis of strata containing the highest proportion of competing risks ($p = 0.203$) (Table 3).

Increasing the ratio of competing risks to events of interest was also not associated with an increase in the amount of overestimation of revision risk by the Kaplan-Meier method.

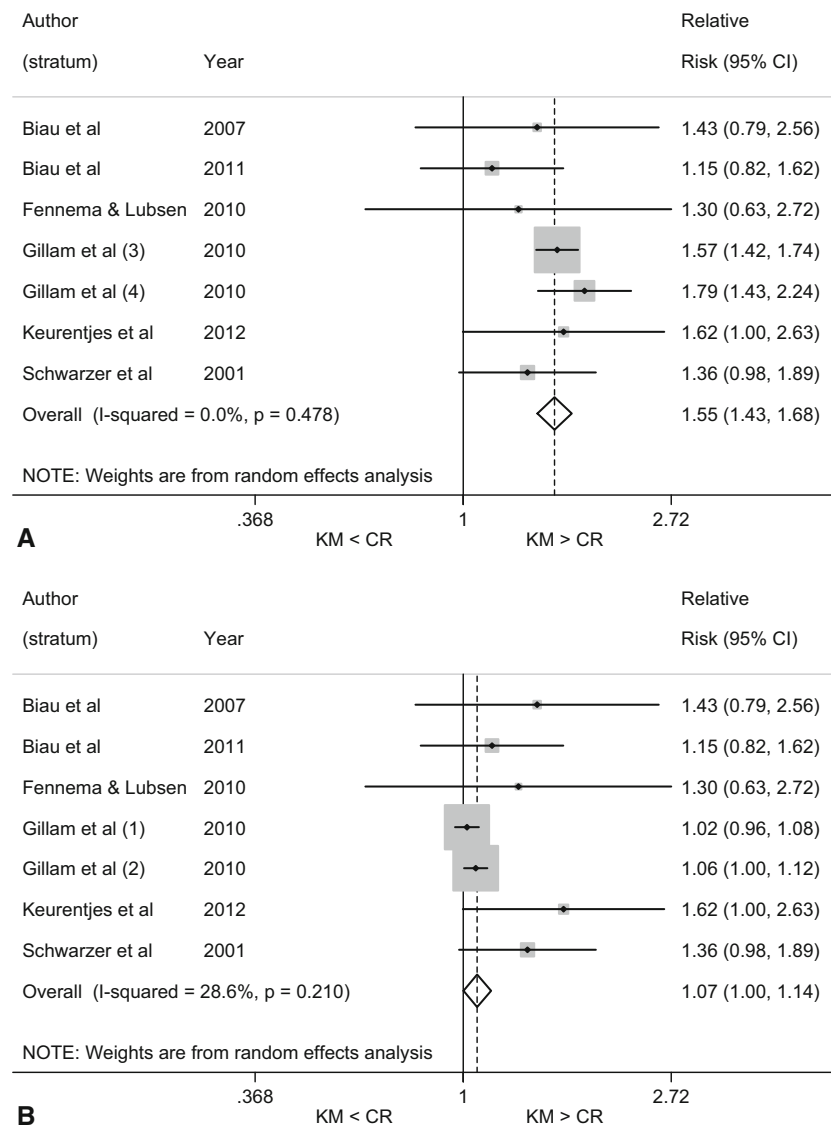


Fig. 2A–B Forest plots of RRs compare the cumulative incidence of revision after hip or knee arthroplasty obtained using the Kaplan-Meier method versus competing-risks method for seven strata (six studies*) containing (A) the highest ratio of competing events to events of interest; and (B) the largest number of revisions. *Gillam et al. [17] estimated the cumulative incidence of revision after THA for three nonmutually exclusive subsets of data. The subset with the

largest number of events of interest included two mutually exclusive strata: patients with osteoarthritis aged < 70 years and patients with osteoarthritis aged \geq 70 years. The subset with the highest rate of competing risks included two mutually exclusive strata (cementless Austin Moore prostheses and cemented Thompson prostheses). KM = Kaplan-Meier; CR = competing risks.

