Construction and Validation of a Parametric Model for Predicting Implant Survivorship Beyond Observed Data in Total Joint Arthroplasty

Jennifer Cao[1], Katie Miller[1], Jing Xie[1]
[1]Biomet Orthopedics, 56 East Bell Drive, Warsaw, IN 46581

Abstract

Background In the orthopaedic space, long-term implant survivorship is a critical measure to determine product performance. Construction of a sound statistical model that will predict implant survival beyond the time window of available data is of strong interest and would allow manufacturers to conduct more proactive surveillance on product performance.

Methods An accelerated failure time model was fit to short-term survival data from a multi-center clinical trial for a total joint replacement product. The model was used to estimate the survival probability at all time points, including those outside the time window of available data. The model building, model selection, and model diagnostics were demonstrated. Model validation (both internal and external) was performed.

Results An Exponential model (intercept only) was found to be the best-fit model for the data, and satisfied the model assumptions and goodness-of-fit. The internal validation confirmed that the estimates were relatively stable although censoring patterns appear to affect the estimates to some extent. External validation confirmed that this model reliably estimates survivorship within the range of available follow-up time and for approximately 8 years past this interval. Model-predicted survivorship and associated 95% confidence intervals were calculated within the time window of available data and for 8 years outside this interval.

Conclusions This methodology is a viable alternative to the traditional data acquisition methods, and may allow for faster attainment of long-term survivorship information and hence more proactive surveillance on product performance.

Key Words: Orthopaedic, total joint arthroplasty, survival, revision, exponential distribution, extrapolation, prediction

1. Introduction

1.1 Clinical Rationale

In the orthopaedic medical device industry, there is an increasing demand for clinical data with specific focus on long-term implant survivorship data. This data is needed to support product performance, fulfill post-market surveillance requirements, and gain market acceptance for products. On the other hand, changes in patient population and resulting new demands require device manufacturers to be responsive and continue to advance technology and enhance product designs to address any unmet clinical needs. Often new products are developed largely based on design features and material

characteristics of previous generations of the product, and so in many cases it is reasonable to expect that the performance of the new generation of product is very similar to that of its predecessor.

The combination of these two factors presents a challenge to new product development and launch as manufacturers need to obtain long-term data on new products in a relatively short time frame. Traditionally this is addressed by collecting clinical data on the new device and waiting until an appropriately-sized cohort of patients have been followed to long-term time points to see how the product is performing. However, this takes significant time and resources, and more importantly prevents device manufacturers from taking the new products to the market and providing the best care to patients. This paper explores the possibility of using a predictive model built using short term data to predict survival probabilities beyond the time window of available data.

1.2 Statistical Rationale

In general, survival analysis provides estimation of survival probability within the time window of available data. Extrapolation of statistical models beyond available data must be done with caution, because models built using data within a certain time frame may not continue to follow the same trend. However, predictive modeling has been successfully applied in several areas of clinical research [3,4,7], and when used appropriately can be very helpful. For example, Jackson, et al. [3] described the validation and use a parametric model for predicting median survival times for Cystic Fibrosis birth cohorts. In Chu, et al, a semi parametric method was proposed to extrapolate survival curve to estimate expected years of life lost for cancer patients, under high-censored rates [4]. Royston, et al described methods for external validation of a prognostic survival model, with practical examples using datasets in cancer [7].

In this analysis, a parametric model was used to estimate long-term survivorship for a total joint replacement product. Mid-term data on a cohort of patients followed to 7 years post-surgery was obtained from a multicenter clinical trial ("primary data set") and was used to build this model. Because extrapolation beyond the time of available follow-up can be somewhat tricky and unreliable, the model was carefully validated. The validation taken here includes external validation using long-term data from an older, but fundamentally similar, device. Internal validation was also performed [3,7].

