Extrapolation of parametric survival curves: a case study evaluation using updated data

Charlotte Davies[1], Andrew Briggs[2], Paula Lorgelly[2], Henrik Malchau[3]

[1]Health Economics Group, School of Medicine, University of East Anglia
[2]Public Health and Health Policy, University of Glasgow
[3] Sahlgrenska University Hospital, Sweden

ABSTRACT

Background: With the continuing growth in demand for total hip replacement surgery (THR), there is a clear need for economic evaluation of alternative prostheses. A key requirement for such evaluation is information on survival rates, but typically previous studies have been constrained by the absence of long term data.

Aims: The current paper forms the first part of a PhD, comparing the cost effectiveness of alternative hip/knee prostheses, using UK data from the National Joint Register, also incorporating the indirect costs of patients and society.

Methods: In this first step, however, we focus narrowly on assessing the reliability of methods for extrapolating survival curves to predict lifetime costs and benefits of an intervention, where data are relatively scarce. We return to a previous study by Briggs et al (2004)[1], but now with an extended dataset from the original source, the Swedish Hip Registry. We replicate the analysis, with the benefit of eight years' extra data, employing simple non-parametric methods, before re-examining the chosen parametric option of the Weibull distribution. We assess (i) success in predicting the actual outcomes revealed by more recent data, (ii) the robustness of its estimated parameters; (iii) revised predictions about long-term survival.

Discussion: It is hoped that this is relevant not just to THR but also more generally to the debate about forecasting future survival outcomes using parametric methods with only short follow up data.

Corresponding Author:
Charlotte Davies, Health Economics Group, School of Medicine, University of East Anglia, Norwich, NR4 7TJ, charlotte.davies@uea.ac.uk, Tel: 01603 593665

1 Introduction

1.1 Background

Total hip replacement & alternative prostheses

Total Hip Replacement surgery (THR) is one of the most commonly used interventions to treat joint failure. The beneficial effects of surgery can be measured in terms of improved quality of life of the patient and the durability of the hip prosthesis before further revision surgery is needed. In the long-term, the primary reason for prosthesis failure is loosening of the prosthesis itself. Revision surgery is more expensive than primary surgery and is reported to involve greater technical difficulty with inferior clinical outcomes [2].

In 1998 more than 60 hip prostheses manufactured by 19 companies were listed on the market [3] with total NHS expenditure on hip prostheses alone of approximately £53 million [4] (£70 million in 2007/8 prices[5]). By 2006 the number of hip prostheses on the market had risen to 155 brands of acetabular cups and 176 brands of femoral stems, manufactured by 33 companies [6]. The prosthesis manufacturing industry has responded to the increase in demand for THR surgery by investing significant amounts of money in developing new, more durable, prostheses.

In the UK, The Orthopaedic Data Evaluation Panel [7] provides a rating for prostheses based on data submitted by the manufacturers, for example, the Charnley cemented cup and stem both have a rating of 10A [6], referring to strong clinical evidence of effectiveness at 10 years (the NICE benchmark [8]).

Survival analysis and health economics

Survival analysis is concerned with the time until the occurrence of an event (in this case, revision surgery) and is widely used in medical research. Right censoring of the data occurs where many patients in the data set have not had the event of interest, and standard non-parametric survival analysis methods can be used to handle this, such as the Kaplan Meier survivor function. The Cox proportional hazards, although widely used, is of limited use for extrapolation purposes because it leaves the baseline hazard function unspecified. Instead, parametric models which assume a baseline hazard can be fitted to the data and used to extrapolate into the future. There are a range of alternative parametric distributions which can be used which vary in how the baseline hazard is assumed.

Comparing the survival rates of the different prostheses implanted in THR surgery is necessary in order to assess their relative cost-effectiveness [8]. However, long term survival rates are rarely known. Very few randomised controlled trials comparing one or more prostheses have been conducted as they require large numbers of patients, and a long follow-up period before the survival rate can be accurately assessed. Thus there is often a need to extrapolate models estimated on relatively short time periods to make predictions over a lifetime horizon.

1.2 Objectives

The objective of this paper is to assess the effectiveness of survival analysis in projecting future revision rates beyond the sample estimation period. We use a previous well-known study[1,9] as a case study, for which we have been able to extend the original database to include eight years more data. Specifically, we assess how well extrapolations of the original estimated parametric curves predict the actual outcomes over the subsequent years, and if they do not (which is the case here), we try to understand what is the cause of non-robustness. In particular, was the specific distribution chosen (the Weibull) inappropriate, and an alternative performs better? Or were the estimated parameters of the Weibull distribution unstable, i.e. the new data can still be accurately described by a Weibull, but with significantly different parameters than those estimated on the original data? Or is there some clinical explanation (in terms of changes in the effectiveness of the prostheses, or a significantly different cohort of patients in the later period?)

2 Briggs et al (2003)[9] and the Swedish Hip Registry (hereafter SR)

The purpose of Briggs et al [1,9] (hereafter AB) was to build a probabilistic cost-effectiveness model for primary total hip replacement. It employed data from the Swedish National Hip Arthroplasty Register (SR) to generate transition probabilities for use in a Markov model to evaluate the cost-effectiveness of two alternative prostheses – Charnley and Spectron. Their analysis also modelled health outcomes in terms of Quality Adjusted Life Years (QALYs) and included cost data from the prosthesis manufacturers and published NHS reference costs. The purpose of the current study is confined to just the survival analysis component of Briggs et al.