Again, this may be due to the small number of studies that met the inclusion criteria and conservative variance approximation. When we considered the seven strata with the largest number of revisions, there were no differences between the RR comparing the Kaplan-Meier and competing-risks estimators for studies with a ratio of competing risks to events of interest less than one compared with the RR for studies with ratios between one and 10 ($p = 0.342$) or greater than 10 ($p = 0.581$) (Table 3). Similarly, when we considered the seven strata that contained a high proportion of patients who

had died during the followup period, there were no differences between the RRs obtained for studies with a ratio of competing risks to events of interest between one and 10 compared with the RR for studies with ratios greater than 10 ($p = 0.161$). There were no strata with ratios less than one for our analysis of strata containing the highest proportion of patients who died.

Applying an alternative variance approximation (defined in the [Materials and Methods](#)) produced similar results for all analyses (data not shown).

Table 3. Meta-analysis and univariate metaregression results for identifying covariates to explain heterogeneity in the estimated pooled RRs: Kaplan-Meier versus competing-risks method*

Strata	Largest number of EI			Highest ratio of CR to EI		
	Number of strata	Meta-analysis RR (95% CI)	Metaregression p value	Number of strata	Meta-analysis RR (95% CI)	Metaregression p value
Followup						
< 10 years	3	1.05 (0.99–1.12)		3	1.59 (1.45–1.73)	
≥ 10 years	4	1.31 (1.03–1.66)	0.125	4	1.31 (1.03–1.66)	0.203
Ratio of CR to EI [†]						
< 1	1	1.02 (0.96–1.08)		0		
1–10	5	1.18 (1.01–1.38)	0.342	4	1.33 (1.09–1.62)	
> 10	1	1.30 (0.63–2.72)	0.581	3	1.60 (1.46–1.75)	0.161

RR = risk ratio; EI = event of interest; CR = competing risks; CI = confidence interval.

$$RR = \frac{\text{Cumulative Incidence}_{\text{Kaplan-Meier}}}{\text{Cumulative Incidence}_{\text{Competing-risks}}}$$

* n = 6 studies, 7 strata; Gillam et al. [17] estimated the cumulative incidence of revision after THA for three nonmutually exclusive subsets of data; the subset with the largest number of EI included two mutually exclusive strata (patients with osteoarthritis aged < 70 years, and those aged ≥ 70 years); the subset with the highest rate of CRs included two mutually exclusive strata (cementless Austin Moore prostheses, cemented Thompson prostheses).

$$^{\dagger} \text{ Ratio of CR to EI} = \frac{\text{Number of competing risks observed}}{\text{Number of events of interest observed}}$$

Discussion

The rapidly increasing demand for joint replacements has placed growing importance on our ability to accurately monitor the cumulative incidence of revisions to assess implant quality, predict future demand for revisions, and inform clinical and health policy decisions [10, 26, 28]. Because the Kaplan-Meier method theoretically overestimates the cumulative incidence of events in the presence of competing risks, alternative competing-risks methods provide more accurate estimates of the cumulative incidence of revisions [18, 22]. However, competing-risks methods have yet to be widely reported within the orthopaedic literature and in joint replacement registries [29]. Our systematic review and meta-analysis aimed to determine the degree of overestimation of the Kaplan-Meier method compared with the competing-risks method when estimating the cumulative incidence of revision and to examine whether followup time and the rate of competing risks influenced this bias.

The articles included in our study conducted analyses of cohort and joint replacement registry data. Although randomized controlled trials are considered the highest level of evidence, registries have recently gained recognition as credible data sources [13, 19, 32, 37]. However, our assessment of the quality of reporting of these studies identified deficiencies similar to those previously identified in a review of survival analyses [3]. For example, only 43% of studies included in our review reported the number of patients that were at risk of revision at each followup time,

estimates of precision (such as SEs or CIs), or the statistical software used. Only three of the nine quality of reporting criteria assessed were fulfilled by all studies included in our review, reflecting the need for adherence to and strict enforcement of guidelines, perhaps through the development of a checklist, to improve the standards of reporting of survival analyses. However, it is important to note that, given that the goal of the studies included in our review was to summarize differences between the Kaplan-Meier and competing-risks methods, several studies did not conduct a full survival analysis using original data. Therefore, our assessment may underestimate the quality of reporting. Furthermore, given that our findings are based on a small number of studies (n = 7), caution is needed in interpreting these results. Nevertheless, clear guidance on the reporting of survival analyses is needed, specifically to address complications that arise in the analysis of competing-risks data. For example, reporting the number of patients at risk of revision becomes ambiguous in competing-risks situations as a result of differences in the censoring procedures between the Kaplan-Meier and competing-risks methods. Although the Kaplan-Meier method censors and removes patients from the risk set at their time of death, the competing-risks method includes patients who die in the risk set for the remainder of the observation period.