2. Process Overview

The general process and methods described in this paper can be seen as a continuous cycle. Although it begins with a certain amount of short-term data on a newer implant design and long-term data on an older, but fundamentally similar, design, as time goes on, more data will be acquired, and the model can be re-evaluated to see whether it still fits and predicts well with the updated data. In addition, the short-term data on the newer design will eventually mature and become long-term data, and the newer design may eventually be improved upon, resulting in another new device with similar characteristics. In this way, the cycle will repeat.

The process can be summarized in this way:

1) Obtain short or mid-term data on new device, denoted as "Primary Data"

2) Fit accelerated failure time model to the Primary survival data
3) Perform internal validation of model using the same data used to fit model
4) Obtain long-term results on older generation device, denoted as "Secondary Data"
5) Perform external validation of model using long-term secondary data
6) Use predictive model to obtain long-term predictions of product performance

This paper will use a real-data example to describe and demonstrate each of these steps.

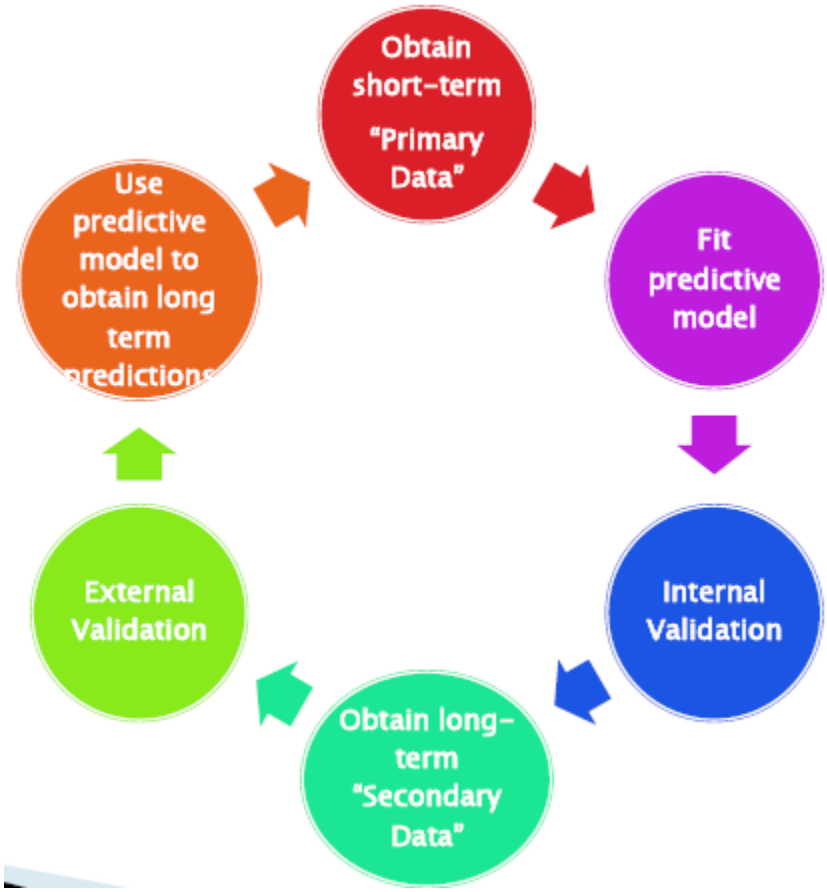

Figure 1: Overview of Predictive Modeling Process

3. Patient Populations – Primary and Secondary Data

The patient population used in external validation ("secondary data") should be similar to the one used for model building ("primary data"). The relevant patient factors for both groups used in this data example are described in Table 1. It can be seen that the primary and secondary populations are similar with regard to primary indication, gender, age, BMI, and percentage of censored values.