The registry includes data of all hip replacement operations performed in Sweden, but AB concentrate entirely on the Charnley and Spectron hip prosthesis, selected as illustrations of cemented and cementless prostheses. Their period of study was 1992-2000 and includes data on 20,495 patients: 18,505 (90%) received the Charnley and 1,990 (10%) the Spectron. Their survival analysis employed simple non-parametric methods, and an appropriate parametric form of the hazard function and consideration of dual hazard functions to generate a 'bathtub' failure curve for extrapolation.

The Swedish Hip Arthroplasty Register (SR) was established in 1979. From 1992 onwards the data on primary hip replacement was reported on an individual rather than aggregate form. It is one of the pioneering and most widely used joint registries, and has been followed by many more national registries, in such countries as Canada, Norway, Australia, New Zealand, Denmark and most recently in England and Wales (the National Joint Registry, established in 2003).

The SR contains data on prosthesis type, surgical procedure, patient characteristics, a number of health outcome measures (such as the EQ-5D) and time to event of interest (prosthesis failure and revision surgery).

The SR's 2007 annual report [10] shows that 284,630 primary THR were carried out between 1979 and 2007. Prosthesis data is reported separately for the cup and the stem components as well as according to prosthesis type: cemented/cementless/hybrid; survival rates are reported at 5 and 10 years. The 2007 annual report tabulates the 15 most implanted prostheses for cup and stem, for which the Charnley is the 4th most common, with 6.1% of the market share (1979-2007). However, Table 1 shows that its rate of implantation has declined sharply since 2001. In comparison, the Spectron is not listed as one of the leading prostheses over the period as a whole, but in comparison to the Charnley, its use has increased since 1995. This is in keeping with the reported increased implantation of cementless stems, accounting for 1/5th of the total implantation by 2007 ([10], p.11). The Lubinus All-Poly

(Lubinus SP II) is reported to account for 36.1% of the total market share and with implantation year on year seeming fairly constant.

3 Replicating the original dataset and results

For the present paper, we have had access to data on all patients in the SR receiving either Charnley or Spectron over the years 1992-2007, thereby extending AB's database by eight more years.

As a first step, we attempted to recover from this 16 year extended database, the patients who would have appeared in the original AB dataset (the replication of this dataset is called repAB). These would be those patients who had had their primary surgery (Charnley cup and stem or Spectron cup and stem) by 2000, but recording as revisions only those who had revised by 2000.

In the event, it was not possible to literally duplicate the original AB dataset, because there is some uncertainty as to when exactly in 2000 the AB dataset ends, and (more important) the Registry appears to have changed its classification scheme for different variants of the Spectron cup component, such that the grouping used in AB is no longer available. This has the effect of increasing the number of Spectron implants (to 2949 compared to the 1990 reported in AB). Comparability was closest by employing 31st December 1999 as the cut-off point, the number of Charnley implants in repAB is almost identical to the number recorded in AB (see Table 2).

Table 2 shows that, with the exception of the increased number of Spectron implants, our data for 1992-1999 displays very similar descriptive statistics on patient characteristics – in terms of age, gender and initial diagnosis. The overall revision rates are also almost identical (3% as opposed to 2.8%.)

With broad comparability of the patient characteristics of the two datasets confirmed, we next attempted to replicate the original non-parametric survival analysis. Figures 1 and 2 show the Kaplan-Meier survival curves by prosthesis for the original analysis (AB) and for repAB (Figures 1 and 2 respectively): these appear to be virtually indistinguishable from visual inspection of the curves (although the survival curve for the Charnley is slightly lower in Figure 2). Crucially, as in AB, there is a clear separation of the two survival curves. The significance of this separation is confirmed using a log-rank test for equality of survivor functions, which shows a highly significant difference ($p<0.0000$). Table 3 compares the three and five year survival rates from the Kaplan-Meier in the repAB dataset with the original results (AB).

Examination of the impact of the sample covariates was then conducted (as in AB) by including them in a Cox Proportional Hazards model (Table 4). As can be seen, the hazard ratios are almost identical for age, gender and fracture in the two datasets, with all coefficients highly significant ($p<0.0001$). Most important, the hazard ratio on the Spectron in the replicated dataset still indicates a highly reduced risk of revision, although the coefficient is now slightly smaller than in the AB analysis (0.366 – repAB, 0.435 – AB).

As in AB, we next quantify the baseline risk and extrapolate beyond the observed eight year data to a lifetime time period by conducting parametric analysis, employing four alternative distributions. As reported by AB, the Weibull distribution continues to outperform the alternatives: exponential; Gompertz; log-logistic and log-normal. However, again as in AB, the fits of the four distributions are very close, in terms of Akaike's information criterion (AIC) and when plotting the Cox-Snell residuals. As with the original analysis, the estimated Weibull (see Table 5) is monotonically decreasing over time (γ= 0.872), with increasingly lower hazard rates in the future. To aid interpretation below, recall that the Weibull distribution is characterised by two parameters: λ (scale) and γ (shape), with a hazard function: $h(t) = \lambda\gamma t^{\gamma-1}$, which reduces to the exponential distribution ($h(t) = \lambda$), if γ =1.

