Kaplan-Meier and Cox Regression Are Preferable for the Analysis of Time to Revision of Joint Arthroplasty

Thirty-One Years of Follow-up for Cemented and Uncemented THAs Inserted From 1987 to 2000 in the Norwegian Arthroplasty Register

Stein Atle Lie, PhD, Anne Marie Fenstad, MSc, Stein Håkon L. Lygre, PhD, Gard Kroken, MSc, Eva Dybvik, PhD, Jan-Erik Gjertsen, PhD, Geir Hallan, PhD, Håvard Dale, PhD, and Ove Furnes, PhD

Investigation performed at the Norwegian Arthroplasty Register, Department of Orthopedic Surgery, Haukeland University Hospital, Bergen, Norway

Background: Previous studies have suggested that the probability function of 1 minus the Kaplan-Meier survivorship overestimates revision rates of implants and that patient death should be included in estimates as a competing risk factor. The present study aims to demonstrate that this line of thinking is incorrect and is a misunderstanding of both the Kaplan-Meier method and competing risks.

Methods: This study demonstrated the differences, misunderstandings, and interpretations of classical, competing-risk, and illness-death models with the use of data from the Norwegian Arthroplasty Register for 15,734 cemented and 7,867 uncemented total hip arthroplasties (THAs) performed from 1987 to 2000, with fixation as the exposure variable.

Results: The mean age was higher for patients who underwent cemented (72 years) versus uncemented THA (53 years); as such, a greater proportion of patients who underwent cemented THA had died during the time of the study (47% compared with 29%). The risk of revision at 20 years was 18% for cemented and 42% for uncemented THAs. The cumulative incidence function at 20 years was 11% for cemented and 36% for uncemented THAs. The prevalence of revision at 20 years was 6% for cemented and 31% for uncemented THAs.

Conclusions: Adding death as a competing risk will always attenuate the probability of revision and does not correct for dependency between patient death and THA revision. Adjustment for age and sex almost eliminated differences in risk estimates between the different regression models. In the analysis of time until revision of joint replacements, classical survival analyses are appropriate and should be advocated.

Level of Evidence: Prognostic Level III. See Instructions for Authors for a complete description of levels of evidence.

Consider the situation that 1 surgeon performs total hip arthroplasties (THAs) at 2 different hospitals. Both hospitals utilize the same implant, the procedure is performed with the same operating team, and all other conditions are identical; the only difference between settings is that 1 hospital has a patient population with a high mortality rate and the other has patients with low mortality. The quality of the THAs performed is equal, which should be reflected in the statistical analysis. However, because patients at the high-mortality hospital may die before revision is needed, the low-mortality hospital will experience—and need to plan for—more revisions than the high-mortality hospital, even if the quality of the THAs is equal.

Because THAs are followed over time until the implant is revised or lost to follow-up, survival (time-to-event, event history) analyses are needed to analyze the risk of revision.

Competing-risk analyses have been increasingly popular for the assessment of joint-replacement data, and some studies have advocated the use of such analyses for follow-ups of >10 years.

For time-to-revision data, some authors have claimed that the probability function of 1 minus the Kaplan-Meier survivorship overestimates the rate of revision in the presence of patient death.

Recently, there have been discussions regarding whether patient death is a competing risk relative to THA revision and...
Kaplan-Meier and Cox Regression Are Preferable for the Analysis of THA Time to Revision

Materials and Methods

To illustrate distinctions between statistical approaches, we used a data set from the Norwegian Arthroplasty Register, which started registration of THAs in September 1987, with a completeness of up to 97%\(^{15}\). Based on the Norwegian personal identification number, we linked primary and revision operations, with complete data on death or emigration\(^{16}\). Our primary exposure was 2 categories of fixation (i.e., cemented and uncemented for both components). We limited data to primary THAs implanted between 1987 and 2000. During that period, several poorly performing uncemented THAs\(^{16,17}\) and inferior cement types\(^{18}\) were utilized in Norway. Uncemented THAs were commonly utilized in younger patients and cemented THAs in elderly patients. Hence, patients with cemented THAs had higher mortality than those with uncemented THAs, which influences the results from the different statistical approaches. All 7,867 uncemented THAs were included, whereas a random subset of 15,734 of 112,873 cemented THAs, twice the number of uncemented procedures, were included.