Individual patient data are considered the “gold standard” for meta-analyzing survival data [8, 39, 44]. Thus, the use of summary data is a limitation of our study. As a result of a lack of individual patient data, we were unable to examine factors that may have impacted the magnitude

of overestimation such as patient age and comorbidities. Furthermore, we were unable to derive a variance estimate for our effect measure. We therefore used a variance approximation primarily for the purpose of assigning weights to the individual studies in our meta-analysis. Because this approximation does not take into account the covariance between the Kaplan-Meier and competing-risks estimators (which are correlated given that both estimators are calculated using the same data), it likely overestimated the variance and width of the associated CIs around the RR estimates. This overestimation reduces the chance of observing statistically significant differences between the Kaplan-Meier and competing-risks estimators, resulting in a conservative estimate of our findings. For instance, the 95% CI around the RR of 1.30 obtained for Fennema and Lubsen [16] ranges from 0.63 to 2.72. The lower bound of this CI suggests that the Kaplan-Meier estimator may be less than the competing-risks estimator. This is mathematically incorrect given that the KM estimator must always be greater than or equal to the competing-risks estimator. Thus, the CIs around individual study and our pooled RR estimates must be interpreted with caution.

A lack of standardized analysis and reporting of revision rates within the arthroplasty literature and among joint replacement registries currently limits our ability to accurately monitor and compare outcomes across patient populations [29, 33]. Although registries have begun reporting the cumulative incidence of revision rather than person-time incidence rates [5], given the former provides more information regarding how the risk of revision changes over time, the International Society of Arthroplasty Registries has recently called for improvements in the standardization of survival analysis methods used to estimate these measures [20]. Because the choice of method depends on the study objective and audience (such as health policy planner versus patient perspective), establishing detailed guidelines for the approach to survival analysis may help address these issues. However, because questions have been raised regarding whether the difference between the Kaplan-Meier and competing-risks methods is clinically significant [17, 38], there is first a need for consensus among experts regarding the appropriateness of these methods.

Our study provides evidence that the Kaplan-Meier method overestimates the cumulative incidence of revision after hip or knee arthroplasty compared with the competing-risks method. The overestimation observed in individual studies ranged from 2% to 79% and in aggregate was approximately 55%. The magnitude of this bias is consistent with what has been observed across several other clinical areas when estimating the time to an event in the presence of competing risks [2, 4, 15, 40, 42, 46]. An alternative approach to exploring the magnitude of bias of

the Kaplan-Meier method in the setting of competing risks that might be considered would be to compare the original Kaplan-Meier estimate with an estimate obtained when all patients who experienced a competing event (such as death) before the event of interest (like revision) are presumed to have infinite followup. These patients would therefore be included in the risk set for the remainder of the followup period after their death, which is similar to how patient deaths are handled using the competing-risks method.

In general, we did not find the rate of competing risks or the duration of followup to influence the degree of overestimation. This may be the result of the small number of studies included in our meta-analysis or the conservative variance approximation that likely biased our results toward showing nonsignificant differences. It should be noted that the duration of followup directly influences the rate of competing events given that studies that follow patients over a longer period of time are more likely to follow patients until death rather than administrative censoring (that is, being unrevised at the end of the study period). However, based on the RR point estimates obtained for our stratified analyses, we speculate that the incidence of competing risks has a greater influence on the overestimation of the Kaplan-Meier method compared with the length of followup. For instance, in our analysis of strata containing the highest ratios of competing risks to revisions, the RR point estimate indicated the overestimation of the Kaplan-Meier method was greater for followup times less than 10 years compared with 10 or more years. This may be the result of the relatively high incidence of competing risks for the Gillam et al. [17] stratum compared with the other strata, despite a relatively shorter followup.

Overestimation of the Kaplan-Meier method may have important implications when monitoring the incidence of revision after hip or knee arthroplasty and is particularly concerning when using these estimates to inform health-care planning and policy decisions. Although our study provides strong support for increased use of competing-risks methods to more accurately estimate the absolute risk of revision, further investigation into factors influencing the overestimation of the Kaplan-Meier method such as the rate of competing events is required to better understand in which circumstances the bias of the Kaplan-Meier method becomes significant. We agree with the recommendations that *Clinical Orthopaedics and Related Research*® made in their editorial earlier this year on this topic [45], which suggest the use of competing-risks estimators when competing risks (such as death) might preclude the occurrence of important events of interest (such as revision surgery). Going forward, we urge journals to develop and encourage improved survival analysis guidelines to ensure appropriate methods are applied to produce unbiased estimates of the

risk of revision that can be used to monitor the safety of joint replacements and deliver relevant information to patients, clinicians, and health policymakers.

