Topics in mathematical and applied statistics

Pas, S.L. van der

6

Bilateral patients in arthroplasty registry data

6.1 Introduction

Worldwide more than 3 million total hip and knee arthroplasties are performed annually, and this number is predicted to increase substantially within the next decades (Pabinger and Geissler, 2014; Pabinger et al., 2015). Data on total joint arthroplasties (TJAs) are collected in a growing number of arthroplasty registries around the world, and the resulting data has proven to be valuable in improving the outcome of TJA (Graves, 2010).

This chapter is based on analyses of total hip arthroplasty (THA) data from the LROI (Landelijke Registratie Orthopedische Implantaten / Dutch Arthroplasty Register), which has been recording patient and implant characteristics of all hip and knee replacements in The Netherlands since its establishment in 2007. A large number of THAs is performed in The Netherlands each year; the LROI registered about 28.000 primary THRs in 2014. Osteoarthritis is the most common reason for THA: 87% of THAs were performed after a diagnosis of osteoarthritis (LROI, 2014).

Benefits of THA include improved mobility, increased hip joint functionality, and pain relief (Wilcock, 1978). A hip implant does not last forever however, and a patient may

This chapter contains material from two papers. The first has been submitted as: S.L. van der Pas, R.G.H.H. Nelissen and M. Fiocco. Staged bilateral total joint arthroplasty patients in registries. Immortal time bias and methodological options. The second is in preparation, with R.G.H.H. Nelissen, B.W. Schreurs and M. Fiocco, and titled 'Risk factors for early revision after unilateral and staged bilateral total hip replacement in the Dutch Arthroplasty Register'.

need to undergo revision surgery, which we define as any change to the implant. Revision places not only a burden on healthcare costs, but on the patient as well, and is associated with higher risk of adverse outcomes than the primary surgery (Mahomed et al., 2003; Ong et al., 2006). Incidence of revision has been linked to many demographic, clinical, surgical and health care provider related factors, including gender, age, race, body weight, American Society of Anesthesiologists (ASA) score, underlying diagnosis, type of fixation and hospital volume (Prokopetz et al., 2012). We investigate risk factors for revision within the first 8 years of follow-up.

Three methodological issues need to be taken into consideration during the statistical analysis. The first and second are due to the presence of (staged) bilateral patients in the data. With "bilateral patients", we refer to patients with two THAs, and we refer to patients with a THA on one side as "unilateral patients". The first issue is that each bilateral THA patient contributes two dependent observations, violating the independence assumptions underlying most methods. Secondly, the time that usually passes between two successive THAs renders a patient's bilaterality status time-dependent. The number of patients with bilateral THAs is not negligible; in The Netherlands 20% of total hip arthroplasty surgeries in 2014 concerned the placement of a second prosthesis, in Sweden 20.5% of patients became staged bilateral between 1992-2014, and in Norway, 23.6% of patients became bilateral within 10 years (Lie et al., 2004; LROI, 2014; SHAR, 2014).

The third issue is that a patient may die before experiencing revision of the implant. If this competing risk of death is not appropriately accounted for, the risk of revision surgery will be overestimated (Keurentjes et al., 2012; Ranstam et al., 2011).

Although this chapter is written in the context of total hip replacement, the considerations and results are relevant to registry data of any body part of which a human has at least two, such as knees, ankles, shoulders, eyes, fingers and teeth.

The structure of this chapter is as follows. Methods for handling the competing risk of death are briefly reviewed in Section 6.2. The complications stemming from the bilateral patients are discussed in Section 6.3. The data structure is then introduced in Section 6.4. This Chapter concludes with preliminary results on the LROI data in Section 6.5.

6.2 Competing risk of death

THA is most commonly done in elderly patients; the average age of the patients in the hip replacement data set is 69 years. A patient may die before experiencing revision. Indeed, out of the 161,434 hips in the data set 3,897 hips were revised, while it was not possible to observe revision for 7,179 hips due to death of the patient. Thus, death should be considered a competing risk. Estimating the probability of revision by Kaplan-Meier would be inappropriate, as it is designed for a single outcome (in this case, revision), which is possibly not observed due to censoring. Deaths are treated as censored observations and not as events. However, considering deceased patients as censored observations violates the independence of the censoring distribution assumption underlying Kaplan-Meier (Putter et al., 2007). By the independent censoring assumption, a dead patient would have the same hazard of revision as a patient who is still alive and has not experienced revision yet. Since Kaplan-Meier treats dead patients as if they could still experience revision, the probability of revision is overestimated.

In a competing risk setting, the functions of interest are the cumulative incidence functions. The cumulative incidence of cause k at time t is the probability that failure due to cause k has occurred by time t. There are methods available to estimate the cumulative incidence of any event in the competing risks setting. We consider three of them, and first introduce some notation.

We assume right-censored data. We have n observations, and each observation i has failure time T_i and censoring time C_i associated to it. Define $X_i = \min\{T_i, C_i\}$, and $\Delta_i = \mathbf{1}\{T_i \leq C_i\}$ and let $\varepsilon_i \in \{1,\ldots,K\}$ be the causes of failure, for $i = 1,\ldots,n$. Let Z_i be a $p \times 1$ bounded and time-independent covariate vector. We assume that $(X_i, \Delta_i, \Delta_i\varepsilon_i, Z_i)$ are independent and identically distributed for $i = 1,\ldots,n$. Denote the observed, distinct event times by $t_1 < t_2 < \ldots < t_m$.

With this notation, the cumulative incidence of cause k is given by:

$$F_k(t) = Pr(T \leq t, \varepsilon = k), \quad k = 1,\ldots,K.$$

A cumulative incidence function is determined by the cause-specific hazards $\lambda_k(t)$, $k = 1,\ldots,K$. The cause-specific hazard is the hazard of failing from cause $k \in \{1,\ldots,K\}$, which is in competition with the other failure causes. It is defined as

$$\lambda_k(t) = \lim_{\Delta t \downarrow 0} \frac{Pr(t \leq T \leq t + \Delta t, \varepsilon = k \mid T \geq t)}{\Delta t}.$$

The cumulative incidence can be expressed in terms of the cause-specific hazards as follows:

$$F_k(t) = \int_0^t S(u) d\Lambda_k(u), \quad k = 1,\ldots,K, \tag{6.1}$$

where $S(t) = \exp(-\sum_{k=1}^K \Lambda_k(t))$ is the overall survival function, and $\Lambda_k(t) = \int_0^t \lambda_k(u)du$ is the cumulative cause-specific hazard. In the following sections, we briefly review three methods for estimating the cumulative incidence: the Aalen-Johansen estimator, cause-specific Cox regression and Fine-Gray regression.

Aalen-Johansen estimator

The unadjusted cumulative incidence can be estimated by the Aalen-Johansen estimator (Aalen and Johansen, 1978), which was defined for more general multi-state models, but in this case reduces to (6.1) with the left-continuous Kaplan-Meier estimate for the survival function, and the Nelson-Aalen estimators for the cumulative cause-specific hazards. Denote the number of failures due to cause k at time t_i by $d_k(t_i) = \sum_{i=1}^n \mathbf{1}\{X_i = t_i, \varepsilon_i = k\}$ and the number still at risk just before time t_i by $n(t_i) = \sum_{i=1}^n \mathbf{1}\{X_i \geq t_i\}$.

The Nelson-Aalen estimators and Kaplan-Meier estimator are given by

$$\widehat{\Lambda}_k(t) = \sum_{t_i \leq t} \frac{d_k(t_i)}{n(t_i)}, \quad \widehat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{\sum_{k=1}^K d_k(t_i)}{n(t_i)}\right),$$

and $F_k(t)$ is estimated by $\widehat{F_k}(t) = \sum_{t_i \leq t} S(t_{i-1}) d_k(t_i)/n(t_i)$.

Cause-specific Cox regression

The Cox proportional hazards model allows a natural extension to the competing risks settings, where the cause-specific hazard for individual i and cause k is modelled as (Holt, 1978):

$$\lambda_k(t; z_i) = \lambda_{0,k}(t) e^{\beta_k^T z_i}. \tag{6.2}$$

Here, $\lambda_{0,k}(t)$ is a cause-specific baseline hazard. All cause-specific hazards are estimated separately and then combined to assess the association of the covariates to the cumulative incidence of the cause of interest. Each cause-specific hazard $\lambda_k(t; z_0)$ is estimated by censoring all individuals who failed due to a cause other than k. At each time at which an individual experiences failure due to cause k, the covariate values of this individual are compared with the covariates of all other individuals who are still event-free and in follow-up. Following Cheng et al. (1998), the cumulative incidence is estimated by plugging in the maximum partial likelihood estimate $\widehat{\beta}_k$ for β_k and the Breslow estimate $\widehat{\Lambda}_{0,k}(t)$ for the cumulative hazard:

$$\widehat{F}_k(t; z_0) = \int_0^t \widehat{S}(u; z_0) d\widehat{\Lambda}_k(u; z_0),$$

where $\widehat{S}(u; z_0) = \exp(-\sum_{k=1}^K \widehat{\Lambda}_k(u; z_0))$ and $\widehat{\Lambda}_k(u; z_0) = \widehat{\Lambda}_{0,k}(u) \exp(\widehat{\beta}_k^T z_0))$.

