Variation in outcome and ranking of hospitals: An analysis from the Swedish knee arthroplasty register

Otto Robertsson, Jonas Ranstam & Lars Lidgren

Variation in outcome and ranking of hospitals

An analysis from the Swedish Knee Arthroplasty Register

Otto Robertsson[1], Jonas Ranstam[2] and Lars Lidgren[1,2]

[1]Department of Orthopedics and [2]National Swedish Competence Center for Musculoskeletal Disorders, Lund University Hospital, SE-221 85 Lund, Sweden
Correspondence OR: otto.robertsson@ort.lu.se
Submitted 05-08-18. Accepted 05-11-30

Background   Hospital-specific variation in outcome is generally considered to be an important source of information for clinical improvement. We have measured the magnitude of this variation.

Methods   We determined the revision risk in 37,642 cemented primary total knee arthroplasties inserted as a result of osteoarthritis from 1993 through 2002 at 93 hospitals in Sweden. We used 2 essentially different methods to estimate risk of revision: a fixed-effects model (Cox's proportional hazards model) and a random-effects model (shared gamma frailty model).

Results   The 2 models ranked hospitals differently. As expected, the fixed-effects model provided more dispersed estimates of hospital-specific revision rates. In contrast to the random-effects model, chance events can easily cause overly optimistic or pessimistic outcomes in the fixed-effects model. Although the revision risk varied significantly between hospitals, the overall revision risk was still low.

Interpretation   Assessment of variation in outcome is an important instrument in the continuing effort to improve clinical care. However, regarding revision rate after knee arthroplasty, we do not believe that such analyses necessarily provide valid information on the current quality of care. We question their value as information source for seeking personal healthcare.

■

The Swedish Knee Arthroplasty Register (SKAR) was established in 1975 (Bauer et al. 1980). Details of knee arthroplasties that have been performed are regularly reported by participating clinics to the registry in Lund, which prospectively monitors primary arthroplasties regarding revision.

After validation and updating of the register in 1997 (Robertsson et al. 1999), 94% of the revisions of the prospectively followed patients were considered to have been accounted for. Prior to the update, 80–85% of all primary arthroplasties were estimated to have been included, but with improved routines and cooperation the participation has increased. Checking against official databases has indicated that since 1997, more than 95% of all operations have been reported to the register.

Using survival statistics, the Register has been able to provide information on surgical results and how these vary depending on the implant used, the type of disease, the surgical technique, and so on. In recent years, the quality of healthcare provided has gained more attention, which has led the Swedish National Board of Health and Welfare to encourage investigations of hospital-specific variability in outcome as a means of quality improvement and guidance.

Two basically different methods can be used to estimate variation in revision risk among hospitals: fixed- and random-effects models. The conventional Cox's proportional hazards model represents a fixed-effects model. One commonly used random-effects counterpart is known as the shared gamma frailty model.

It is known that random-effects models generally produce more reliable estimates of overall variability than fixed-effects models, which overestimate variability (Efron and Morris 1977).

The purpose of this study was to describe variation in revision rates between hospitals and the differences in results of the two different methods when applied to SKAR data, to discuss their potential consequences, and to suggest which method is preferable.

Material and methods

For all 37,636 cemented total knee prostheses operated for osteoarthritis in Sweden between 1993 and 2002 (primary operations), we determined the cumulative revision rate using revision for any reason as endpoint. The dataset consisted of 24,719 prostheses from females and 12,917 from males. Mean age was 72.2 (18–96) years.

937 knees were revised during a mean follow up of 3.7 (1–11) years. In all, the dataset included observations on 140,451 prosthesis years. The incidence rate of revision was thus 67 per 10,000 prosthesis years.

Hospital-specific revision rates were estimated using both Cox's proportional hazard's model (a fixed-effects model) and a shared gamma frailty model (a random-effects model). The analyses were performed with prosthesis as analysis unit, disregarding bilaterality phenomena, as validated in an earlier study (Robertsson and Ranstam 2003). The analyses in both models included adjustment for differences in age (as a continuous variable), sex, and use of patellar button.

The methodology has been described in detail by Therneau and Grambsch (2000). A short description follows.

Cox's proportional hazard's model

Cox's proportional hazard's model is often used to analyze survival with adjustment for potential confounding factors. The model can be described as

$$h_i(t) = h_0(t)\exp(\beta^T x_i), \ \ 0 \leq t \leq \infty$$

where t represents time, h0(t) is the ith individual's baseline hazard function, β^T is a vector of coefficients, and xi the ith individual's vector of covariate values, i.e. identifies the hospital where the operation was performed, the patient's age at primary operation, sex, and use of patellar button.