Table 1: Patient Demographic Information for Primary and Secondary Data

	Primary	***Secondary***
Number of Cases	958	15207
Female/Male	609/349	9151/6056
Age at Surgery	69.1+/-9.48	68.4+/-9.28
BMI at Surgery	31.8+/-6.10	37.1+/-8.46
Percent Censoring	98.2%	96.1%
Primary Indication		
Osteoarthritis	96.7%	97.5%
Rheumatoid arthritis	2.2%	2.1%
Other	1.1%	0.36%

The log rank test comparing the KM curves for primary and secondary datasets from surgery to Year 4 (the time of the last "event" for the primary data set) gives a p-value of 0.1556. This gives some evidence that the two generation devices are not statistically significantly different with respect to implant survivorship, which is also seen graphically in Figure 2.

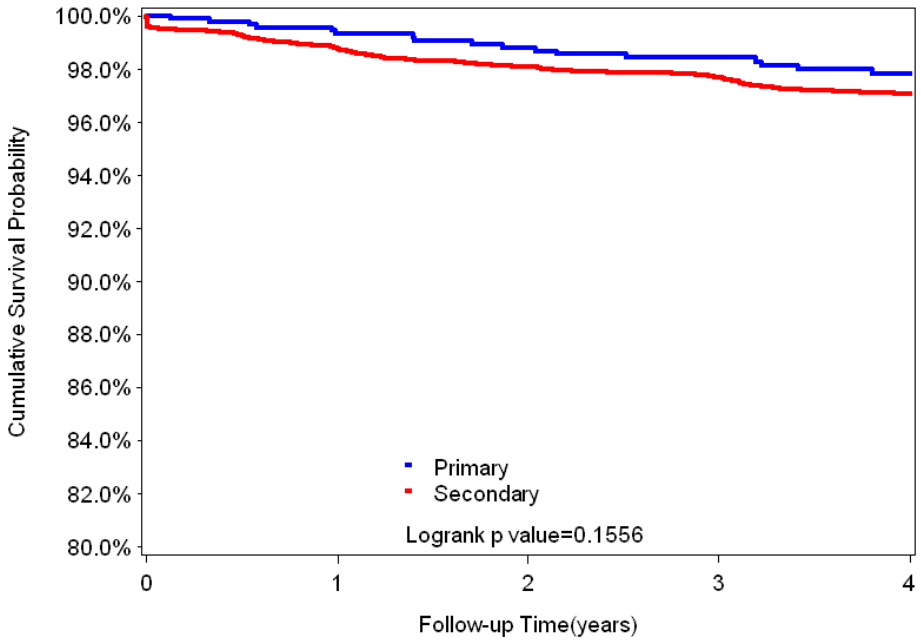

Figure 2: Kaplan Meier Plots for Primary and Secondary Data Censored at Latest Event Time for Primary Data (4 Years)

4. Statistical Model Building

4.1 Model Building

Parametric survival regression modeling using distributions such as Weibull, exponential, log-normal, and log-logistic, are often used for survival predictions [1,4,5]. In survival analysis, the parametric regression models have this form [1]:

$$Y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sigma\varepsilon$$

Where Y is either T (time to failure) or log(T), x_j are covariates, ε is a random error, β and σ are parameters to be estimated. When log(t) is used, the model is known as an accelerated failure time (AFT) model. In SAS, the LIFEREG procedure is used to calculate the maximum likelihood estimators of the parameters. Akaike Information Criterion (AIC) and Schwarz's Bayesian Criterion (SBC) are the criteria used for comparing model fit among the different types of AFT models.

Using the primary dataset, ten survival distribution functions were fitted and compared using PROC LIFEREG in SAS. Time is calculated as years from date of surgery to date of revision (uncensored) or date of last follow up (censored). The exponential distribution gave the smallest AIC and SBC, suggesting that the exponential distribution is the best fit for the data among the ten survival distribution functions selected.

Assuming the data follows an exponential distribution as suggested above, patient risk factors including gender, age at time of surgery, BMI at time of surgery, primary indication for surgery, and joint side were checked. A backward elimination model building process was utilized to select significant covariates for inclusion in the model. If more than one main effect remained in the model, interaction terms were added one at a time to check for significant interactions [6]. For this model, the order of removal for the selected covariates (based on the p-value > 0.05 for the Type III Analysis of Effects, Wald Chi-square test) was: primary indication, BMI, age at surgery, joint side, and gender (Table 2).