Fine-Gray regression

Fine-Gray regression (Fine and Gray, 1999) is a Cox model like (6.2), but for the subdistribution hazard $h_k(t; z_0)$ instead of the cause-specific hazard. The subdistribution hazard is the instantaneous risk of failing from cause k given that the individual has not failed from cause k:

$$h_k(t; Z_i) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr(t \leq T \leq t + \Delta t, \varepsilon = k \mid T \geq t \cup (T \leq t \cap \varepsilon \neq k), Z_i).$$

Fine-Gray regression is designed to model only one subdistribution hazard at the same time (Beyersmann et al., 2012, Section 5.3.4). The model is given by:

$$h_k(t; z_i) = h_{0,k}(t) e^{\beta_k^T z_i}, \tag{6.3}$$

where $h_{0,k}(t)$ is a subdistribution baseline hazard and k refers to the single cause of failure under consideration. An appealing property of the subdistribution hazard is that it satisfies

$$F_k(t; z_0) = 1 - e^{-\int_0^t h_k(u; z_0) du}. \tag{6.4}$$

The model (6.3) for the subdistribution hazard thus allows direct assessment of the relationship between a covariate and the cumulative incidence of the cause of interest. The risk set corresponding to the subdistribution hazard is counterintuitive however, as it contains those individuals who have already failed from a cause different than k, and are thus not able to fail from cause k anymore (Fine and Gray, 1999). The coefficients β in (6.3) are estimated by a weighted partial likelihood approach and the cumulative subdistribution baseline hazard is estimated by a Breslow-type estimator. A test for equality of cause-specific cumulative incidence functions is available (Gray, 1988).

Choice of method and model checking

Usually, only either cause-specific Cox regression or Fine-Gray regression is selected for adjusted analyses. This seems natural, as they estimate different quantities, although both can be used to estimate cumulative incidences. There is a relationship between the subdistribution hazard $h_k(t; z_0)$ and cause-specific hazard $\lambda_k(t; z_0)$, which follows from combining (6.1), (6.4) and differentiating with respect to t (Beyersmann and Scheike, 2013; Beyersmann and Schumacher, 2007):

$$h_k(t; z_0) = \frac{S(t; z_0)}{1 - F_k(t; z_0)} \lambda_k(t; z_0).$$

If the proportionality assumption holds for one of the hazards, the other model will thus typically be misspecified (Beyersmann and Schumacher, 2007; Latouche et al., 2013, 2007). Grambauer et al. (2010) found that the subdistribution hazards and cause-specific hazards for cause 1 are numerically quite close if a covariate has no effect on the remaining cause-specific hazards, or when there is heavy censoring.

Cause-specific Cox regression provides insight into the relationship of covariates on the hazard of, in this case, revision or death. Fine-Gray regression yields in a sense a summary, indicating the association between a covariate and the cumulative incidence of revision. Grambauer et al. (2010) and Latouche et al. (2013) recommend presenting the results from both the cause-specific Cox model and Fine-Gray regression side by side, for all causes. In any case, it is prudent to report results on model fit. There are several options available. Three aspects of the models are evaluated (Lin et al., 1993):

1. The proportional hazards assumption;
2. The functional forms of covariates in the exponent of the model;
3. The link function.

An overview of diagnostic tests for the Cox model is given in Chapter 11 of Klein and Moeschberger (2003), while Li et al. (2015) discuss a number of tests for each of the three aspects listed above for Fine-Gray regression. In addition, Andersen and Pohar Perme (2010) review methods for assessing goodness-of-fit using pseudo-values, which can be applied to the Cox model and the Fine-Gray model. Omnibus tests for all three aspects are available for the Cox model (Lin et al., 1993; McKeague et al., 2001) and the Fine-Gray model (Li et al., 2015), and an R package for the latter is under development (Li et al., 2015).

6.3 Dependence between hips and the time-dependent bilateral status

Approximately 20% of THAs undertaken in The Netherlands concern the placement of a second hip implant (LROI, 2014). Thus, the LROI data contains a sizable proportion of bilateral patients. Both hips can be placed simultaneously, but more commonly, the interoperative time is several months or years. In the latter case, the patients are referred to as "staged bilateral patients". Their presence poses a problem to the statistical analysis of

arthroplasty data, as has been recognized in the orthopaedic literature (Bryant et al., 2006; Lie et al., 2004; Ranstam et al., 2011). The focus of those papers has been on the dependence of the two observations contributed by a bilateral patient. There is little recognition for a second problem however, namely that a patient's bilaterality status is time-dependent.

We review some methods for handling the dependence between two hip implants within a patient in Section 6.3.1, and discuss methods which incorporate the time-dependent status in Section 6.3.2. We give some remarks on practical relevance in Section 6.3.3.

6.3.1 Methods for dependent observations

Most of the methods proposed in the orthopaedic literature only account for the dependence between two hips within one patient, and do not consider the time between two successive THAs. The interoperative time is not relevant for patients undergoing simultaneous bilateral hip replacement ("same-day bilateral patients"), and those are the patients we will have in mind in this section. We now review the methods that have been proposed in the orthopaedic literature. These methods are intended to be used in combination with the competing risks methods discussed in Section 6.2.

Subgroup analysis

One recommendation by Bryant et al. (2006) is to analyse patients with bilateral THA as a separate subgroup, which is done occasionaly in practice, or the bilateral patients are excluded altogether (Buchholz et al., 1985; Gillam et al., 2010; SHAR, 2014; Visuri et al., 2002). The unilateral observations will all be independent, but the bilateral patients' observations are still dependent, so the dependence issue is not completely resolved by subgroup analysis. In addition, when subgroup analysis is done with staged bilateral patients, the analysis is at risk of being affected by immortal time bias, as will be explained in Section 6.3.2.

Excluding the second joint

Bryant et al. (2006) suggested excluding the second joint, and this option is used in practice (Maurer et al., 2001; Morris, 1993; NJR, 2015). Only using each patient's first THA ensures independence of the observations used in the analysis. A disadvantage is that not all data is used, although this may not be a serious problem in arthroplasty registry studies where the amount of data can run into the hundred thousands. At first glance, a second drawback may be that the conclusions only hold for a patient's first THA and not the second, but this may actually be sensible given that the outcomes for the second implant may be different compared to the first implant.

Selecting a random joint

A third suggestion by Bryant et al. (2006) is to select a random hip for each bilateral patient, and this was previously implemented by Visuri et al. (2002). The analysis is carried out using all unilateral observations, and one randomly selected observation from each bilateral patient. In this way, alle observations in the sample are independent. However,

this raises other issues. The first problem is that the sensitivity of the results to the particular sample should be assessed. The second is that it is unclear what is being estimated. If the outcomes of a patient's first and second THA are different, the interpretation of the estimate resulting from this procedure is difficult.

Resampling techniques

Closely related to the selection of a random joint per patient is the idea of within-cluster resampling. Each patient is viewed as a cluster, containing either one or two THAs. Ranstam et al. (2011) suggested to apply the methodology of Hoffman et al. (2001), which is valid for data with clusters of nonignorable size, meaning that the risk for the outcome is related to the cluster size. For within-cluster resampling, a large number of data sets is created by randomly selecting one observation per individual. The estimator is computed on each data set, after which all estimates are averaged, resulting in the within-cluster resampling estimator.

Hoffman et al. (2001) prove asymptotic normality in the context of generalized linear models, and the main proof concept can be adapted to the competing risks setting, when combined with results in Cheng et al. (1998); Fine and Gray (1999) and Lin (1997). This extension would require assuming that both hips follow the same model, which seems unlikely to be true. The resulting estimator would represent the cumulative incidence of revision for a randomly sampled hip from a randomly sampled patient, and again, it is not clear how meaningful this would be in practice.

The within-cluster resampling procedure is reminiscent of the block or cluster bootstrap, but these methods differ in execution and aim. Suppose we have observed C clusters. The resampled datasets of the cluster bootstrap arise by sampling C clusters with replacement (Davison and Hinkley, 1997), while for within-cluster resampling, exactly one observation is sampled from each cluster. In the arthroplasty example, the cluster bootstrap would be performed by sampling the patients with replacement, while within-cluster sampling proceeds by sampling one hip per patient.

Regarding the difference in aim, the cluster bootstrap is intended to find the sampling distribution of the estimator, which would in our example be the variance of the estimated cumulative incidence of revision for a randomly sampled hip from the population. The two methods coincide only when there is no correlation between units in a cluster.

Shared gamma frailty model

A shared gamma frailty model was proposed to model the within-patient correlation (Ripatti and Palmgren, 2000), and has been applied since (Robertsson and Ranstam, 2003; Schwarzer et al., 2001). A disadvantage of these models is that the correlation is explicitly modeled, and the underlying assumptions do not necessarily hold for arthroplasty data. In particular, only positive correlation between the two THAs can be induced (Wienke, 2003). Not much is known about the correlation between two THAs in one patient. A positive one is possible, e.g. if the patient is very active, both prostheses are prone to earlier failure. However, two prostheses in one patient can be negatively correlated. If the patient favors one of the prostheses, then the prosthesis bearing the most stress is likely

to fail early while the other prosthesis is likely to survive longer. Thus, the shared gamma frailty model does not seem to be entirely adequate.