Finally, the data were split into early and late failures in order to test whether we could replicate the ‘bath-tub’ shaped failure curve discussed in the AB analysis (initially high failure rates, stabilising then rising again with ‘wear and tear’). We classified reason for revision surgery according to the AB categories and found the same results (59% late failures) for reasons for revision. We then estimated two separate Weibull functions for late and early failures (Table 6). Exactly as in AB, early failures have a monotonically decreasing hazard function (γ = 0.5) while late failures have an increasing hazard function (γ=1.5), i.e. there is a failure curve with properties which are 'bath-tub' shaped exactly as in the AB analysis.

Although it is not possible to identify a sample which corresponds *exactly* to the AB analysis, we have identified a sample, for 1992-9, which generates very similar results to AB, in terms of the Kaplan-Meier, Cox proportional and Weibull survivor functions. Most importantly, this sample leads to exactly the same headline conclusion as in AB: in terms of survival over the first 8 years, the Spectron is unambiguously and significantly superior to the Charnley. This replication of AB's results justifies the analysis in the following sections, in which we assess the success of extrapolations from the above 1992-9 Weibull function over the years which followed, 2000-7.

4 Re-estimation over the new extended dataset (1992-2007)

The full dataset contained data on 31,925 patients undergoing a primary total hip replacement (23,261 received a Charnley and 8664 received a Spectron). The summary statistics are reported in Table 7[1] and show that the characteristics of the patients are almost unchanged for age, gender and initial diagnosis when compared with the dataset for 1992-7 (Table 2). However, there are clearly more Spectron implants relative to Charnleys implanted – as might be expected, given the sharp decline in the use of the Charnley as shown earlier in the SR (see Table 1.)

Kaplan-Meier survival curves for the two prostheses are reported in Figure 2, and these cross at about 10 years. This result is quite startling as it clearly shows that, contrary to expectations from AB, the Spectron survives **less** well after 10 years, and increasingly so thereafter. In fact, this finding is consistent with the published SR annual report, which shows survival rates at 10 years of 92.7% for the Charnley and 92% for the Spectron. Although the log-rank test for equality of the two survivor functions still shows a significant difference between the two curves at the 2% level ($p>0.001$), this test has no real meaning, given the crossing of the curves.

[1] Once mixed cup and stem observations were excluded

Similarly, when the covariates were included in a Cox proportional hazards model (Table 8), although the Spectron still shows a lower risk of revision (hazard ratio of 0.888), this is now much closer to 1 and not significantly different from 1 at the 5% level. All other coefficients are still highly significant ($p<0.001$)[2]. Again, however, since we know that the Kaplan-Meier curves cross, this result has no real meaning since the proportional hazards assumption is clearly not satisfied.

Turning to parametric methods, the same five distributions were fitted to the data. Judged by the AIC (Table 9), there is little to choose between them, apart from the Log-Normal which again provided the poorest fit. However the Gompertz now has the lowest result AIC, although it is only marginally better than the Weibull.

While the Cox Snell Residuals plots (Figure 4) for the Weibull also suggest a marginally closer fit for the Gompertz, for the purposes of comparison, the analysis of the remainder of the present paper continues to employ the Weibull.

For comparability with AB, we again split the data into late and early failures. As can be seen from Table 10 there are proportionately more late failures in the sixteen year period than in the replicated AB dataset – as would be expected, given the longer time period of observation.

Table 11 shows the results for fitting the Weibull for late and early failures: $\gamma = 0.5$ for early failures (monotonically decreasing hazard) and $\gamma = 1.5$ for late failures (monotonically increasing hazard), confirming the 'bath-tub' effect, as found previously. To illustrate graphically, Figure 5, depicts the implied Weibull hazards for Charnley compared to Spectron for a woman, without an initial diagnosis of fracture aged 60 to 70 for late and early failures. These confirm the message from the Kaplan-Meier above, with a crossing confirming that the Charnley outperforms the Spectron, here the crossing of the survival curves – in this case from about 6 years on for late failures.

[2] In order to check that fracture is still the most important predictor of revision status, the model was run again including all diagnoses, it showed that there were other important predictors of revision status, such as Osteoarthritis, Rheumatoid Arthritis and so on. Treatment interactions between Spectron, age, sex and diagnosis covariates failed to show any significant interaction effects.

Contrary to the headline result in AB, the addition of eight years more data find that it is no longer the case that the survival of the Spectron is superior to the Charnley. After about 10 years, the Charnley survives increasingly better than the Spectron. The Weibull no longer displays superiority of fit to the data compared to other distributions, notably the Gompertz.

5 Assessment of Weibull extrapolations from curves fitted to the first eight years, in the light of eight years more data

The final stage in the analysis was to assess the forecasts (i.e. out of sample extrapolations) of the replicated AB Weibull functions against the actual outcomes for the eight year subsequent years. Since these predictions clearly depend on the characteristics of each individual patient, the most simple method for comparison is to compare against the within-sample fitted values from the full-period Weibull for these years. In order to do this, we compare the Weibull results shown in Tables 6 and 11, for late failure using an illustrative patient (female, age 60, non-fracture.) Figure 6 shows this comparison.