Table 2: Type III Analysis of Effects based on Wald Chi-Square Test

Predictor	*Primary Indication*		*BMI*	*Age at surgery*	*Joint Side*	*Gender*
	DJD	*OA*				
P value	0.99	0.99	0.685	0.585	0.190	0.159

None of the covariates were statistically significant predictors. The exponential model, fit without any covariates, has

$$\lambda = 5.444$$

The corresponding predicted survivorships can be calculated by:

$$S(t)\ \text{hat} = \exp(-t*\exp(-5.444))$$

Where t = time in years from surgery until revision, death, or date of most recent follow-up.

4.2 Model Diagnostics

The Exponential distribution assumes a constant hazard rate. Based on a plot of log(-log(S(t))) against log(t), it is reasonable to assume that there is constant hazard rate. In addition, the Cox-Snell residuals were distributed as approximately unit exponential; therefore the exponential model without predictors appears to be adequate. The plot of scaled score residuals against follow-up time in years was used to gauge the influence of each observation on the intercept. The Martingale residuals plot was also obtained to check the influence of individual observations on the model. No extreme values of the residuals were observed; therefore it can be concluded that no influential observations were found.

5. Statistical Model Validation

5.1 Internal Validation

In order to assess the performance of the model in predicting long-term survival after only a short duration of follow-up with heavily censored data, the impact of censoring needed to be explored. The observed survivorship data was censored beginning at 1 year, so that only follow-up from Year 0 to Year 1 were used in the model. The estimated 10-year and 15-year survivorships and associated 95% confidence intervals were calculated. The follow-up period was extended a year at a time until year 7 (the last interval for which data from the primary data set was available) [3]. A new model was fit each time and predicted 10-year and 15-year survivorships were re-estimated based on the model. Figure 3a and 3b show the predicted 10-year and 15-year survivorship (respectively) for each iteration. The 10-year predicted survival ranges from 88% when data is censored at Year 3, to 95% when data is censored at Year 7. The 15-year predicted survival ranges from 82% when data is censored at Year 4, to 93% when data is censored at Year 7. The confidence intervals (calculated using the Delta Method) overlap, with the exception of Year 4 and Year 7.

10-year and 15-year survivorship estimates are lowest when the censoring occurs at Year 3 and Year 4 (respectively). Year 4 is also when the last event occurs within the primary data set; although follow-up still occurred from years 5-7, no revisions occurred during this time. In addition, censoring has increased to ~30% by year 4 and to ~50% by year 5.

Although the model predictions are relatively stable, results show that the amount and pattern of censoring likely has some impact on the model predictions. External validation of the predictions based on all data from the primary dataset is considered in order to make a better conclusion about whether the estimates produced by the model are accurate for predicting long-term survivorship.

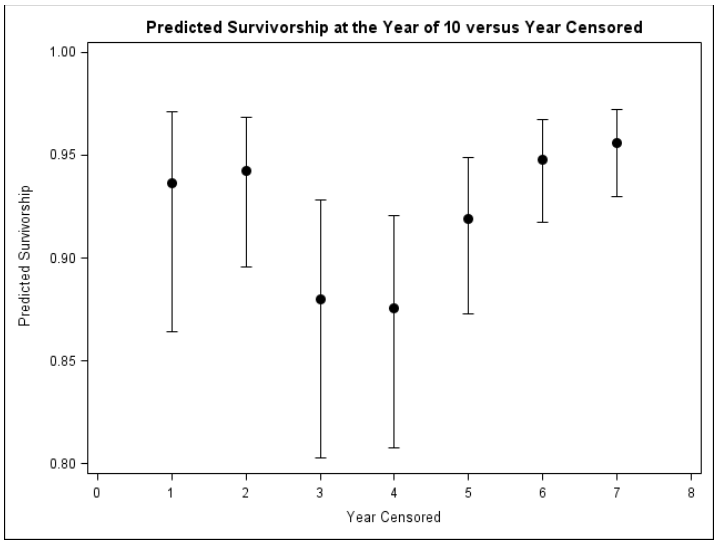

Figure 3a: Iterative prediction of 10-year survival (per Jackson et al [3])