The end of follow-up was December 31, 2018. Only 0.2% of patients emigrated, and these were considered censored for all models.

Statistical Methods

Classical Survival Analysis

In classical models of time to revision (Fig. 1), the hazard is defined as the immediate risk of revision just after time \(t\), given that the implant is at risk at time \(t\). In Figure 1, the hazard \(h_{01}(t)\) is indicated as an arrow between state 0 and state 1, specifying the immediate risk of shifting states. For the classical survival model (Fig. 1), the Kaplan-Meier method\(^{19}\) estimates the cumulative probability of being in state 0. One minus the Kaplan-Meier will hence estimate the cumulative probability for state 1, and there is a 1-to-1 relationship between the hazard and the probabilities of being in each of the 2 states\(^{20}\), assuming that time to revision and censoring are independent. Thus, analyses of hazards (e.g., Cox regression\(^{21}\)) will be analyses of probabilities\(^{22}\). Importantly, the sum of the 2 probabilities will be 1 (100%). In the classical survival model (Fig. 1), loss of the patient to follow-up (including death or emigration) or end of follow-up in 2018 will lead to censoring of the follow-up time for the THA.

Competing Risks

In the competing-risk model (Fig. 2), patient death is included as a separate state. Hence, there are now 2 mutually exclusive hazards \((h_{01}(t)\) and \(h_{02}(t)\)) simultaneously influencing the probabilities for being in each of the 3 states (0, 1, or 2). The survival probability (i.e., the probability of state 0) can directly be calculated (applying the Kaplan-Meier method) based on the sum of the 2 mutually exclusive hazards \((h_{\text{sum}}(t) = h_{01}(t) + h_{02}(t))\). The cumulative probability of being in the 2 competing states (i.e., the cumulative incidence functions) is based on each of the hazards and the total survival\(^{23}\). For the competing-risk model, analysis of each hazard (e.g., using Cox regression) is still valid, but does not reflect the probability of being in the states.
In the competing-risk model, calculation of cumulative incidence functions can be performed with use of the Aalen-Johansen estimator or via direct methods. The Aalen-Johansen method is based on the increments for the cumulative hazard (e.g., Nelson-Aalen method). Fine and Gray suggested a regression model for the cumulative incidence functions (i.e., the likelihood of being in the state of interest; e.g., state 1), expressed as sub-hazard ratios, with the possibility to adjust for covariates.

Illness-Deaths Models
Extending the competing risk model with a hazard (h_{12}(t)) for changing from revision THA (state 1) to patient death (state 2), we have an illness-death model (Fig. 3). Still, each hazard can be analyzed separately. Hence, Cox regression models will still be valid for the shifts between states but will not reflect probabilities for being in any of the states.

The Use of Pseudo-Data
With use of pseudo-data, a variety of analyses of classical survival, competing-risk, illness-death, and elaborate multistate models can be performed. The calculation of pseudo-data is described elsewhere. Pseudo-data can be expressed as several quantities, such as probabilities of being in different states or restricted mean duration in a given state (e.g., restricted mean survival or restricted mean failure time). Pseudo-data are typically analyzed with use of generalized estimating equations with an appropriate link-function and distribution.

In the present study, pseudo-data were utilized to calculate differences in probabilities (hazard scale, using a complementary log-log link function) and restricted mean differences in time spent (using an identity link function) in state 1 (restricted mean failure times). Pseudo-data and probability curves were calculated with use of the R statistical framework (R Foundation for Statistical Computing). Regression models were analyzed with use of Stata (version 16; StataCorp). Significance was set at 0.05.

Source of Funding
There was no external funding for this study.

Results
The mean age at the time of the primary procedure was higher for cemented THAs than for uncemented THAs, and the frequency of males was higher for uncemented THAs (Table I). The numbers of revisions and patient deaths for cemented and uncemented THAs are presented in Table II. For the classical survival model, follow-up was censored at the time of patient death.