Acknowledgments We thank Diane Lorenzetti, research librarian in the Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada, and the Institute of Health Economics, Edmonton, Alberta, Canada, for her assistance in the development of our literature search strategy.

References

1. Abaira V, Muriel A, Emparanza JI, Pijoan JI, Royuela A, Plana MN, Cano A, Urreta I, Zamora J. Reporting quality of survival analyses in medical journals still needs improvement. A minimal requirements proposal. *J Clin Epidemiol*. 2013;66:1340–1346.e5.
2. Alberti C, Metivier F, Landais P, Thervet E, Legendre C, Chevret S. Improving estimates of event incidence over time in populations exposed to other events—application to three large databases. *J Clin Epidemiol*. 2003;56:536–545.
3. Altman DG, De Stavola BL, Love SB, Stepniowska KA. Review of survival analyses published in cancer journals. *Br J Cancer*. 1995;72:511–518.
4. Andersen PK, Geskus RB, de Witte T, Putter H. Competing risks in epidemiology: possibilities and pitfalls. *Int J Epidemiol*. 2012;41:861–870.
5. Australian Orthopaedic Association National Joint Replacement Registry. Annual Report 2013. Adelaide, Australia: AOA; 2013. Available at: <https://aoanjrr.dmac.adelaide.edu.au/annual-reports-2013>. Accessed February 4, 2014.
6. Biau DJ, Hamadouche M. Estimating implant survival in the presence of competing risks. *Int Orthop*. 2011;35:151–155.
7. Biau DJ, Latouche A, Porcher R. Competing events influence estimated survival probability—when is Kaplan-Meier analysis appropriate? *Clin Orthop Relat Res*. 2007;462:229–233.
8. Chalmers I. The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Ann N Y Acad Sci*. 1993;703:156–163; discussion 163–165.
9. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. *Br J Cancer*. 2003;89:232–238.
10. Cram P, Lu X, Kates SL, Singh JA, Li Y, Wolf BR. Total knee arthroplasty volume, utilization, and outcomes among Medicare beneficiaries, 1991–2010. *JAMA*. 2012;308:1227–1236.
11. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177–188.
12. Dignam JJ, Kocherginsky MN. Choice and interpretation of statistical tests used when competing risks are present. *J Clin Oncol*. 2008;26:4027–4034.
13. Dreyer NA, Garner S. Registries for robust evidence. *JAMA*. 2009;302:790–791.
14. Egger M, Davey Smith G, Altman DG, eds. *Systematic Reviews in Health Care: Meta-analysis in Context*. London, UK: BMJ Publishing Group; 2001.
15. Evans DW, Ryckelynck JP, Fabre E, Verger C. Peritonitis-free survival in peritoneal dialysis: an update taking competing risks into account. *Nephrol Dial Transplant*. 2010;25:2315–22.
16. Fennema P, Lubsen J. Survival analysis in total joint replacement: an alternative method of accounting for the presence of competing risk. *J Bone Joint Surg Br*. 2010;92:701–706.
17. Gillam MH, Ryan P, Graves SE, Miller LN, de Steiger RN, Salter A. Competing risks survival analysis applied to data from the Australian Orthopaedic Association National Joint Replacement Registry. *Acta Orthop*. 2010;81:548–555.
18. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med*. 1999;18:695–706.
19. Graves S. The value of arthroplasty registry data. *Acta Orthop*. 2010;81:8–9.
20. International Society of Arthroplasty Registries. Bylaws (revised March 2013). Available at: <http://www.isarhome.org/statements>. Accessed February 2, 2014.
21. Hamadouche M, Boutin P, Daussange J, Bolander ME, Sedel L. Alumina-on-alumina total hip arthroplasty: a minimum 18.5-year follow-up study. *J Bone Joint Surg Am*. 2002;84:69–77.
22. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York, NY, USA: John Wiley; 1980.
23. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53:457–481.
24. Keurentjes JC, Fiocco M, Schreurs BW, Pijls BG, Nouta KA, Nelissen RG. Revision surgery is overestimated in hip replacement. *Bone Joint Res*. 2012;1:258–262.
25. Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: irrelevance or ignorance? *Stat Med*. 2012;31:1089–1097.
26. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J Bone Joint Surg Am*. 2007;89:780–785.
27. Kurtz SM, Lau E, Ong K, Zhao K, Kelly M, Bozic KJ. Future young patient demand for primary and revision joint replacement: national projections from 2010 to 2030. *Clin Orthop Relat Res*. 2009;467:2606–2612.
28. Kurtz SM, Ong KL, Schmier J, Mowat F, Saleh K, Dybvik E, Karrholm J, Garellick G, Havelin LI, Furnes O, Malchau H, Lau E. Future clinical and economic impact of revision total hip and knee arthroplasty. *J Bone Joint Surg Am*. 2007;89(Suppl 3):144–151.
29. Lacy S, Bohm E, Hawker G, Powell J, Marshall DA. Disjointed? Assessing the comparability of hip replacement registries to improve the monitoring of outcomes. *Osteoarthritis Cartilage*. 2014;22:S213–S214.
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
31. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol*. 2009;170:244–56.
32. Maloney WJ. The role of orthopaedic device registries in improving patient outcomes. *J Bone Joint Surg Am*. 2011;93:2241.
33. Marshall DA, Pykerman K, Werle J, Lorenzetti D, Wasylak T, Noseworthy T, Dick DA, O'Connor G, Sundaram A, Heintzbergen S, Frank C. Hip resurfacing versus total hip arthroplasty: a systematic review comparing standardized outcomes. *Clin Orthop Relat Res*. 2014;472:2217–2230.
34. Pepe MS, Mori M. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Stat Med*. 1993;12:737–751.
35. Pfirrmann M, Hochhaus A, Lausaker M, Saeuele S, Hehlmann R, Hasford J. Recommendations to meet statistical challenges arising from endpoints beyond overall survival in clinical trials on chronic myeloid leukemia. *Leukemia*. 2011;25:1433–1438.
36. Pintilie M. *Competing Risks: A Practical Perspective*. West Sussex, UK: John Wiley & Sons Ltd; 2006.