To interpret the two curves, note that for the full dataset, all fitted values are within sample, but for the replicated AB dataset, the fitted values are within-sample for the first eight years, but out of period extrapolations for the second eight. As can be very clearly seen, the two curves increasingly diverge into the later years. Table 12 tabulates the comparison in numerical form, at two-yearly intervals. As can be seen, the difference between the two curves is small in the early years, but increases steadily in the forecast period, such that, by year 16, the extrapolations of the replicated AB dataset over-predict survival by as much as 6 percentage points. Such an 'error' could be very important in any cost-effectiveness evaluation.

The Weibull distribution appears to predict Spectron survival fairly poorly. We considered whether there are any other better performing distributions, but when comparing the Cox Snell residual plots for both the eight year and sixteen year dataset, the Weibull appears to have the best fit, although (as shown in Figure 5), the Weibull itself still does not appear to be a very good fit. We also re-examined the

AICs for the repAB and sixteen year datasets and it appears that none of the other distributions provides a better fit than the Weibull, but nevertheless it is not predicting Spectron survival very well.

6 Possible hypotheses to explain results

It appears then that the extrapolations from the replicated AB dataset yield poor predictions of future revision dates beyond the sample estimation period. We return to our original hypotheses to seek the explanation.

Hypothesis I: Is the Weibull distribution inappropriate for this case study?

There is not strong evidence that any of the alternative distributions suggested by AB would have performed any better. In terms of fit (judged AIC and plots of Cox Snell residuals), the Weibull still performs better than the exponential, log-normal and log-logisitic and no worse that the Gompertz.

Hypothesis II: Is the Weibull distribution still appropriate for this case, but its parameters unstable?

The estimated γ coefficient appears to be robust between the eight and sixteen year data for late revisions (1.48 and 1.45 respectively, see tables 6 and 11). However, the λ coefficient does appear to have changed not due to any major changes in covariate coefficients, but the coefficient on the Spectron itself, which has risen from 0.176 (eight year dataset) to 0.937 (16 year dataset). This suggests perhaps that any change might be due to the prosthesis itself rather than the Weibull distribution. It is clear that the proportional hazards assumption, which is inherent in the way that the Weibull model is estimated, is violated due to the crossing of survival curves for the two prostheses. Allowing the γ shape parameter of the Weibull model to vary by arm (effectively parameterising the γ as well as the λ parameter) would relax the proportional hazards assumption.

7 Discussion

This research is still very much in a work progress and we envisage further work in at least 4 directions. We would be very grateful for any comments or suggestions you might have on the following:

1. At present we have assessed forecasts from the initial period by comparing them with fitted values from the full period. Is there some more recognised approach to evaluating forecasts and actuals in survival analysis?
2. Further examination of the instability of the Spectron coefficient in the estimated Weibull in terms of potential changes of the prosthesis itself.
3. Further work surrounding the visual inspection of the Weibull distribution and the appropriateness of allowing the ancillary (γ) parameter to vary as a function of covariates.
4. Consideration of the Royston Parmar flexible parametric model[11,12]. Would this have predicted any better?

References

1. A Briggs MS, J Dawson, R Fitzpatrick, D Murray and H Malchau. The Use of Probabilistic Decision Models in Technology Assessment - The Case of Total Hip Replacement. *Applied Health Economics and Health Policy 2004;3-4:79-89.*

2. Burns AWR, Bourne RB. (vi) Economics of revision total hip arthroplasty. *Current Orthopaedics 2006;20-3:203-7.*

3. Murray D W CAJ, Bulstrode C J. Which primary total hip replacement? *J Bone Joint Surg 1995;77-Br:520-7.*

4. NAO. Hip Replacements: Getting it right first time. *Report by the Comptroller and Auditor General 2000;http://www.pasa.nhs.uk/PASAWeb/NHSprocurement/AboutNHSPASA/LandingPage.htm.*

5. Curtis L. *Unit Costs of Health and Social Care.* University of Kent, 2008.

6. NJR. 4th Annual Report. *The National Joint Registry 2006/7;http://www.njrcentre.org.uk/njrcentre/AbouttheNJR/Publicationsandreports/Annualreports/tabid/86/Default.aspx-Online.*

7. Orthopaedic Data Evaluation Panel. http://www.supplychain.nhs.uk/portal/page/portal/Products%20new/Orthopaedics/ODEP%20(Orthopaedic%20Data%20Evaluation%20Panel). NHS Supply Chain.

8. NICE. Guide to the Methods of Technology Appraisal. *National Institute of Clinical Excellence 2004:1 - 54.*

9. Briggs. A, Sculpher. M, Dawson. J, Fitzpatrick. R, Murray. D, Malchau. H. Modelling the cost-effectiveness of primary hip replacement: how cost-effective is the Spectron compared to the Charnley prosthesis? *CHE Technical Paper Series 2003;28:1-51.*

10. Swedish Hip Athroplasty Register. Annual Report 2007. Department of Ortopedics Sahlgrenska University Hospital, 2008.