The estimated probabilities, calculated with use of the Kaplan-Meier method, show that the cumulative risk of revision was less for cemented (18.1%; 95% confidence interval [CI], 15.2% to 21.0%) than for uncemented THAs (42.2%; 95% CI, 39.5% to 44.9%) at 20 years (Figs. 4-A and 4-B).

Including patient death as a competing risk, we observed that the probability (cumulative incidence function) of observing a patient with a revised THAs (10.9%; 95% CI, 8.5% to 13.2%) was more influenced by mortality among patients with cemented THAs than uncemented THAs (36.4%; 95% CI, 33.7% to 39.0%) at 20 years (Figs. 4-C and 4-D).

In the illness-death model, with a transition from revision THAs to patient death, the probability of being alive at 20 years with a revised cemented THA (5.6%; 95% CI, 3.9% to 7.3%) was substantially lower than that of patients with a revised uncemented THA (31.2%; 95% CI, 28.7% to 33.8%) (Figs. 4-E and 4-F).

Risk estimates comparing cemented THA with uncemented THA with use of the unadjusted Fine and Gray model were larger.
Kaplan-Meier and Cox Regression Are Preferable for the Analysis of THA Time to Revision

than estimates from the unadjusted Cox model because of the high mortality for patients with cemented THAs. These 2 estimates were more similar following adjustment for age and sex (Table III).

Classical and competing-risk models for the probability of revision THA based on pseudo-data corresponded to the results from the Cox and Fine and Gray methods, respectively (Table IV). Additionally, in the illness-death model, the difference in probability of a patient being alive with a revision (state 1) was even larger between patients with uncemented and cemented THAs than for the 2 former approaches (Table IV).

Using pseudo-data to analyze the restricted mean failure time based on 25 years of observation, we observed that uncemented THAs were revised at an average of 4.13 years earlier than cemented THAs (Table V). Including patient mortality, we observed that patients with uncemented THAs underwent revision at a mean of 4.31 years earlier than patients with cemented THAs. Furthermore, patients with uncemented THAs lived an average of 4.23 years longer following revision. Adjusting for age and sex, the different regressions performed with use of pseudo-data had similar results to each other (Tables IV and V).

**Discussion**

There is an important difference between the classical, competing-risk, and illness-death models. For the classical analyses, we emphasize that revision of the implant is analyzed. This is also highlighted in the naming of the 2 states in the Figure 1 legend: “Primary THA” and “Revised THA.” For the competing-risk and illness-death models, which include patient death as a competing risk relative to THA revision will always attenuate the probability of THA revision and consequently does not correct for dependency between time to death and time to revision. This was clearly illustrated by Sayers et al. The sum of probabilities for the different states will always equal 1 (100%); therefore, adding a new state will always attenuate the probabilities for the other states. The assumption of independence between the times to revision, death, and censoring is still present in competing-risk analysis. Methods to adjust the Kaplan-Meier estimate for dependency exist, such as utilizing inverse probability weights to adjust for the censoring process (e.g., see page 175 in the book by Aalen et al.).

For analysis of patient mortality after joint replacement surgery, survival analysis is needed. Overall mortality is often of primary interest. If multiple causes of death are studied, competing-risk analysis is appropriate, and each cause of death (respiratory disorders, circulatory disorders, etc.) represents competing causes/risks. Hence, for analysis of 1 cause, considering the other causes as competing risks, using cumulative

| TABLE II Transitions Between Primary THA, Revision THA, and Death for the Cemented and Uncemented THAs* |

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cemented</td>
<td></td>
</tr>
<tr>
<td>Primary THA (state 0)</td>
<td>15,734</td>
</tr>
<tr>
<td>Revision THA (state 1)</td>
<td>1,167</td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>Uncemented</td>
<td></td>
</tr>
<tr>
<td>Primary THA (state 0)</td>
<td>7,867</td>
</tr>
<tr>
<td>Revision THA (state 1)</td>
<td>2,695</td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

*Values are given as the count with the percentage in parentheses.
incidence functions, will cause the mortality estimates to properly sum to the total patient mortality.

In contrast, utilizing patient death as a competing risk relative to revision of a primary THA, the sum of probabilities for the 2 competing risks (i.e., revision THA and patient mortality) is meaningless. Considering revision as a competing risk relative to patient death is also meaningless. This is not an argument against patient death as a competing risk relative to THA revision, but demonstrates that this analysis deviates from the original intention of competing-risk analysis.