The coefficients of this model, β^T, represent log hazard ratios and are usually presented in exponentiated form, interpreted as relative risks. Hospital effects estimated with this model relate to the mean of hospital-specific effects.

The shared gamma frailty model

Assuming that patients operated at the same hospital share a common hospital-specific risk (or frailty), wj, and that risks of different hospitals are independent, the shared gamma frailty model is written

$$h_{i(j)}(t) = h_0(t)w_j\exp(\beta^T x_i), \ \ 0 \leq \tau \leq \infty$$

where i(j) denotes that individual i has been operated at hospital j. The w's are assumed to follow a Gamma distribution with mean 1 and variance θ.

In this model, hospital effects are estimated using the w's, not the β^Ts, which here identify only the patient's age at primary operation, sex, and use of patellar button. The w's can be interpreted as relative risks directly.

Fixed and random effects

In the Cox's proportional hazards model, hospital effects are estimated as mean effects. In the shared gamma frailty model, they are based on estimation of variance of effects.

The general terms for effect estimates calculated using these principles are random and fixed, respectively. This terminology was used as early as by Eisenhart (1947), although Fisher developed the methodology itself during the 1920s.

Random effects models of hospital-specific risks are based on the assumption that the hospitals studied represent a larger, common, population of hospitals that could, in principle, have been sampled randomly. In fixed-effects models, on the other hand, hospitals are considered to be unique entities instead, representing only themselves.

Effect estimates of random-effects models are "shrunken" as compared to fixed models estimates. This can be thought of as an adjustment for the well-known effect "regression to the mean". The magnitude of shrinkage is precision-dependent; the use of a random effect model thus reduces the problem of overinterpreting randomly-produced high and low risks in hospitals with relatively few patients.

Table 1. Cumulative revision rate for all cemented primary TKA inserted during 1993–2002

Follow-up (year)	Numbers of knees			Cumulative revision rate	
	total	revisions	censored	rate	95% CI
1	37,638	298	6,563	0.009	(0.008–0.010)
2	30,777	285	5,427	0.019	(0.017–0.020)
3	25,065	146	4,685	0.025	(0.023–0.027)
4	20,234	79	4,176	0.029	(0.027–0.031)
5	15,979	47	4,274	0.033	(0.030–0.035)
6	11,658	38	3,575	0.037	(0.034–0.039)
7	8,045	22	2,903	0.040	(0.037–0.043)
8	5,120	12	1,992	0.042	(0.039–0.046)
9	3,116	6	1,692	0.044	(0.041–0.049)
10	1,418	4	1,414	0.050	(0.044–0.057)

The shared gamma frailty model describes hospital effects as deviation from an overall mean value. When using Cox models, hospital effects are usually estimated in relation to a reference hospital. To facilitate comparisons between the two methods, we used Södertälje Hospital as reference for the Cox model because it has about the same relative hospital effect as the frailty model's reference and a relatively large number of observations.

Computing
We calculated crude hospital-specific revision rate estimates using the Kaplan-Meier method and used the lifetable method to calculate overall annual revision risk estimates. Parameter estimates of Cox's proportional hazard's model were calculated using the partial likelihood method, while those of the shared gamma frailty model were calculated with the penalised partial likelihood method. We used a p-value of below 0.05 to indicate statistical significance.

We used the statistical software STATA version 9.0 and R version 2.1 (http://www.r-project.org/) for the calculations.

Results
The 10-year cumulative revision rate for all cemented total knee arthroplasty implants inserted for osteoarthritis in Sweden between 1993 and 2002 was 5.0% (95% CI: 4.4–5.7) (Table 1). The revision risk declined with increasing age by 4.1% (95% CI: 3.4–4.7) for each calendar year. Men

had a 6% higher revision risk than women (95% CI: –7–21) but this was not a statistically significant finding. Use of a patellar button reduced risk by 20% (95% CI: 3–35).

Hospital-specific effect estimates
The study population had been operated on at 93 different hospitals (primary operations). The average number of patients per hospital was 405 (1–1693). Unadjusted cumulative 10-year revision rates varied from 0% to 37% between hospitals. Adjustment for age and sex in a Cox's regression model produced hospital-specific relative revision risk estimates between ~0 and 5.9 (Figure).