Cluster Fine-Gray

An extension of the Fine-Gray proportional subdistribution hazards model for clustered data is available (Zhou et al., 2012). More details on standard Fine-Gray regression can be found in Section 6.2. The cluster version has, to the best of our knowledge, not been applied to arthroplasty data yet. The cumulative incidence is estimated using standard Fine-Gray methodology under an independence working assumption, after which the variance is estimated using a sandwich variance estimator. The method was designed for settings where there are unobserved shared factors across individuals, such as multicenter trials or family studies. The correlation structure remains unspecified, making this method more attractive than a frailty model for arthroplasty data.

6.3.2 Methods for the time-dependent bilaterality status

In the terminology of Kalbfleisch and Prentice (2002), a patient's bilaterality status can be viewed as an *internal* time-dependent covariate, meaning that the possibility of its observation depends on the survival status of the patient. Internal time-dependent covariates pose a challenge in competing risks analysis, as their very observation at some time point t informs us that the probability of survival up until time t conditional on the time-dependent covariate is equal to one. It is possible to estimate cause-specific hazards, but prediction of cumulative incidences is not possible when an internal time-dependent covariate is included (Andersen et al., 1993; Cortese and Andersen, 2009). This makes the method of Lie et al. (2004), who propose to include a time-dependent covariate that contains information on a patient's bilaterality status and revision status of the opposite hip, unsuitable for our purposes.

If one's goal is to study the entire patient population, without any specific interest in the bilateral patients, the time-dependence problem can be avoided by only including each patient's first THA in the analysis, as discussed in Section 6.3.1. In this section, we discuss methods for the situation where the goal is to study bilateral patients specifically, or when the loss of data resulting from excluding the second limb is considered prohibitive.

Cortese and Andersen (2009) discuss three methods to incorporate a time-dependent covariate: a multistate model with additional transient states, the landmark analysis of Van Houwelingen (2007), or an extended competing risks model in which all possible combinations between the levels of the time-dependent covariate and cause-specific events are included as final states. These alternatives require a change of research question: the multistate model takes the per-patient point of view as opposed to the per-hip point of view, landmark analysis yields estimates conditional on event-free survival up until a landmark time, as does the extended competing risks model. Before reviewing these three options, we discuss the potential for immortal time bias.

Immortal time bias

A basic principle in survival analysis is that subgroups defined by patient characteristics that are not known at the start of follow-up (such as receiving a second THA), can only be compared with the greatest caution. The reason is the immortal time bias, a well-known phenomenon in observational studies, resulting from flawed statistical analysis (Lévesque et al., 2010; Suissa, 2007). Immortal time refers to a period of follow-up during which the study outcome, which may be death or another event (e.g. revision surgery), cannot occur. It was first described in the context of heart transplant data, when it was noted that the observed improved survival of heart transplant patients was due to selection bias: only patients who survived long enough to receive a heart transplant were included in the transplant group (Gail, 1972).

Analyses of arthroplasty data risk being affected by the immortal time bias as well. The immortal time bias arises when patients with staged bilateral THA are studied as a separate subgroup, because only those patients who survive long enough to be able to receive the second implant are observed. The bias occurs both when revision of one of the implants or death are taken as the endpoints. With arthroplasty data, when the outcome of interest is revision, the bias is subtle. Revision of the first hip does not prevent a patient from joining the staged bilateral group, and thus there is no obvious immortal time bias. However, there is the competing risk of death.

The underlying mechanism of the immortal time bias is illustrated through an artificial example, in which 50% of patients will become staged bilateral exactly 2 years after their index surgery. The first-placed implants of unilateral and bilateral patients are compared. The implants of all patients behave the same: they have a 30% probability of revision after exactly 3 years. In addition, each patient has a 20% probability of dying after 1 year. All percentages are chosen for illustrative purposes and are not meant to be realistic. We assume independence for all events. The process is visualized in Figure 6.1.

When the unilateral and staged bilateral subgroups are created at the end of follow-up, patients that would have become staged bilateral at the 2-year mark but died before realizing that potential, are observed to be unilateral. This leads to an estimate of a zero probability of death for staged bilateral patients, while the cumulative incidence of death is overestimated for unilateral patients. The reverse happens for revision: the cumulative incidence of revision is overestimated for staged bilateral patients, as the competing risk of death is not observed, while it is underestimated for unilateral patients, because the risk set is made artificially large by the inclusion of patients who would have become staged bilateral if they had not died before the second surgery could take place.

The severity of the effect of the immortal time bias depends on the revision, mortality and bilaterality rates, and also on the research question. With 5% revision, 5% mortality and 20% bilaterality, the bias in the artificial data example is inconsequential for the cumulative incidence of revision, but still relatively large for death. Moreover, statistically significant differences in implant survival between two groups can be very small, even less than 1%, when the follow up is short. In such a case, even a small bias may be large enough to give the false impression of a difference between subgroups where there is actually none. In addition, the Swedish Hip Arthroplasty Register reports 23-year revision rates of up to 38.5% for men who are 50-59 years at index surgery (SHAR, 2014). After such a long follow up, immortal time bias may significantly affect analyses, and thus clini-

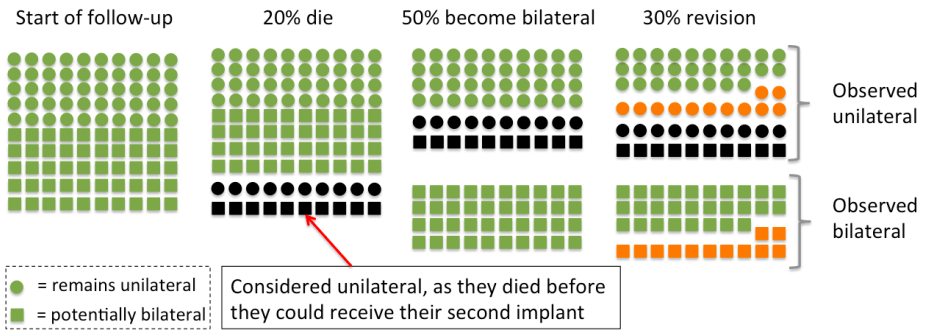

Figure 6.1: Visualization of the artificial data example. Circles denote patients who will only have one implant, while squares indicate patients who will become bilateral at the 2-year mark. Green indicates event-free patients, black patients who die before experiencing revision, and orange patients whose prosthesis has been revised. The subgroup analysis ignores the fact that some patients will have died before realizing their potential of becoming bilateral, and thus some potentially bilateral patients will be considered unilateral.

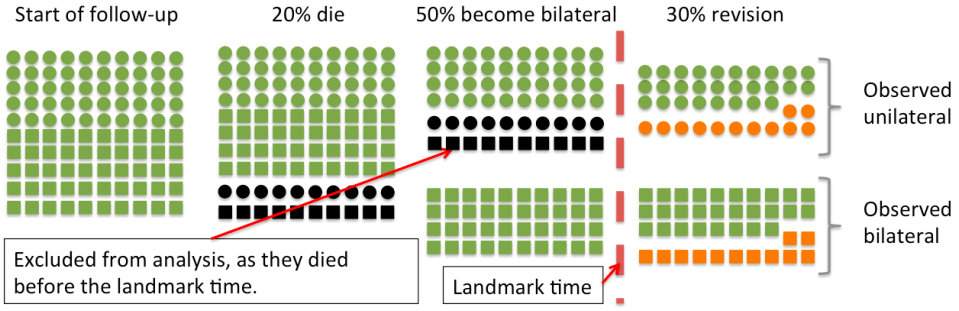

Figure 6.2: A landmark time is chosen, in this case after the patients become bilateral. All patients who died or were revised before the landmark time are excluded from the analysis.

cal results based on subgroup analysis with staged bilateral patients should be interpreted with caution.

Landmark analysis

Landmark analysis allows for comparison of unilateral and staged bilateral patients without the risk of immortal time bias (Cortese and Andersen, 2009; Van Houwelingen, 2007). The first step is to choose a landmark time of for example 2 years. The choice of landmark time should be guided by the research question. Only patients who are still alive and have not experienced revision yet at the landmark time are included in the analysis. This ensures a fair comparison between the two groups, as both need to survive for the same minimum amount of time in order to be included in the analysis.

The next step is to create the subgroups: patients who have become bilateral by the landmark time, and patients who were unilateral at the landmark time. As only each patient's status at the landmark time is considered, the latter group includes patients who may receive a second implant after the landmark time. The procedure is illustrated in Figure 6.2.

When the landmark subgroups have been made, the cumulative incidence can be estimated, for example by using one of the methods described in Section 6.2. The interpretation of the resulting models is conditional on the landmark time. Thus, conclusions can be drawn for comparison of unilateral and staged bilateral patients, conditional on the fact that these patients were still alive and did not undergo revision by the landmark time. This is a limitation to the method: the conclusions only hold for patients who are still alive and unrevised by the landmark time point. This is not a negative per se, as this question will be of interest to a patient who has survived some time unrevised since the primary or index THA. However, excluding the first few postoperative months or years from analysis may not be satisfactory in a situation where mortality or revision risk are especially high immediately following surgery.

Extended competing risks model and multistate models

The second approach discussed by Cortese and Andersen (2009) is an extended competing risks model, which has all possible combinations between internal covariate levels and cause-specific events as final states. In the case of arthroplasty data, such a model could be represented as in Figure 6.3.

The change in status from unilateral to bilateral comes bundled with the introduction of a second outcome: revision of the second hip. Thus, the outcome "revision" needs a more precise definition, such as "revision of the first THA", in which case the outcome of the second THA is disregarded.