Fig. 4
Estimated probabilities: for a classical survival model for cemented (A) and uncemented (B) THAs, for the states described in Figure 1; for a competing risk model for cemented (C) and uncemented (D) THAs, for the states described in Figure 2; and for an illness-death model for cemented (E) and uncemented (F) THAs, for the states described in Figure 3. Red = state 0 (primary THA or patient alive with primary THA), green = state 1 (revised THA or patient with revised THA), and blue = state 2 (dead patient).
Bilateral THA is another complicating issue. In the classical approach, Kaplan-Meier estimates can be adjusted to account for multiple implants per patient. Furthermore, multilevel approaches can be utilized to analyze multiple implants within a single patient (e.g., finger arthroplasties, dental implants, or dental fillings).

For the competing-risk approach, considering multiple or bilateral implants becomes problematic. Simple cumulative incidence functions or regression models, ignoring bilateral implants, are inappropriate because the risks of revision for the 2 THAs are related.

Including random effects in the analysis will lead to a wrong dependency between the competing risks (e.g., see page 260 in the book by Aalen et al.). The appropriateness of utilizing robust variance estimates in regression models or calculating cumulative incidence functions based on patient-averaged hazards is also debatable.

For bilateral THAs, each THA may experience patient death, resulting in the patient counted as dying twice on the same date. Such “double counting” results in a strong dependency between the bilateral THAs for the time to patient death. van der Pas et al. discovered that this problem can be exacerbated by consecutive implants, as these will always have a shorter time to patient death, resulting in unnecessary bias. This problem also increases when a patient has >2 implants (e.g., in the fingers). These examples represent multiple unnecessary problems related to competing-risk models that are not present in the classical methods.

The same problems as in competing-risk models are present for illness-death models. Additionally, if a patient dies with 1 revised and 1 intact primary THA, a problematic dependency occurs between the 2 hazards (shifts) to patient death.

To give an example: for patients with pacemakers, failure of the pacemaker may be 1 of several causes of death. Therefore, adding the time of pacemaker failure as a competing risk for patient death is sensible. However, for an analysis of the longevity of pacemakers, death of the patient (as a result of causes other than pacemaker failure) leads to a loss of follow-up (censoring) of the pacemaker (i.e., the pacemaker was still working at the time of patient death). For the pacemaker itself, an empty battery, loose cables, etc., may be competing risks of failure.

For some situations, competing-risk analyses may at first glance seem odd, but they may also hold additional value. Consider the lifetimes of infants with special needs. Death of the mother as a competing risk can be a sensible inclusion from a resource-planning perspective because extra resources may be required to care for the infant following the death of the parent.

Thus, analysis of revision THA with patient death as a competing risk can be useful from a resource-planning perspective because it helps to predict the future burden of

### Table III Cox and Fine and Gray Regression Models for Risk of Revision

<table>
<thead>
<tr>
<th></th>
<th>Cox Model for $h_{01}(t)$</th>
<th>Fine and Gray Model for State 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unadjusted</td>
<td>Adjusted†</td>
</tr>
<tr>
<td>Uncemented§</td>
<td>2.69 (2.57 to 2.81); &lt;0.001</td>
<td>1.56 (1.47 to 1.65); &lt;0.001</td>
</tr>
<tr>
<td>Age, per 5 yr</td>
<td>0.87 (0.87 to 0.88); &lt;0.001</td>
<td>1.46 (1.41 to 1.52); &lt;0.001</td>
</tr>
<tr>
<td>Male sex#</td>
<td>1.46 (1.41 to 1.52); &lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>

*Values are given as the hazard ratio with the 95% CI in parentheses, followed by the p value. †Adjusted for age and sex. §Compared with cemented. #Compared with female sex.