As expected, the frailty model shrank the hospital-specific revision risk estimates, which varied between 0.36 and 1.94 (Figure). The hospital variance component of the frailty model was estimated to be 0.181, which was statistically significant (p < 0.001); thus, there was a statistically significant heterogeneity between the hospitals. The 2 models ranked the hospitals differently. Revision risk estimates with 95% confidence intervals and the ranks from the two models are presented in Table 2.
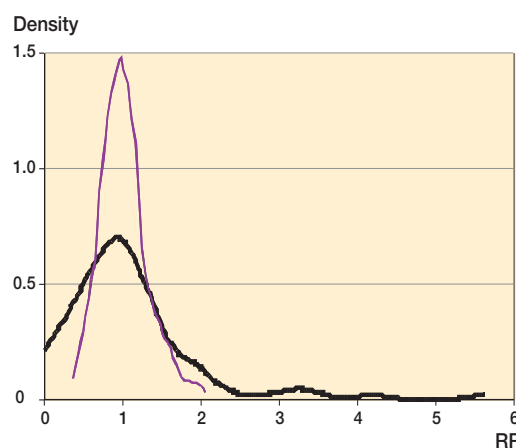
Density plot of fixed-effect (thick line) and random-effect (thin purple line) estimates of hospital-specific effects on revision-free survival, i.e. relative risks. Fixed- effects estimates show greater variation; the "shrinkage" of random effects is apparent.

Table 2. Relative revision risk (RR) and rank per hospital, 1993–2002, with the shared gamma frailty model. Results of Cox's model are included on the right