37. Pivec R, Johnson AJ, Mears SC, Mont MA. Hip arthroplasty. *Lancet*. 2012;9855:1768–1777.
38. Ranstam J, Karrholm J, Pulkkinen P, Makela K, Espehaug B, Pedersen AB, Mehnert F, Furnes O, NARA Study Group. Statistical analysis of arthroplasty data. II. Guidelines. *Acta Orthop*. 2011;82:258–267.
39. Sargent DJ. A general framework for random effects survival analysis in the Cox proportional hazards setting. *Biometrics*. 1998;54:1486–1497.
40. Schuh R, Kaider A, Windhager R, Funovics PT. Does competing risk analysis give useful information about endoprosthetic survival in extremity osteosarcoma? *Clin Orthop Relat Res*. 2014 May 28 [Epub ahead of print].
41. Schwarzer G, Schumacher M, Maurer TB, Ochsner PE. Statistical analysis of failure times in total joint replacement. *J Clin Epidemiol*. 2001;54:997–1003.
42. Southern DA, Faris PD, Brant R, Galbraith PD, Norris CM, Knudtson ML, Ghali WA. Kaplan-Meier methods yielded misleading results in competing risk scenarios. *J Clin Epidemiol*. 2006;59:1110–1114.
43. Williamson PR, Smith CT, Hutton JL, Marson AG. Aggregate data meta-analysis with time-to-event outcomes. *Stat Med*. 2002;21:3337–3351.
44. Williamson PR, Smith CT, Sander JW, Marson AG. Importance of competing risks in the analysis of anti-epileptic drug failure. *Trials*. 2007;8:12.
45. Wongworawat MD, Dobbs MB, Gebhardt MC, Gioe TJ, Leopold SS, Manner PA, Rinnac CM, Porcher R. Editorial: Estimating survivorship in the face of competing risks. *Clin Orthop Relat Res*. 2015 Feb 11 [Epub ahead of print].
46. Yan Y, Moore RD, Hoover DR. Competing risk adjustment reduces overestimation of opportunistic infection rates in AIDS. *J Clin Epidemiol*. 2000;53:817–822.