The disease process of a patient can be more fully captured by a multistate model with transient states. Such a multistate model allows inclusion of a patient's second THA in a natural manner. Another advantage of these extended models is that they allow us to take a per-patient point of view, which is more useful to the orthopaedic surgeon than the classical per-hip point of view. See Figure 6.4 for an example of such a model.

There is a Markovian assumption behind this model, which can be relaxed. It may be the case that the probability of transitioning from for example "bilateral" to one of the

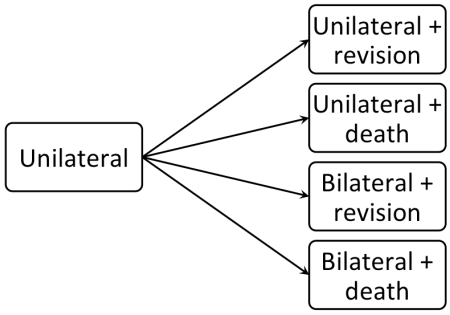

Figure 6.3: Extended competing risks model for Total Hip Arthroplasty.

revised states, depends on the amount of time spent as a unilateral patient. It is possible to model the intensity regulating the transitions as a function of the time spent as a unilateral patient, resulting in a semi-Markov model (Cortese and Andersen, 2009; Putter et al., 2007).

While multistate models have been rarely used in orthopaedic studies, there has been a successful application to the data of the Australian Orthopaedic Association National Joint Replacement Registry (Gillam et al., 2013, 2012).

6.3.3 Clinical relevance

Two reviews of arthroplasty studies found that is commonly believed that the bilateral patients do not affect the results of the analyses too much, and thus the dependence of their observations is often ignored (Bryant et al., 2006; Ranstam et al., 2011). Robertsson and Ranstam (2003) find that the effect of subject dependency in total knee arthroplasty is negligible, and explain this by saying that the source of the bias generated by ignoring dependency consists solely of bilateral patients with revisions on both sides, of which there are very few. Findings of Ripatti and Palmgren (2000), Schwarzer et al. (2001), Visuri et al. (2002) and Lie et al. (2004) for THA are similar. A contributing factor is that hip implant survival is very high.

The findings that ignoring the within-patient dependence does not significantly affect results are all within the context of questions about the entire patient population. Whether ignoring the presence of bilateral patients is problematic depends on the goal of the analysis, and on the similarity of the outcomes for the two prostheses. If one is interested in the time to revision for any hip, then ignoring the dependence may be a pragmatic solution if the first and second THAs have similar survival properties and similar associations with the covariates, and especially if implant survival is high in general. In that case, the ignored dependence will likely only affect the confidence intervals. However, if the implants of bilateral patients have different survival properties than unilateral prostheses, grouping everyone together without extra consideration does not make much sense. In that cases, studying unilateral and bilateral patients separately will provide more useful clinical insights.

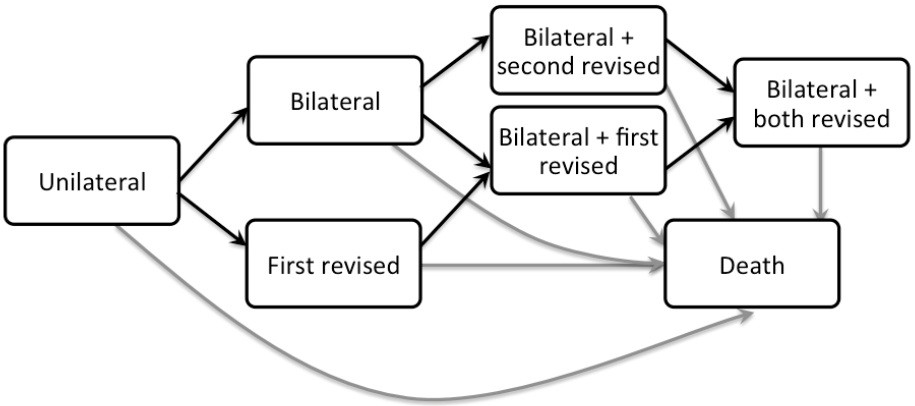

Figure 6.4: Multistate model for Total Hip Arthroplasty.

When the research question is concerned with bilateral patients, caution is warranted. The main potential pitfalls lie in the time-dependence of a patient's bilateral status. Extra care needs to be taken when the proportion of bilateral patients is high, and when the patients who become staged bilateral tend to do so relatively long after the first surgery. A naive subgroup analysis may be affected by immortal time bias. Landmark analysis or a multistate model seem appropriate solutions in this case.

In any analysis of arthroplasty registry data, researchers should carefully consider the impact the bilateral patients may have on their results, define their research population precisely, and select the statistical method accordingly.

6.4 Data structure

The data set contains data on 161,434 primary total hip arthoplasties, undertaken between 2007 and 2014. Arthroplasties after tumors or fractures, and hemiarthroplasties were not included. The survival information is captured in the following four variables:

1. **Status_revision**: indicates whether the hip was revised.
2. **Status_death**: indicates whether the patient has died.
3. **Surv_revision**: time at risk until revision.
4. **Surv_death**: time at risk until death.

In order to illustrate the time to event structure of the data, consider the following patients (only status indicators and time at risk shown for clarity):

Patient	Status_revision	Status_death	Surv_revision	Surv_death
103867	0	0	7.67	7.67
99702	1	0	3.97	6.56
88945	0	1	0.18	0.18
6645	1	1	0.13	0.52

Patient 103867 was under follow up for 7.67 years and was still alive at the end of follow up, without revision of his or her implant. Patient 99702 was under follow up for 6.56 years. After 3.97 years, his or her hip implant was revised. The patient was still alive at the end of follow up. Patient 88945 died after 0.18 years of follow up, without revision of his or her implant. The implant of patient 6645 was revised after 0.13 years, and the patient died 0.39 years later, at 0.52 years of follow up.

Each line in the data set corresponds to one hip. However, there are bilateral patients included in the data set, with a hip implant on both sides. Some examples in the data:

Patient	Status_revision	Status_death	Surv_revision	Surv_death
5	0	0	3.16	3.16
5	0	0	2.30	2.30
3044	0	1	1.63	1.63
3044	1	1	0.15	1.13
22112	0	0	4.47	4.47
22112	1	0	1.36	2.86

Patient 5 received a second hip implant 0.86 years after the first, and was then followed for another 2.30 years. During that time, none of the implants were revised, and the patient was still alive at the end of follow up. Patient 3044 received his or her second implant after 0.5 years. The second implant was revised 0.15 years after its placement. The first was never revised. The patient died 1.63 years after the first prosthesis was implanted. Patient 22112 received his or her second implant after 1.61 years, and it was revised 1.36 years later. The patient was still alive, without revision of the first implant, at the end of follow up at 4.47 years.

The statistical complications associated with the presence of bilateral patients in the data set are discussed in Section 6.3. Before proceeding to the data analysis in Section 6.5, we describe the remaining variables in the data set. The variables used in the model are listed below.

1. **Age**: age of patient at index surgery.
   Converted to the five age categories used by the LROI: younger than 50, 50-59, 60-69, 70-79, 80 and older.

2. **GENDER**: gender of patient.

3. **ASACLASH**: American Society of Anesthesiologists (ASA) classification.
   1: A normal healthy patient. 2: A patient with mild systemic disease. 3: A patient with severy systemic disease. 4: A patient with severe systemic disease that is a constant threat to life (ASA, 2014).

4. **DIAGH**: diagnosis.
   The nine diagnoses in the data set were combined into five diagnostic groups, following the recommendation of the clinician: osteoarthritis, post-Perthes and dysplasia, rheumatoid and inflammatory arthritis, osteonecrosis, and late posttraumatic combined with all other diagnoses.

5. **FIXH_incl_rev**: type of fixation of the hip implant.
   Cementless, hybrid, cemented or reversed hybrid.

6. **Hospitaltype**: type of hospital.
   General, academic, or private. Taken as a proxy for unmeasured confounders, outcome not reported.

The distribution of the patient characteristics is shown in Figure 6.5.

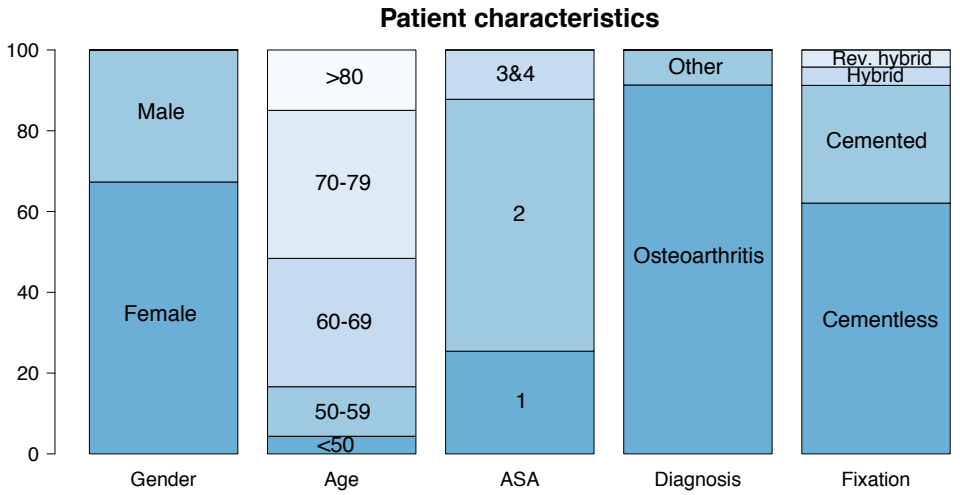

Figure 6.5: Barplot of patient characteristics in the LROI data set.