11. Royston. P, Mahesh.K. B. Parmar. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine 2002;21-15:2175-97.*

12. Royston P. Flexible parametric alternatives to the Cox model, and more. *Stata Journal 2001;1-1:1-28.*

Figure 1: Kaplan-Meier survival curves by prosthesis type for original analysis (AB)[9]

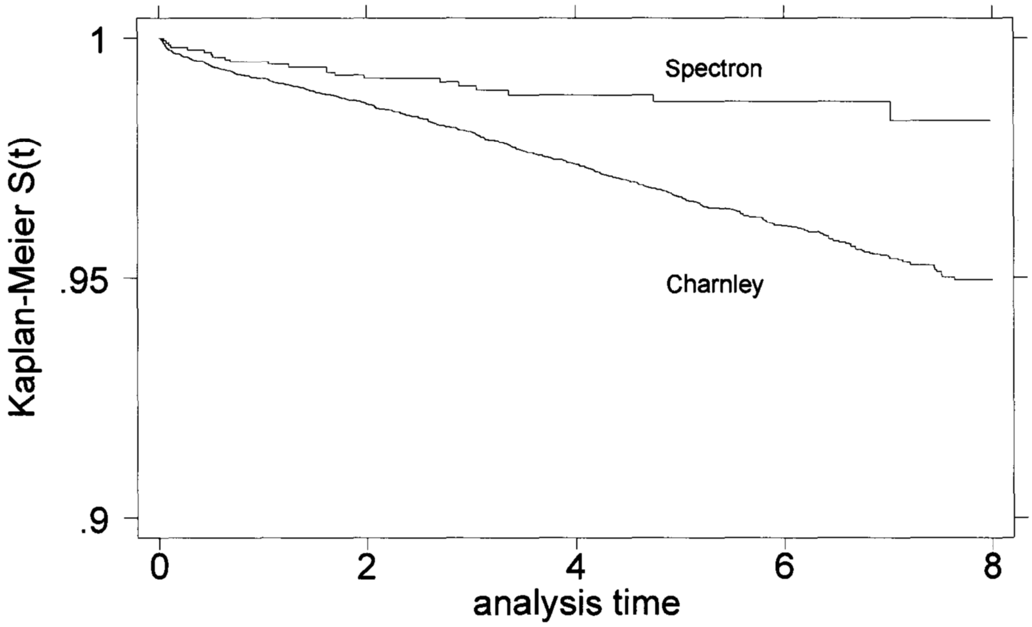


Figure 2: Kaplan-Meier survival curves by prosthesis type 1992-1999: (rep AB)

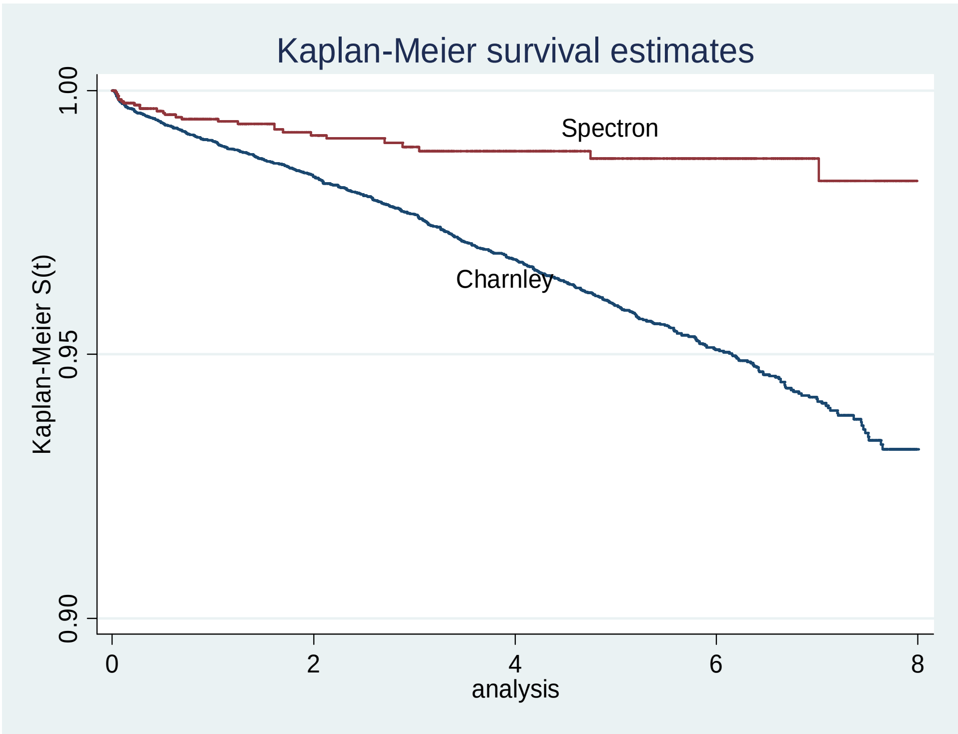


Figure 3: Kaplan-Meier estimate of survivor function by prosthesis type (16 year dataset)