### Table IV Probability Models for Revision THA (State 1), Based on Pseudo-Data

<table>
<thead>
<tr>
<th></th>
<th>Classical Survival Model</th>
<th>Competing-Risk Model</th>
<th>Illness-Death Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unadjusted</td>
<td>Adjusted†</td>
<td>Unadjusted</td>
</tr>
<tr>
<td>Uncemented†</td>
<td>2.60 (2.42 to 2.78); &lt;0.001</td>
<td>1.67 (1.5 to 1.83); &lt;0.001</td>
<td>3.53 (3.31 to 3.76); &lt;0.001</td>
</tr>
<tr>
<td>Age, per 5 yr</td>
<td>0.88 (0.87 to 0.90); &lt;0.001</td>
<td>0.86 (0.85 to 0.87); &lt;0.001</td>
<td>0.85 (0.78 to 0.93); &lt;0.001</td>
</tr>
<tr>
<td>Male sex§</td>
<td>1.08 (0.99 to 1.17); 0.080</td>
<td>0.95 (0.88 to 1.03); 0.202</td>
<td>0.85 (0.78 to 0.93); &lt;0.001</td>
</tr>
</tbody>
</table>

*Values are given as the hazard ratio with the 95% CI in parentheses, followed by the p value. †Adjusted for age and sex. §Compared with cemented.
TABLE V Models for Differences in Restricted Mean Failure Time for Years Spent in the Revision THA State (State 1), Based on Pseudo-Data *

<table>
<thead>
<tr>
<th>Classical Survival Model</th>
<th>Competing-Risk Model</th>
<th>Illness-Death Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted</td>
<td>Adjusted†</td>
<td>Unadjusted</td>
</tr>
<tr>
<td>Uncemented‡</td>
<td>4.13 (3.84 to 4.42); &lt;0.001</td>
<td>4.31 (4.10 to 4.52); &lt;0.001</td>
</tr>
<tr>
<td>Age, per 5 yr</td>
<td>–0.45 (–0.52 to –0.39); &lt;0.001</td>
<td>–0.63 (–0.68 to –0.58); &lt;0.001</td>
</tr>
<tr>
<td>Male sex§</td>
<td>0.71 (0.42 to 1.00); &lt;0.001</td>
<td>0.17 (–0.04 to 0.38); 0.110</td>
</tr>
</tbody>
</table>

*Values are given as the beta coefficient (indicating the difference in time, in years) with the 95% CI in parentheses, followed by the p value. †Adjusted for age and sex. ‡Compared with cemented. §Compared with female sex.

revisions. However, when assessing implant survival and the quality of the THA procedure, competing-risk analysis cannot be compared with and interpreted as being equal to classical survival analysis. Presenting 1 minus the Kaplan-Meier estimate with cumulative incidence functions from such analyses is misleading and should be avoided. Patients alive with a THA revision may need special care, and the illness-death model will indicate the prevalence of such patients.

For the analyses in the present article, we showed that competing-risk and illness-death models sometimes have added value. However, these approaches do not circumvent assumptions from the classical approaches, but introduce additional problems not easily accounted for. Competing-risk and illness-death models are special cases of the multistate models described by Gillam et al., ignoring that bilateral THAs are not independent for both time to revision and patient death.

Statistical consultants guiding the analysis of implant survival should be aware of the important differences between models and of the pitfalls when patient death is added as a competing risk. In the case of THA survival, the Kaplan-Meier method is designed to account for loss to follow-up for any reason. If the patient dies, emigrates, withdraws from the study, is excluded because of a change in health condition, or is otherwise is lost from the study, the follow-up time of the THA will be censored. However, for the analysis of patient death and different causes of death, 1 minus the Kaplan-Meier estimate is incorrect for the probability of dying from a single cause. Likewise, if different causes of revision are analyzed, 1 minus the Kaplan-Meier estimate gives incorrect estimates for revision for each cause. Hence, for both these situations, the competing causes—of patient death or THA revision—must be considered. When analyzing the overall risk of revision following THA, the Kaplan-Meier method and Cox regression are appropriate and correct, and should therefore be advocated.

References

Kaplan-Meier and Cox Regression Are Preferable for the Analysis of THA Time to Revision

13. Van Der Pas S, Nelissen R, Ficocci M. Different competing risks models for different questions may give similar results in arthroplasty registers in the presence of few events. Acta Orthop. 2018 Apr;89(2):145-51.