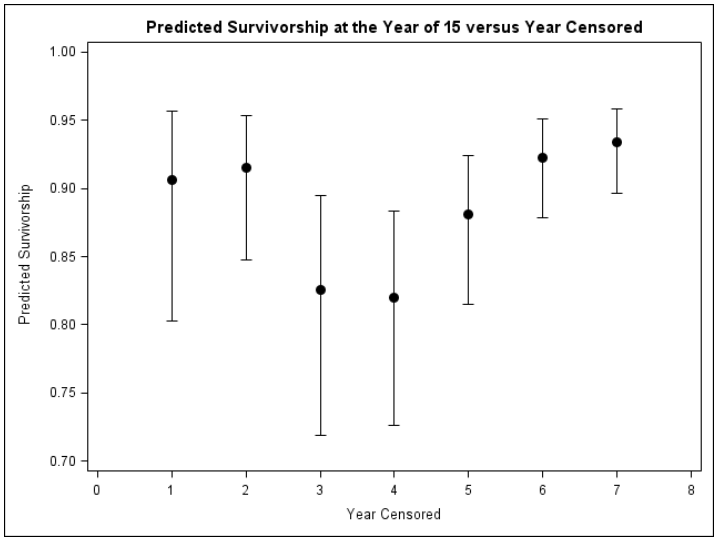

Figure 3b: Iterative prediction of 15-year survival (per Jackson et al [3])

5.2 External Validation

In practice, the best way to validate a model is to use an external source of data to check the predictive ability of the model, as opposed to internal validation methods such as data splitting, cross-validation and bootstrap resampling on the primary data set. Unfortunately, there is little statistical literature on formal techniques for external validation of models for time-to-event data [3]. In general, external validation uses an independent dataset in which a completely different cohort of patients is used from the one used for building the model. The patients should have a set of relevant factors in common, including the same clinical condition within similar settings and the same definition of the time-to-event outcome [7]. In addition, similar patient populations as

indicated by similar distributions of primary indication for surgery, gender, age, BMI are needed.

Using the estimated Exponential model parameter from the model built using the primary dataset, estimated survival probabilities for the secondary dataset were obtained and compared with the actual Kaplan-Meier estimates for the secondary dataset to assess the predictive ability of the exponential model. The secondary data set has a longest observed event time of approximately 24 years, compared with 7 years of the longest event time of the primary set of data used to build the predictive model. Figures 4 and 5 show the survival estimates based on the Exponential and Kaplan-Meier models for the primary and secondary data sets, respectively. It can be seen from the plots that the estimated survivorships from the model match the observed Kaplan-Meier very well within the time window of follow-up (approximately 7 years) and for a few years outside this window. Not surprisingly, as the predictions get farther from the observed window of data, the estimates are not as closely matched by the external validation model. Also adding to the uncertainty at these later time points are the many censored observations caused by low patient follow-up at later time points in both primary and secondary datasets.

In addition, the yearly average relative bias between Kaplan-Meier estimates and Exponential Regression Estimates of Survivorship were calculated, and are listed in Table 3 [4]. Relative bias is calculated by dividing Kaplan-Meier estimate by the difference between Kaplan-Meier estimate and exponential estimate. The average relative biases before 17 years are less than 5%, which is very small [4,9], and much smaller than those of longer follow-up time in years. We would conclude with a reasonable level of confidence that the survivorship can be predicted to at least a few years beyond the time window of observed data.