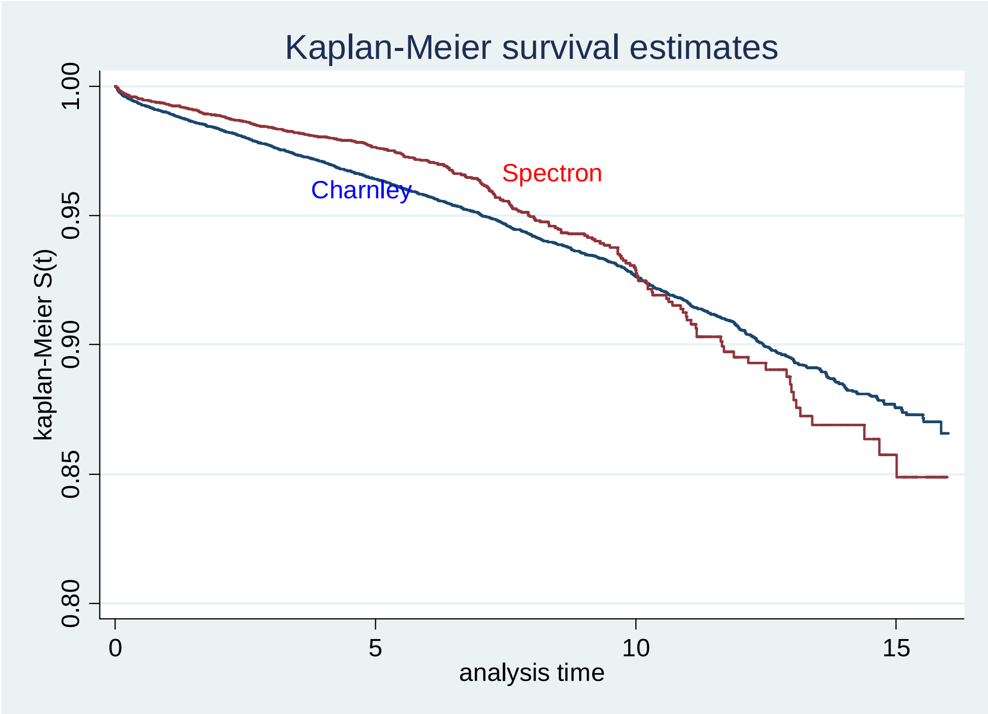


Figure 4: Residual plots for the Weibull and Gompertz parametric survival models (16 year data)

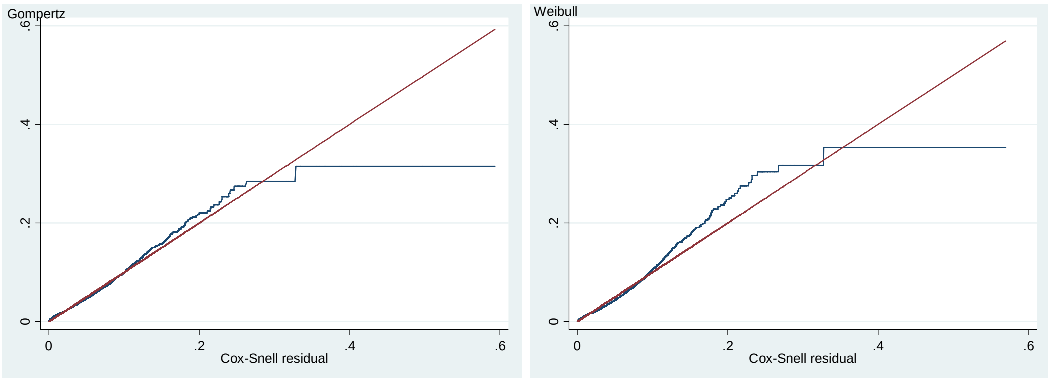
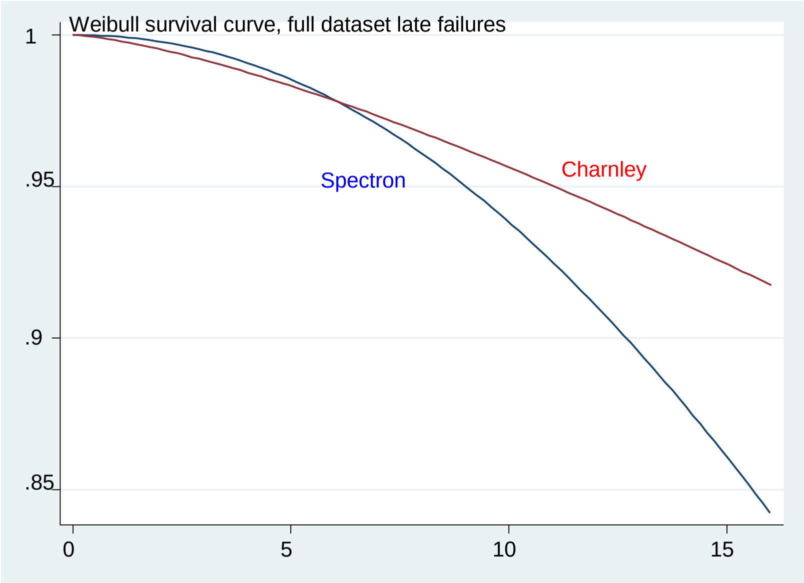


Figure 5: Late and Early failures, survival curves: Charnley vs Spectron for a female, non-fracture, age 60 to 70 (Weibull distribution)