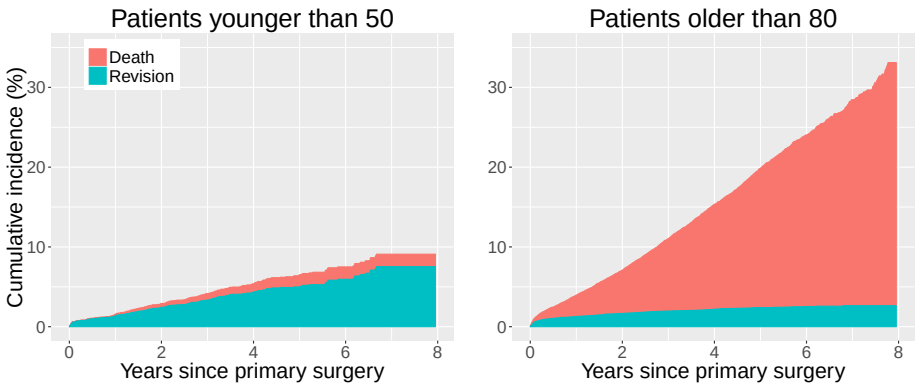

Figure 6.6: Aalen-Johansen estimator of the cumulative incidences of revision and death, for the youngest and oldest patients. Estimated using each patient's first THA.

Besides those variables, the data set contained information on the side of operation, year of operation and revision, and whether the hip had been operated on before. These were left out of the model because of redundancy or, in case of previous operations on the same hip, low completeness.

6.5 Results on the LROI data

This section contains results on data from the Dutch Arthroplasty Register (LROI). The analysis is based on total hip arthroplasty (THA) and subsequent revision surgeries performed on a total of 161,434 hips in 144,513 patients, between January 1st, 2007 and December 31st, 2014. We aim to identify variables associated with the cumulative incidence of revision, for all patients and in particular the bilateral patients.

A limitation of these analyses is that it requires identifying whether a THA was a patient's first or second. This is not problematic for patients whose both surgeries took place after 2007, but it is for bilateral patients whose first THA took place before the establishment of the LROI. If a patient's first THA happened before 2007, and his or her second THA happened after 2007, then only the second THA is recorded in the data set, and we would require an indicator to alert us to the fact that it is that patient's second, not first THA. Such an indicator exists in the form of the Charnley score, but this score has only been recorded sinds mid 2013, and we do not have access to it at time of writing. This is discussed in more detail in 6.5.3.

We first describe the analyses and present the results. We discuss the results, limitations of these analyses and plans for future analyses in Section 6.5.5.

6.5.1 Competing risks

The need for competing risks methods is illustrated in Figure 6.6. It shows the cumulative incidences of revision and death, estimated separately for the youngest and the oldest patients. The cumulative incidences were estimated using each patient's first THA, noting that for some patients, this will actually be their second, as explained above.

The Figure shows a very strong competing risk of death for patients older than 80. For patients under 50, the competing risk of death in this relatively short amount of follow up is so small as to be negligible. Given the average age of patients undergoing THA, which is 69, the competing risk of death cannot be ignored.

6.5.2 All patients

Before zooming in on the bilateral patients, we consider the entire patient population. Following the recommendations of Grambauer et al. (2010) and Latouche et al. (2013), both Fine-Gray regression and cause-specific Cox regression are performed.

We use Fine-Gray regression for clustered data (Zhou et al., 2012) to find variables associated with revision. As discussed in Section 6.3, this analysis does not account for the time between two THAs for bilateral patients. As a form of sensitivity analysis, standard Fine and Gray regression was also performed on the entire data set, and on all first THAs

Table 6.1: Fine-Gray regression for all patients. Reference category between parentheses.

Variable	Cluster Fine-Gray all data		Fine-Gray all data		Fine-Gray first THAs	
	coefficient	s.e.	coefficient	s.e.	coefficient	s.e.
Gender (female)						
Male	0.080	0.036	0.080	0.036	0.076	0.037
Age (< 50)						
50-59	-0.073	0.083	-0.073	0.082	-0.084	0.087
60-69	-0.285	0.080	-0.285	0.079	-0.277	0.083
70-79	-0.300	0.083	-0.300	0.082	-0.312	0.086
≥ 80	-0.426	0.094	-0.426	0.093	-0.407	0.098
ASA (ASA 1)						
ASA 2	0.090	0.040	0.090	0.040	0.086	0.042
ASA 3 & 4	0.234	0.060	0.234	0.060	0.249	0.062
Diagnosis (Osteoarthritis)						
Osteonecrosis	0.027	0.097	0.027	0.096	0.063	0.098
Post-Perthes/Dysplasia	-0.123	0.110	-0.123	0.110	-0.135	0.115
Late posttraumatic	0.434	0.091	0.434	0.091	0.421	0.093
Rheum./infl. arthritis	-0.085	0.162	-0.085	0.163	-0.140	0.177
Fixation (Cementless)						
Cemented	-0.559	0.047	-0.559	0.046	-0.533	0.049
Hybrid	-0.260	0.089	-0.260	0.088	-0.284	0.094
Reversed hybrid	0.046	0.078	0.046	0.078	0.081	0.081

(which will in some cases be the second THA, as explained above). The results are given in Table 6.1.

Cause-specific Cox regression was done for revision and death, both on the first THAs and on all THAs. The results are given in Table 6.2.

The subdistribution hazard ratios resulting from cluster Fine-Gray, as well as the hazard ratios resulting from cause-specific Cox regression on all patients are given in Table 6.3, together with the p-values and the numbers of revisions and deaths.

The cumulative incidence of revision is associated with gender, age, ASA score, diagnosis and type of fixation. Men are more likely to experience revision than women. The cumulative incidence of revision decreases with age, and increases with ASA score. It is less for hybrid fixation and even smaller for cemented fixation, compared to cementless fixation. This is set in context and discussed in Section 6.5.5.

As a visual check of the proportionality assumptions for Fine-Gray regression and cause-specific Cox regression, nonparametric estimates of the cumulative incidences of revision and the cause-specific cumulative hazards of revision and death are given in Figures 6.7, 6.8 and 6.9.

Table 6.2: Cause-specfic Cox regression for all patients. Reference category between parentheses.

Variable	First THAs				All data			
	Revision		Death		Revision		Death	
	coefficient	s.e.	coefficient	s.e.	coefficient	s.e.	coefficient	s.e.
Gender (female)								
Male	0.085	0.037	0.478	0.027	0.088	0.035	0.476	0.026
Age (< 50)								
50-59	-0.079	0.086	0.687	0.159	-0.068	0.082	0.767	0.155
60-69	-0.270	0.082	1.099	0.150	-0.278	0.078	1.162	0.147
70-79	-0.296	0.085	1.803	0.149	-0.284	0.081	1.866	0.146
≥ 80	-0.365	0.096	2.583	0.149	-0.384	0.092	2.636	0.147
ASA (ASA 1)								
ASA 2	0.091	0.042	0.313	0.039	0.095	0.040	0.315	0.038
ASA 3 & 4	0.281	0.062	1.100	0.044	0.265	0.059	1.112	0.042
Diagnosis (Osteoarthritis)								
Osteonecrosis	0.078	0.096	0.672	0.061	0.040	0.094	0.651	0.060
Post-Perthes/Dysplasia	-0.132	0.115	0.023	0.120	-0.120	0.110	0.049	0.115
Late posttraumatic	0.441	0.092	0.645	0.062	0.453	0.091	0.636	0.062
Rheum./infl. arthritis	-0.132	0.176	0.499	0.106	-0.078	0.163	0.484	0.103
Fixation (Cementless)								
Cemented	-0.530	0.048	0.138	0.029	-0.556	0.046	0.140	0.028
Hybrid	-0.282	0.093	0.195	0.055	-0.258	0.088	0.200	0.053
Reversed hybrid	0.082	0.081	0.048	0.078	0.048	0.078	0.060	0.075

Table 6.3: Numbers of events (all patients). Subdistribution hazard ratios (SHR) from cluster Fine-Gray, and hazard ratios (HR) from cause-specific Cox on all patients. Reference category between parentheses.