Table 3: Yearly Estimates of Average Relative Bias Between Kaplan-Meier Estimates and Exponential Regression Estimates of Survivorship for the Secondary Data Set

Follow-up Time in Years	*Average Relative Bias(%)*	*Follow-up Time in Years*	*Average Relative Bias(%)*
0	0.4991374	12	1.0730166
1	0.8018126	13	1.5535007
2	0.6669383	14	2.1144513
3	0.7812099	15	2.9582373
4	0.5582979	16	3.631268
5	0.5254385	17	5.0697704
6	0.4197201	18	7.0696639
7	0.6494154	19	8.4293814
8	0.6563039	20	10.328814
9	0.5875496	21	12.984082
10	0.6592852	22	18.423965

Follow-up Time in Years	*Average Relative Bias(%)*	*Follow-up Time in Years*	*Average Relative Bias(%)*
11	0.8753437	23	27.867682

6. Point Estimates and Confidence Intervals for Predicted Exponential Survivorship

Table 4 shows the predicted survivorship at selected time periods. Years 1-5 are within the time window of available data, and years 10 and 15 are outside this window. At 10 years, the point estimate for survivorship is 0.958, and at 15 years the estimate is 0.937.

Table 4: Predicted Survivorship at Selected Time Points Using Exponential Regression Parameter Estimates

Follow-up time in Years	*Predicted Survivorship*
1	0.996
2	0.991
3	0.987
4	0.983
5	0.979
10	0.958
15	0.937

Pointwise confidence intervals for the predicted exponential survivorships were constructed for the primary data set. These were developed by transforming S(t) onto a scale which more closely follows a normal distribution [3], that is, log(-log(S(t))), as proposed by Borgan and Liestol 1990[8]. The variance of S(t) was derived using the Delta method [2] as shown below. Note that the parameter β_0 and its standard error are the estimates obtained from the model fitted on the primary data set.

For the intercept only model,

Since

$$S(t) = \exp(-t \times \exp(-\beta_0))$$

Using the Delta method, we have,

$$se(S(t) = \exp(-t \times \exp(-\beta_0)) \times t \times \exp(-\beta_0) \times se(\beta_0)$$

The confidence intervals based on log(-log(S(t))) is

$$[S(t)^{1/\theta}, S(t)^{\theta}]$$

Where $\theta = \exp[z_{\alpha/2}(se(S(t))/[S(t) \times \log(S(t))]]$

The estimated survivorships and the confidence intervals are plotted as in Figure 4, along with Kaplan-Meier estimates and associated confidence intervals using Greenwood's formula to estimate the variance of S(t). As the number of years from surgery increases, the exponential model increasingly over-estimates the Kaplan Meier survivorship. This may be due in part to the patient drop-out rate, which increases considerably with time. This would be the case, for example, if patients without issues and who are doing fine are less likely to return for follow-up visits than patients who are not doing well, which would result in the Kaplan Meier method underestimating the true survivorship at these time points. This is, however, difficult to prove, since data on these patients is not available.

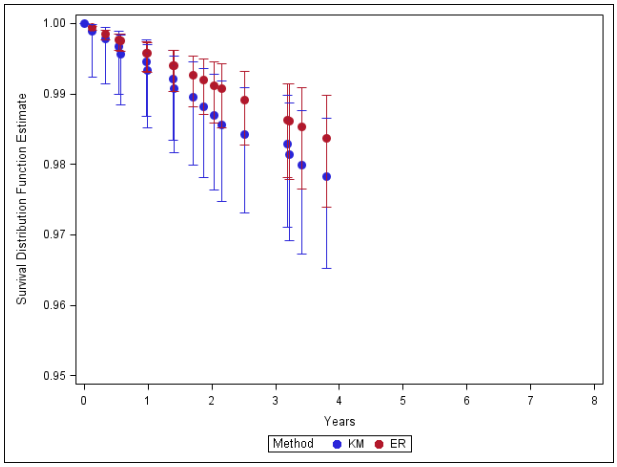

Figure 4: Estimated Survivorships and 95% Confidence Intervals Constructed Using Kaplan Meier and Exponential Regression Method for Primary Data