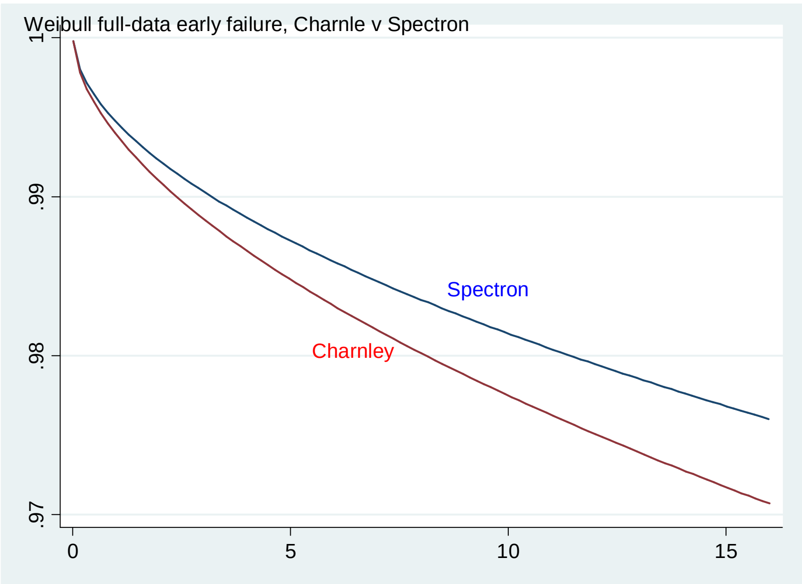


Figure 6: Survival curves for late-failures: Spectron, female, non-fracture, age 60 to 70 (Weibull distribution) - fitted values for sixteen year dataset and extrapolated eight year dataset
(vertical line shows 8 year period, after which all values for the red line are extrapolated)

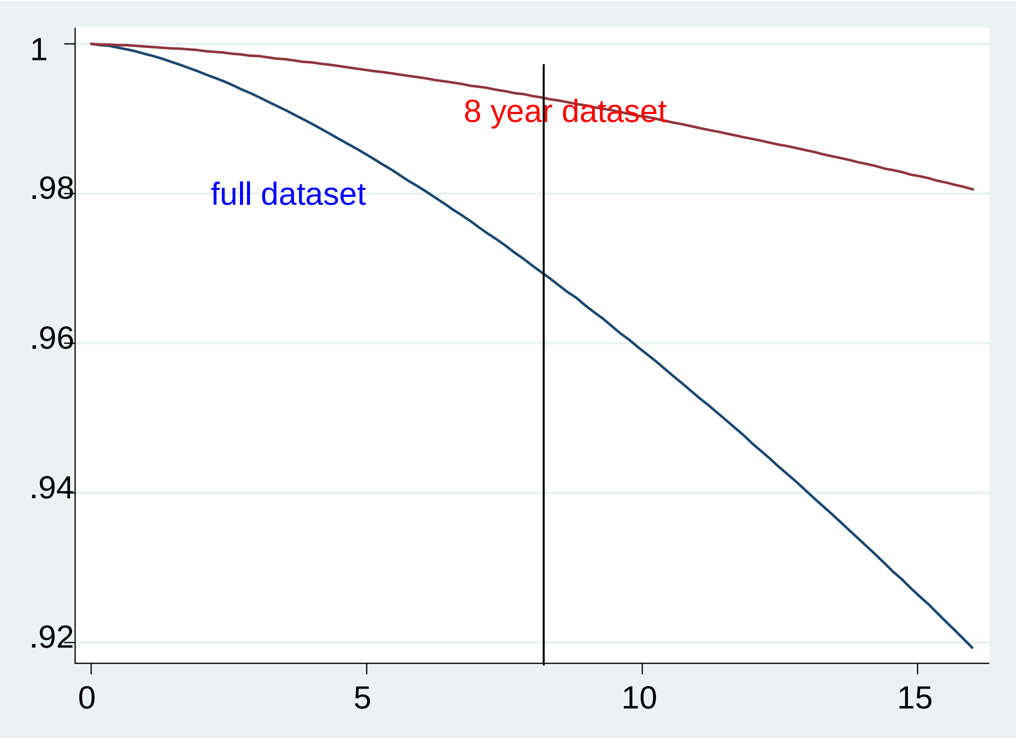


Table 1: Number of Charnley and Spectron implanted recorded in Swedish Hip Arthroplasty Register

Year	Charnley	Spectron	All others
1992	3420	341	1850 (mean/year) (1992–2002)
1993	2685	272	
1994	2373	206	
1995	2052	171	
1996	2402	289	
1997	2120	502	
1998	1912	602	
1999	1772	566	
2000	1622	591	
2001	1600	675	
2002	927	693	
2003	282	889	4712
2004	81	871	5395
2005	8	792	5705
2006	2	677	5529
2007	3	527	5226

Table 2: Summary sample statistics 1992-7 compared with AB (Table 2)

	repAB		AB	
	Charnley	Spectron	Charnley	Spectron
Patients	18,736	2949	18,505	1990
Mean age (sd)	71 (9.2)	72 (9.1)	72 (9.2)	74 (8.1)
Age distribution (%)				
<40 years	72 (0.4)	12 (0.4)	70 (0.4)	5 (0.3)
40-50 years	289 (1.5)	53 (1.8)	264 (1.4)	16 (0.8)
50-60 years	1,607 (8.6)	192 (6.5)	1418 (7.7)	60 (3.0)
60-70 years	5,120 (27.3)	675 (22.9)	4836 (26.1)	391 (19.7)
70-80 years	8,154 (43.5)	1,382 (46.9)	8090 (43.7)	1014 (51.0)
80-90 years	3,341 (17.9)	607 (20.6)	3630 (19.6)	481 (24.2)
>90 years	153 (0.8)	28 (0.9)	197 (1.1)	23 (1.2)
Gender (%)				
Female	12,336 (65)	1,944 (66)	12337 (66.7)	1472 (74.0)
Male	6,500 (35)	1,005 (34)	6168 (33.3)	518 (26.0)
Initial diagnosis (%)				
Osteoarthritis	14,253 (77.8)	2,068 (71.7)	12970 (70.1)	1348 (67.7)
Fracture	1,933 (10.5)	405 (14.1)	1692 (9.1)	319 (16.0)
Other	2,117 (11.58)	409 (14.1)	3843 (20.8)	323 (16.2)
Revisions	629	27	552	22