Variable				Cluster Fine-Gray		Cause-specific Cox			
				revision		revision		death	
	event-free	revisions	deaths	SHR	p-value.	HR	p-value	HR	p-value
Gender (female)	101,121	2,482	4,410	1		1		1	
Male	48,305	1,395	2,734	1.08	0.025	1.09	0.013	1.61	< 0.001
Age (< 50)	6,729	252	55	1		1		1	
50-59	18,816	641	297	0.93	0.38	0.93	0.41	2.15	< 0.001
60-69	48,796	1,256	1,144	0.75	< 0.001	0.76	< 0.001	3.20	< 0.001
70-79	54,819	1,293	2,949	0.74	< 0.001	0.75	< 0.001	6.46	< 0.001
≥ 80	20,894	448	2,724	0.65	< 0.001	0.68	< 0.001	13.96	< 0.001
ASA (ASA 1)	37,105	999	946	1		1		1	
ASA 2	90,014	2,151	3,635	1.09	0.025	1.10	0.017	1.37	< 0.001
ASA 3 & 4	16,371	469	1,910	1.26	< 0.001	1.30	< 0.001	3.04	< 0.001
Diagnosis (Osteoarthritis)	137,592	3,491	6,362	1		1		1	
Osteonecrosis	4,316	136	328	1.03	0.78	1.04	0.67	1.92	< 0.001
Post-Perthes/Dysplasia	3,628	94	84	0.88	0.26	0.89	0.27	1.05	0.67
Late posttraumatic	3,304	134	298	1.54	< 0.001	1.57	< 0.001	1.89	< 0.001
Rheum./infl. arthritis	1,522	41	104	0.92	0.60	0.92	0.63	1.62	< 0.001
Fixation (Cementless)	93,192	2,718	3,254	1		1		1	
Cemented	42,659	752	3,124	0.57	< 0.001	0.57	< 0.001	1.15	< 0.001
Hybrid	6,668	151	448	0.77	0.0033	0.77	0.003	1.22	< 0.001
Reversed hybrid	6,331	192	205	1.05	0.56	1.05	0.54	1.06	0.42

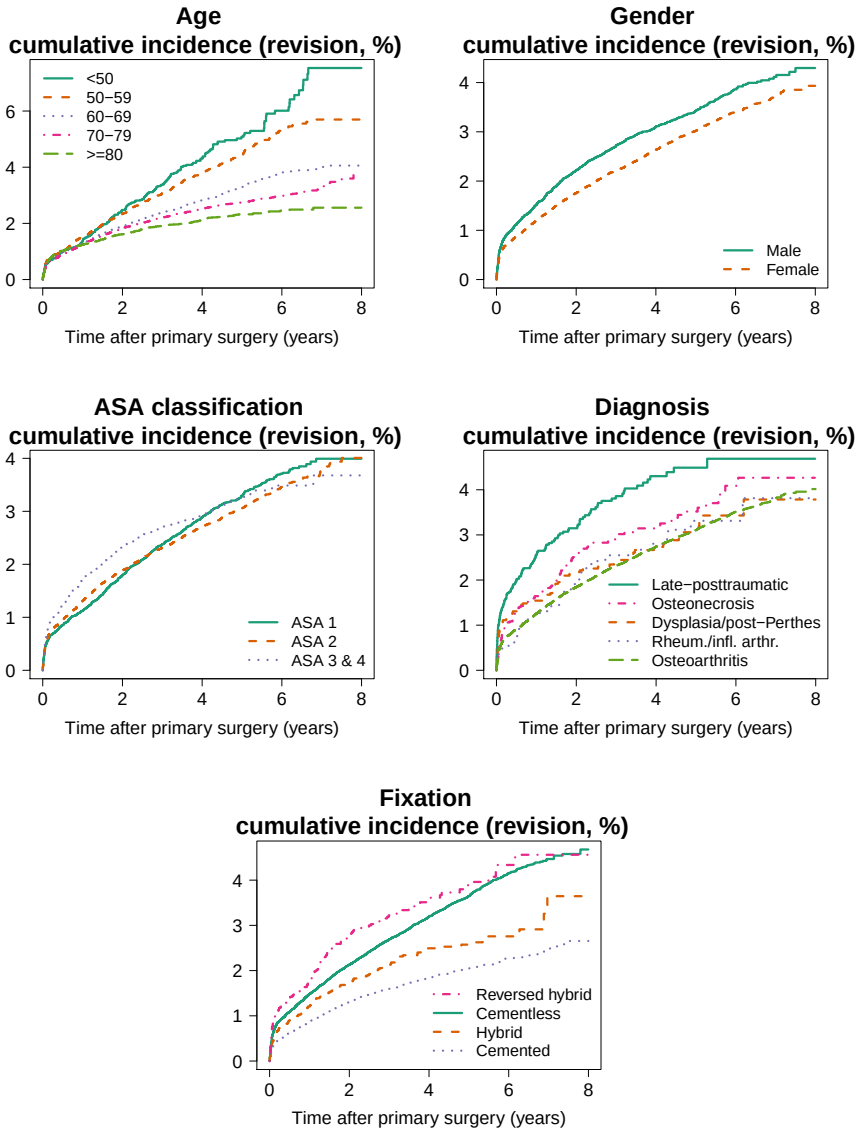

Figure 6.7: Aalen-Johansen estimates of the cumulative incidences of revision, using all hips in the data set.

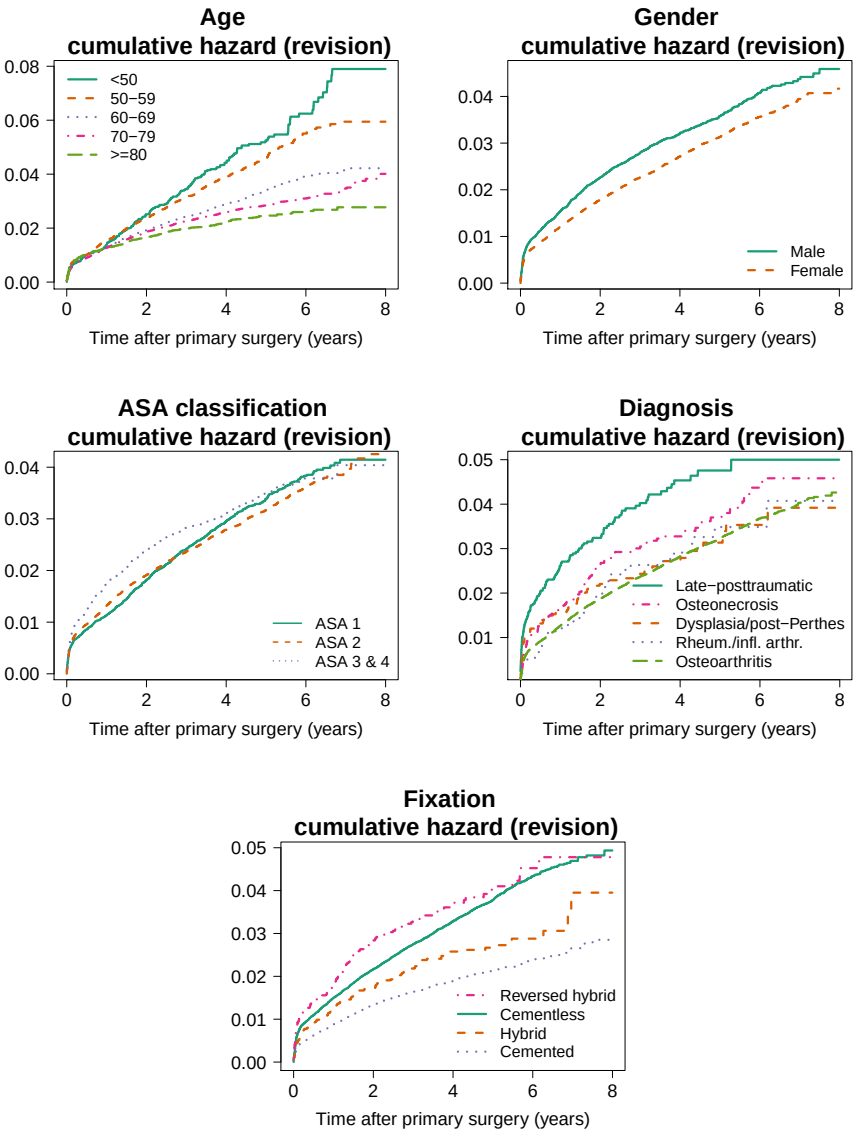

Figure 6.8: Nelson-Aalen estimates of the cumulative hazard of revision, using all hips in the data set.