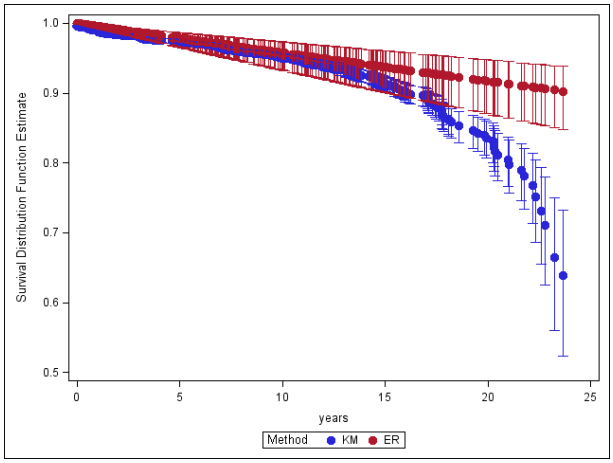

Figure 5: Estimated Survivorships and 95% Confidence Intervals Constructed Using Kaplan Meier and Exponential Regression Method for Secondary Data

7. General Challenges

In general, survivorship data on total joint replacement products often has a very high percentage of censored values. A primary reason for this is that due to the high success rates seen in total joint replacement, many subjects do not actually have the "event", which is revision of the implant for any reason. Although this is good from the

standpoint that the joint replacements are performing as intended, from a statistical analysis perspective, this means that we have less information about the times to event with which to build our predictive model, and therefore more uncertainty about the distribution of times to event at the future time points.

Another common reason for censored values is that there are often many patients who do not return for the later follow up visits. A plausible reason for this is that in many cases, they are doing well and feel they do not need to be seen. Of course, this cannot be definitively proven. As a result, the length of time into the future for which the predictive model can be expected to be accurate is somewhat limited by the adequacy of the patient follow-up at these later time points.

8. Conclusion

The parametric model built using the exponential distribution for censored survival data from a joint replacement implant and procedure works very well to make predictions for implant survivorship for the newer generation of device within the range of available follow-up time as well as for prediction to longer time windows. This model is viable and relatively accurate to project the survival of the implant into the near future beyond the observed time windows. Refinement and update of this predictive model can be made when new data is collected as time progresses. In general, this method can potentially serve as an alternative to the traditional clinical data acquisition methods in order to allow for faster attainment of long-term survivorship information on newer implant designs. It provides a useful tool to predict the survivorship of a product before long-term data is available and can help to more proactively assess product performance.

References

[1] Vadim P. Predictive Modeling Using Survival Analysis. Northeast SAS Users Group 18.

[2] Mahesb k.B. P, David M.(1995). Survival Analysis: A Practical Approach. John Wiley, New York.

[3] Jackson A., Patricia F. etc. Validation and Use of a Parametric Model for Projecting Cystic Fibrosis Survivorship beyond Observed Data: a Birth Cohort Analysis. Thorax 2011;66;674-679

[4] Chu, Po Ching, Jung D.W, Jing Sh.H,Yu Y.Ch. Estimation of Life Expectancy and the Expected Years of Life Lost in Patients with Major Cancers: Extrapolation of Survival Curves under High-Censored Rates. Value in Health 2008; 11:1102-1109.

[5] John P.K, Melvin L.M. (1998). Survival Analysis: Techniques for Censored and Truncated Data. Springer, New York.

[6] Survival Analysis in SAS. http://www.ats.ucla.edu/stat/sas/seminars/sas_survival/

[7] Royston, Patrick, Mahesh K B P. External validation and updating of a

prognostic survival model. Research Report No.307, Department of Statistical Science, University College London. March 2010

[8] Borgan, Ø. and Liestøl, K. (1990), A Note on Confidence Interval and Bands for the Survival Curves Based on Transformations. ***Scandinavian Journal of Statistics,*** 18, 35–41.

[9] Zheng, Yingye (2005). Partly Conditional Survival Models for Longitudinal Data. *Biometrics*, 61, 379-391.