Table 3: Survival estimates from original (by visual inspection) and replicated datasets

	After 3 years		after 5 years	
	repAB	AB	repAB	AB
Spectron	99	98.5	98.7	98
Charnley	97.5	97.5	96	96.5

Table 4: Cox proportional hazards model

	repAB		AB	
	Hazard Ratio	SE	Hazard Ratio*	SE
Spectron	0.366	0.072	0.435	0.095
Age	0.977	0.003	0.974	0.004
Male	1.738	0.137	1.785	0.150
Fracture	1.733	0.205	1.718	0.221

Table 5: Estimated Weibull model for repAB

	hazard ratio (SE)	p
Spectron	0.361 (0.071)	0.000
Male	1.734 (0.137)	0.000
Age*	0.977 (0.003)	0.000
Fracture	1.718 (0.203)	0.000
Gamma	0.872 (0.030)	

Table 6: Estimated Weibull models for replicated and original AB, separating late and early prosthesis failure

	early failures				late failures			
	repAB		AB		repAB		AB	
	hazard ratio(SE)	p	hazard ratio(SE)	p	hazard ratio(SE)	p	hazard ratio(SE)	p
Spectron	0.607 (0.14)	0.032	0.637 (0.171)	0.093	0.176 (0.67)	0.00	0.258 (0.099)	0.000
Male	1.339 (0.171)	0.023	1.352 (0.183)	0.025	2.07 (0.211)	0.00	2.177 (0.238)	0.000
Age*	0.998 (0.006)	0.807	0.992 (0.007)	0.255	0.965 (0.004)	0.00	0.963 (0.005)	0.000
Fracture	2.025 (0.337)	0.000	2.203 (0.387)	0.000	1.454 (0.248)	0.029	1.303 (0.251)	0.170
Gamma	0.529 (0.030)		0.485 (0.030)		1.48 (0.063)		1.454 (0.069)	

*in the original AB analysis the Age coefficient was age 40

Table 7: Summary sample statistics - 16 year dataset

	Charnley	Spectron
Patients	23,261	8,664
Mean age (sd)	71 (9.3)	72 (9.0)
Age distribution (%)		
<40 years	89 (0.3)	35 (0.4)
40-50 years	342 (1.4)	102 (1.1)
50-60 years	2075 (8.9)	580 (6.6)
60-70 years	6394 (27.4)	2256 (29.5)
70-80 years	9934 (42.7)	3755 (43.3)
80-90 years	4225 (18.1)	1862 (21.4)
>90 years	202 (0.8)	74 (0.8)
Gender (%)		
Female	15210 (65.3)	5667 (65.4)
Male	8051 (34.6)	2997 (34.5)
Initial diagnosis (%)		
Osteoarthritis	18,014 (78.9)	6,465(75.2)
Fracture	2,233 (9.7)	747 (8.6)
Other	2,581 (11.4)	1,385(16.2)
Revisions	1,560	298

Table 8: Cox proportional hazards model – 16 year dataset

Explanatory variable	Hazard Ratio*	SE
Spectron	0.888	0.057
Age	0.970	0.002
Male	1.758	0.083
Fracture	1.712	0.134

Table 9: AIC for parametric distributions – 16 year dataset

Distribution	Exponential	Weibull	LogNormal	LogLogistic	Gompertz
AIC	17180	17176	17435	17198	17156

Note that the gamma distribution was unable to run on current memory level on STATA

Table 10: Using reasons for revision to separate early and late failure

Revision reason	Classification	Percentage (full period)	Percentage (1992-1999)
aseptic loosening	late	69	65
primary deep infection	early	8	11
dislocation	early	11	15
technical problem	early	1	1
other	unknown	11	8
total		100	100

Table 11: Estimated Weibull models for full dataset, separating late and early failure

	early failures		late failures	
	hazard ratio (SE)	p value	hazard ratio (SE)	p value
Spectron	0.794 (0.080)	0.024	0.937 (0.0768)	0.430
Male	1.420 (0.114)	0.000	1.953 (0.113)	0.000
Fracture	2.021 (0.230)	0.000	1.412 (0.115)	0.002
Age	0.999 (0.004)	0.886	0.956 (0.002)	0.000
Gamma	0.565 (0.021)		1.454 (0.002)	

Table 12: Comparing fitted survival rates from the replicated AB period and the full 16 year period*

years	survival rate (estimated from full dataset)	survival rate (estimated from repAB)	difference
2	0.996	0.999	0.002
4	0.989	0.997	0.008
6	0.981	0.995	0.014
8	0.970	0.993	0.022
10	0.959	0.990	0.030
12	0.946	0.987	0.040
14	0.933	0.984	0.050
16	0.919	0.980	0.061

* note that the fitted values are within sample for 1993-9 in the replicated dataset and extrapolations for 2001-2007. All fitted values are within sample for the full dataset