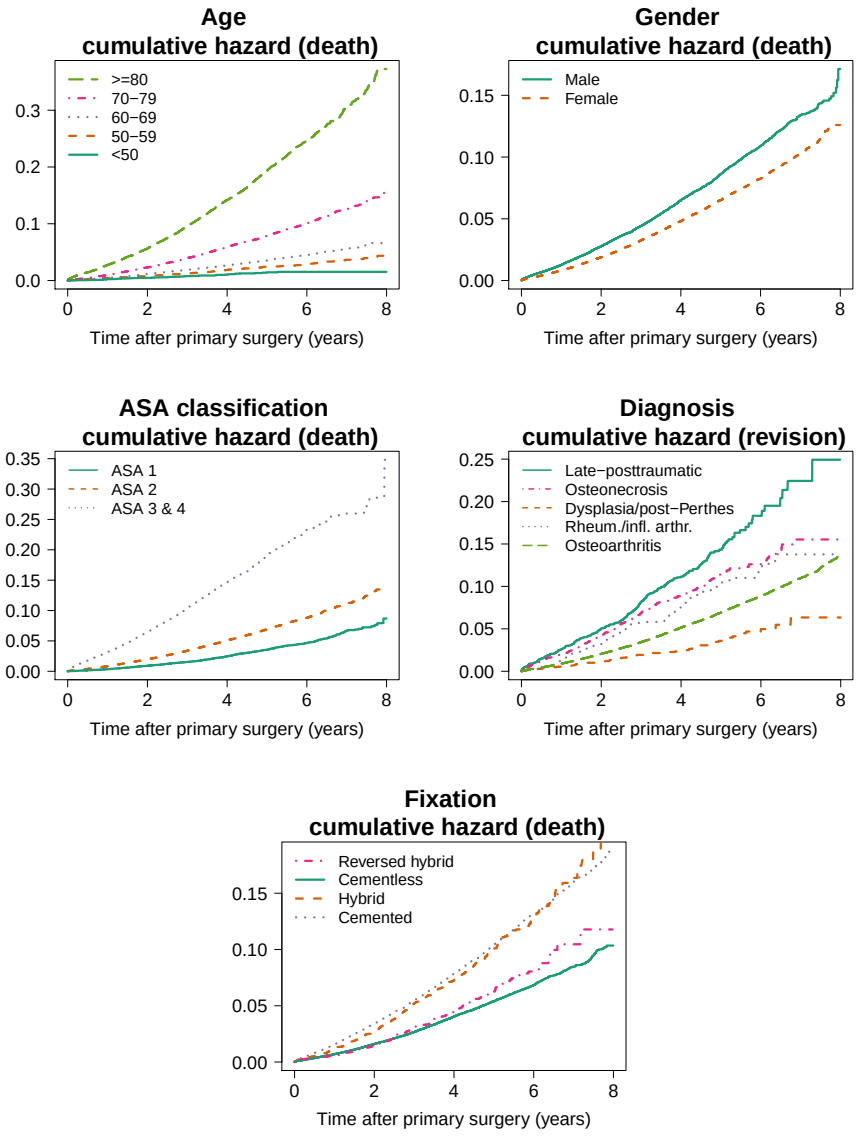

Figure 6.9: Nelson-Aalen estimates of the cumulative hazard of death, using all hips in the data set.

Table 6.4: Number of known second THAs in the LROI data.

	2007	2008	2009	2010	2011	2012	2013	2014
Number of THAs (total)	7,938	13,781	19,976	21,992	22,713	23,971	24,631	26,438
Number of second THAs	197	632	1,304	1,975	2,539	2,964	3,436	3,874
Percentage second THAs	2.5%	4.6%	6.5%	9.0%	11.2%	12.4%	13.9%	14.7%

6.5.3 Comparison of unilateral and bilateral patients

We compare rates of revision for the first implanted hip for unilateral and staged bilateral patients at the landmark time of 1 year, meaning that we include those patients who had not yet experienced revision after 1 year and divide them into groups who have received either one or two prostheses by 1 year.

As explained above, this analysis is problematic, because the "unilateral" group will contain some second THAs from bilateral patients. To get a sense of how many of these second THAs we may miss, we compute for each year the number of THAs that are known to be the second of a bilateral patient, because the corresponding first THA took place in or after 2007. These numbers are given in Table 6.4. For reference, the Charnley score was recorded in 2014, and in that year, 20% of THAs concerned the placement of a second hip (LROI, 2014).

To mitigate the problem of the unidentified second THAs, we only study patients whose first (known) procedure took place in 2010 or later. Based on clinical experience, we perform landmark analysis at the 1 year landmark, for 4.5 years of follow-up. In total, 75,397 patients were included in the unilateral group, and 5,031 in the bilateral group. Gray's test detects a difference in cumulative incidence of revision between patients who are unilateral or bilateral 1 year after the first THA ($p = 0.003$). The estimated cumulative incidences are given in Figure 6.10. As shown in Figure 6.10, the first implanted prosthesis of a patient who has become bilateral at the one year mark is less likely to be revised compared to unilateral prostheses, if we compare patients who have not undergone revision and are still alive one year after the first THR.

6.5.4 Second-implanted hips

For a comparison of the second-implanted hips of staged bilateral patients, no time-dependent covariates are required, as their time point of origin is the time of the second primary THA. We thus compute the unadjusted and adjusted cumulative incidences without any further considerations. Characteristics of the bilateral patients are given in Table 6.5.

The results from the Fine-Gray regression are given in Table 6.6. When we consider the second-placed hips of staged bilateral patients, the amount of time between the two surgeries is a significant variable. The unadjusted cumulative incidence of revision of the second hip is significantly different for patients whose second hip was placed more than one year after the first one, compared to patients whose hips were both placed within one year ($p = 0.009$). This is illustrated in Figure 6.11, which shows that the unadjusted cumulative incidence of revision for patients whose surgeries take place more than one year apart is higher than for patients whose surgeries take place within one year.

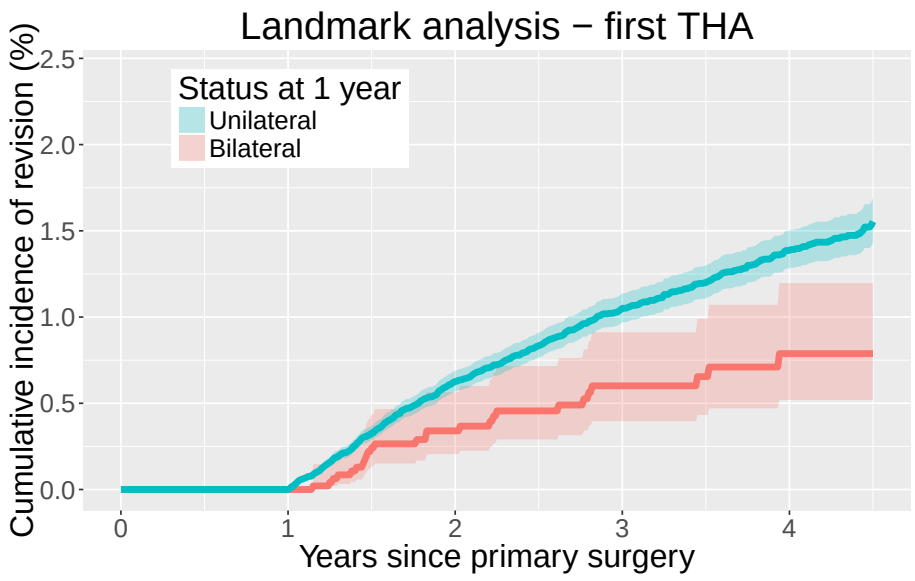

Figure 6.10: Aalen-Johansen estimator of the cumulative incidence of revision of the first hip implant for patients who are unilateral or bilateral and event-free at the 1 year landmark. The cumulative incidence of revision is higher for patients who are (still) unilateral 1 year after their first THA.

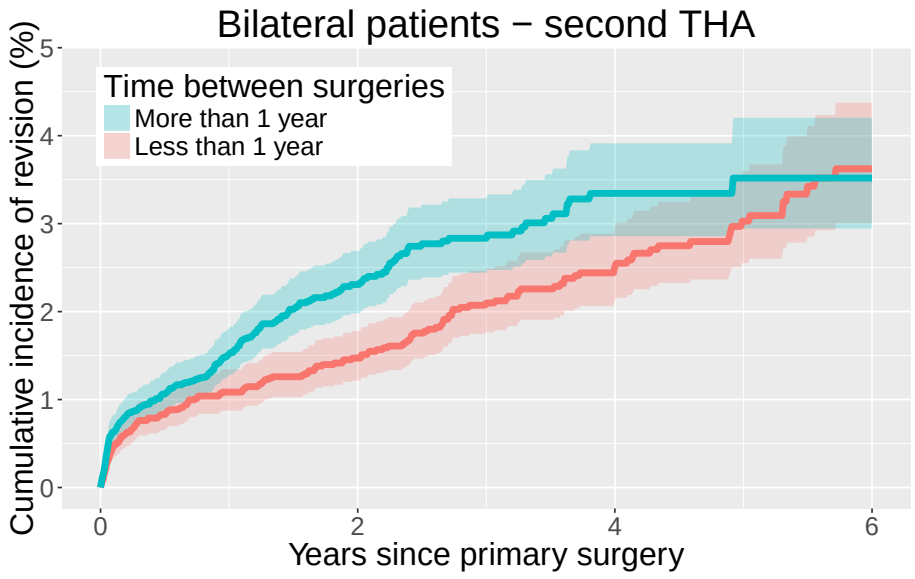

Figure 6.11: Aalen-Johansen estimator of the cumulative incidence of revision for a bilateral patient's second hip implant. The cumulative incidence of revision is higher for patients who receive their second THA after more than 1 year.

Table 6.5: Characteristics of bilateral patients as recorded at the time of the second THA, compared to all patients.

Characteristics	Second hip within 1 year	Second hip after more than 1 year	All patients
Total	8,027	8,894	161,434
Mean age	67.2	70.1	69.0
Female	68.7%	72.8%	67.3%
ASA 1	25.4%	20.9%	25.4%
ASA 2	64.2%	66.6%	62.4%
ASA 3&4	10.4%	12.5%	12.2%
Osteoarthritis	92.3%	95.6%	91.3%
Cementless fixation	64.1%	64.2%	62.1%

When the interoperative time is adjusted for, we find few remaining significant predictors for revision of the second prosthesis. Patients with cemented hips are less likely to experience revision than patients with cementless fixation, and the oldest patients are less likely to experience revision than the youngest. Gender, ASA and diagnosis do not appear to play a significant role when we consider only the second prostheses of bilateral patients.

6.5.5 Discussion and outlook

All patients

Outcome of Fine-Gray regression

The finding that young age, male sex, high ASA score, uncemented prostheses and an earlier trauma are risk factors for revision (Table 6.3) is consistent with previous studies (Prokopetz et al., 2012). Many of those previous analyses were done without accounting for the competing risk of death, but the conclusions still stand when it is corrected for. Explanations are available in the clinical literature. Younger patients are typically more active and heavier, leading to increased stress on the implant components compared to older patients (Johnsen et al., 2006). Higher mechanical stress may also explain the increased risk of revision for men compared to women, together with hip kinematics (Gallo et al., 2010). The ASA score is an indicator of a patient's preoperative health status, and can be predictive of the early functional status (Hooper et al., 2012). Regarding fixation, the lower cumulative incidence of revision for cemented implants compared to uncemented implants is well-documented (Makela et al., 2014). In addition, we find that hybrid prostheses have a lower risk of revision compared to uncemented prostheses, a finding for which previous studies found evidence in either direction (Prokopetz et al., 2012). Of the diagnoses included in this study, only those patients who receive a hip prosthesis long after a trauma have a significantly different risk of revision than patients who have a diagnosis of osteoarthritis. After a trauma, risk of dislocation is increased, as anatomic structures may be compromised (Mallory et al., 1999). Thus, the results are consistent with the clinical literature.

Table 6.6: Fine-Gray regression based on bilateral patients' second THA.

Variable	event-free	revisions	deaths	Subdistribution hazard ratio	95% CI	p-value
Time between surgeries (<1 year)	7,588	168	271	1		
≥ 1 year	8,463	198	233	1.37	1.10-1.70	0.004
Gender (female)	11,374	243	323	1		
Male	4,612	123	180	1.15	0.92-1.44	0.23
Age (< 50)	650	23	2	1		
50-59	1,973	69	29	0.96	0.57-1.62	0.87
60-69	5,444	118	96	0.64	0.38-1.09	0.096
70-79	5,962	132	224	0.74	0.44-1.26	0.27
≥ 80	1,999	24	152	0.44	0.23-0.84	0.013
ASA (ASA 1)	3,652	91	64	1		
ASA 2	10,316	229	277	1.12	0.87-1.45	0.37
ASA 3& 4	1,729	34	137	1.05	0.70-1.59	0.80
Diagnosis (Osteoarthritis)	15,100	341	470	1		
Osteonecrosis	392	7	16	0.66	0.30-1.47	0.31
Post-Perthes/Dysplasia	333	9	7	0.99	0.48-2.03	0.98
Late posttraumatic	58	3	4	2.91	0.92-9.18	0.069
Rheumatoid/inflammatory arthritis	168	6	7	1.35	0.60-3.07	0.47
Fixation (Cementless)	10,277	281	241	1		
Cemented	4,374	53	212	0.45	0.33-0.61	< 0.001
Hybrid	650	17	29	0.97	0.58-1.62	0.89
Reversed hybrid	671	12	16	0.73	0.41-1.30	0.28

Outcome of cause-specific Cox regression
The conclusions from the cause-specific Cox model are in line with the results from Fine-Gray regression. The subdistribution hazard ratios for revision and the cause-specific hazard ratios for revision are numerically very close (Table 6.3). This may be due to the heavy censoring (Grambauer et al., 2010): there were 161,434 hips in the data set and only 11,076 events (3,897 revisions and 7,179 deaths).

The added value of two separate analyses for death and revision is visible for those variables where the coefficients for the two hazards have opposite signs: old age, diagnoses of post-Perthes / dysplasia or rheumatoid/inflammatory arthritis, and cemented or hybrid fixation. In all these cases, the cause-specific hazard of revision is decreased; the cause-specific hazard of death increased, and the cumulative incidence of revision decreased. Besides the explanations already provided above, this analysis makes clear that another effect may be that patients with these characteristics are revised less frequently because the rate of occurence of death is increased.

The proportionality assumptions
We highlight two aspects of Figures 6.7, 6.8 and 6.9. First, all plots in Figure 6.7 are remarkably similar to the corresponding plots in Figure 6.8. We already observed that the subdistribution and cause-specific hazards for revision are numerically very close. This is most likely due to the heavy censoring.

The second aspect is that there is some evidence for violation of the proportional subdistribution/cause-specific hazards assumption. The assumption seems to hold for neither hazard for revision for the ASA score, age (first 2 years) and diagnosis. For age, the violation could be due to the categorization. The proportionality assumption does appear to be reasonable for gender and fixation, and for the cause-specific hazards of death. This can be investigated further using, for example, the methods listed in Section 6.2.

Sensitivity to the presence of bilateral patients
The differences between standard Fine-Gray and cluster Fine-Gray regression are negligible (Table 6.1). The estimated coefficients are the same, and the standard errors only differ on the third decimal place. The coefficients and standard errors estimated using only the first THAs are different compared to cluster Fine and Gray, but the signs of all coefficients are the same, and the same coefficients would be significant at the 5% level. The cause-specific Cox regression is not substantively impacted by the within-subject dependence of the bilateral patients either (Table 6.2); the differences between the estimated coefficients based on all THAs or only the first THAs are minimal, and conclusions based on a 0.05-cutoff for the p-values would be the same.

Bilateral patients

Results
Our results indicate that the cumulative incidence of revision is different for staged bilateral patients than for unilateral patients, and that staged bilateral patients are not a homogeneous subgroup. Interoperative time is an important factor to take into account. If a patient's second THR takes place within 1 year, not only does his or her first prosthesis

survive longer compared to unilaterally implanted prostheses, but his or her second prosthesis is less likely to be revised than the second prosthesis of a bilateral patient whose second THR took place more than 1 year after the first.

The results for the first bilateral implant compared to a unilateral implant correspond with the findings of the Swedish Hip Arthroplasty Register; they report better survival for the first bilateral THA compared to a unilateral implant (SHAR, 2014). However, in homogeneous subgroups consisting of patients with a diagnosis of osteoarthritis, no difference in survivorship of the first bilateral prosthesis compared to the unilaterals was found (Havelin et al., 1995; Lie et al., 2004; Visuri et al., 2002).

Most variables that were significant for all patients, are not significant at the 5% level for bilateral patients, when the time between surgeries is included as a categorical variable (Table 6.6). Only very old age and a cemented fixation remain significant. This may be in part because the time between surgery serves as a proxy for a patient's general health status and activity level, as will be discussed below.

Limitations

We must be careful not to draw causal conclusions, as the data are observational. Furthermore, there are limitations to the comparison of unilateral and bilateral patients. First of all, even after removing the data from 2007-2009, some second THAs will have been included in the "unilateral" group. Two studies indicate that the second THA has better survival than a unilateral implant, but the evidence is limited (Lie et al., 2004; Visuri et al., 2002). It is thus not clear how the presence of unidentified second THAs may have affected the estimates presented in Figure 6.10. A second limiting factor is that the landmark analysis precludes us from drawing conclusions about the risk of revision within the first year.

The analysis of the second THAs does not suffer from these limitations, and suggests that implant survival is better for patients who receive their second THA within 1 year after the first. When interpreting the results in Figure 6.11, the competing risk of death needs to be considered. A bilateral patient who receives his or her second implant after more than 1 year is on average older than a bilateral patient whose second surgery takes place within 1 year, as supported by Table 6.6. Being older, the patient may be at lower risk of revision. Yet Figure 6.11 and Table 6.5 indicate that patients who receive their second implant after more than 1 year have higher risk of revision, lending credence to the hypothesis that the two groups of bilateral patients differ from each other in some other respect.

Timing of the second THA

The protective effect of a shorter time between the two surgeries has been observed before (Havelin et al., 1995; Lie et al., 2004; Möllenhoff et al., 1994; Visuri et al., 2002). The cutoff for significant differences found in each of these studies has been different, and none of the studies accounted for the competing risk of death. The optimal lengths of interoperative time as reported by these studies are within 1 year (Visuri et al., 2002), within 2 years (Lie et al., 2004), or within 1-3 years (Möllenhoff et al., 1994).

Our results suggest that the relevant period may be as short as 1 year. However, again it must be stressed that these data are observational, and the conclusion that bilateral

THAs should be placed as soon as possible cannot be drawn.

We offer some clinical considerations on the observed protective effect of a shorter interoperative time period. One factor may be the relationship between activity levels and revision risk. A patient who receives two implants within 1 year may have other health issues associated with impeded mobility, thus putting less strain on the first replaced hip, leading to longer survival of the implant. Bilateral patients whose two surgeries are more than 1 year apart may have suffered from impaired mobility to a lesser extent, explaining why their implants are more prone to early failure than those of patients who received their second implant soon after the first.

On the other hand, with some diagnoses, patients may elect to have the second THA sooner rather than later. The patients who do so are likely to be in good health, and more satisfied with the outcome of the first THA. This may actually lead to worse survival of the implants, as these are generally more active patients.

A third factor may be that the group of osteoarthritis patients is not homogeneous, and that those who receive a second implant soon after the first represent a subgroup within the group of osteoarthritis patients for whom osteoarthritis should be considered a systemic disease.

Outlook

Only a randomized clinical trial can confirm hypotheses about interoperative time and improved outcomes for staged bilateral patients. The LROI is still relatively young. With the passing of time, more data will become available, allowing more detailed study of bilateral patients. A multistate model has been applied to the data from the Australian National Joint Replacement Registry, with promising results (Gillam et al., 2013, 2012). One insight from the Australian multistate model is that women are more likely than men to experience a second joint replacement surgery, which may be due to the lower mortality risk for women, or because women may have more extensive osteoarthritis. We expect that such a model, applied to the LROI data, would provide more insight into the path a patient may take from unilateral to possible bilateral, revision and/or death. The Dutch hip replacement data can be linked to knee replacement data, allowing for further study of patients with multiple implants.