Code	Hospital	No. of prostheses	Frailty model RR	95% CI	Rank	Cox's model RR	95% CI	Rank
62011	Örnsköldsvik	447	0.36	(0.17–0.78)	1	0.08	(0.01–0.61)	8
53013	Skövde	552	0.45	(0.24–0.84)	2	0.25	(0.09–0.72)	10
64010	Skellefteå	451	0.45	(0.23–0.87)	3	0.22	(0.07–0.73)	9
55012	Lindesberg	429	0.49	(0.25–0.97)	4	0.27	(0.08–0.88)	11
22012	Värnamo	541	0.52	(0.27–0.98)	5	0.32	(0.11–0.91)	12
56012	Köping	470	0.54	(0.27–1.05)	6	0.33	(0.10–1.10)	15
54014	Torsby	435	0.55	(0.28–1.08)	7	0.32	(0.10–1.06)	13
53010	Falköping	321	0.57	(0.29–1.11)	8	0.33	(0.10–1.07)	14
13010	Eskilstuna	425	0.59	(0.32–1.10)	9	0.40	(0.15–1.05)	17
52012	Alingsås	280	0.67	(0.33–1.36)	10	0.39	(0.09–1.62)	16
24010	Västervik	519	0.67	(0.39–1.17)	11	0.55	(0.24–1.27)	21
10484	Sabbatsbergs närsjh.	377	0.68	(0.33–1.40)	12	0.45	(0.11–1.89)	18
52011	Borås	614	0.69	(0.41–1.17)	13	0.58	(0.27–1.26)	23
50010	Östra sjukhuset	658	0.70	(0.43–1.16)	14	0.63	(0.30–1.29)	26
57012	Köping	65	0.71	(0.31–1.64)	15	~0	–	4
54013	Säffle	426	0.73	(0.38–1.38)	16	0.57	(0.20–1.67)	22
56010	Västerås	398	0.74	(0.42–1.28)	17	0.62	(0.27–1.41)	25
13011	Nyköping	308	0.75	(0.41–1.37)	18	0.59	(0.23–1.52)	24
22010	Jönköping	614	0.76	(0.47–1.24)	19	0.70	(0.35–1.40)	29
11002	Huddinge	549	0.79	(0.49–1.26)	20	0.69	(0.35–1.37)	28
13012	Kullbergska sjukhuset	407	0.79	(0.43–1.45)	21	0.68	(0.26–1.77)	27
65014	Kalix	100	0.80	(0.35–1.83)	22	~0	–	7
52013	Skene	423	0.80	(0.46–1.40)	23	0.73	(0.32–1.67)	31
55010	Örebro	427	0.80	(0.47–1.37)	24	0.71	(0.32–1.56)	30
23010	Växjö	349	0.81	(0.46–1.45)	25	0.75	(0.31–1.82)	35
10016	Ortopediska huset	228	0.81	(0.38–1.75)	26	0.52	(0.07–3.82)	19
42010	Halmstad	619	0.81	(0.48–1.38)	27	0.75	(0.35–1.62)	34
25011	Oskarshamn	391	0.82	(0.47–1.43)	28	0.73	(0.32–1.68)	32
65016	Sunderby Hospital	151	0.83	(0.38–1.78)	29	0.53	(0.07–3.93)	20
10010	Sabbatsberg	31	0.83	(0.36–1.91)	30	~0	–	1
51012	Kungälv	476	0.84	(0.49–1.43)	31	0.86	(0.39–1.90)	42
21001	Linköping	649	0.84	(0.55–1.31)	32	0.78	(0.42–1.46)	37
51010	Uddevalla	546	0.85	(0.51–1.40)	33	0.84	(0.41–1.73)	39
41001	Lund	244	0.86	(0.48–1.53)	34	0.74	(0.31–1.79)	33
25010	Kalmar	745	0.86	(0.57–1.31)	35	0.82	(0.45–1.50)	38
28011	Ängelholm	551	0.87	(0.53–1.41)	36	0.86	(0.43–1.73)	41
62010	Sundsvall	628	0.91	(0.59–1.41)	37	0.91	(0.49–1.70)	44
42011	Varberg	785	0.91	(0.61–1.37)	38	0.90	(0.50–1.63)	43
54010	Karlstad	588	0.93	(0.57–1.50)	39	0.84	(0.42–1.71)	40
21480	Linköping med. cent.	9	0.94	(0.41–2.16)	40	~0	–	2
57013	Fagersta/Västerås	8	0.94	(0.41–2.17)	41	~0	–	5
11014	Nacka/Södersjukhuset	166	0.95	(0.51–1.74)	42	0.75	(0.29–2.00)	36
27011	Karlshamn	409	0.95	(0.54–1.65)	43	1.04	(0.45–2.38)	49
50001	Sahlgrenska	343	0.95	(0.56–1.63)	44	0.96	(0.43–2.12)	45
41010	Landskrona	425	0.96	(0.56–1.63)	45	1.07	(0.49–2.35)	53
11011	Södertälje	367	0.97	(0.55–1.68)	46	1.08	(0.47–2.47)	54
63010	Östersund	447	0.97	(0.59–1.61)	47	0.98	(0.48–2.03)	46
50080	Sergelkliniken Gbg	26	0.99	(0.43–2.29)	48	~0	–	3
65011	Luleå	1	1.00	(0.43–2.29)	49	~0	–	6
10013	Södersjukhuset	943	1.01	(0.72–1.42)	50	1.00	reference	47
11012	Norrtälje	351	1.01	(0.59–1.74)	51	1.06	(0.47–2.36)	51
21014	Motala	351	1.01	(0.62–1.67)	52	1.04	(0.50–2.14)	50
65010	Boden	285	1.02	(0.62–1.67)	53	1.01	(0.49–2.07)	48
41012	Helsingborg	447	1.02	(0.63–1.63)	54	1.07	(0.54–2.10)	52
55011	Karlskoga	319	1.05	(0.61–1.79)	55	1.17	(0.53–2.57)	62
11010	Danderyd	938	1.05	(0.71–1.55)	56	1.08	(0.61–1.94)	55
61011	Bollnäs/Söderhamn	285	1.05	(0.62–1.79)	57	1.15	(0.53–2.53)	58

Table 2. Continued

Code	Hospital	No. of prostheses	Frailty model			Cox's model		
			RR	95% CI	Rank	RR	95% CI	Rank
28012	Hässleholm	1037	1.05	(0.73–1.51)	58	1.16	(0.68–2.00)	59
65013	Piteå	183	1.06	(0.56–2.00)	59	1.36	(0.48–3.90)	71
26010	Visby	398	1.06	(0.64–1.76)	60	1.10	(0.52–2.32)	56
10015	Sophiahemmet	312	1.08	(0.62–1.88)	61	1.41	(0.61–3.26)	74
64001	Umeå	313	1.08	(0.66–1.78)	62	1.11	(0.54–2.29)	57
61012	Hudiksvall	397	1.08	(0.67–1.76)	63	1.17	(0.58–2.34)	60
50071	Frölunda spec. sjukhus	86	1.10	(0.51–2.36)	64	3.18	(0.43–23.60)	89
11001	Karolinska	458	1.11	(0.67–1.83)	65	1.29	(0.62–2.66)	66
64011	Lycksele	219	1.11	(0.64–1.93)	66	1.28	(0.56–2.93)	65
27010	Karlskrona	404	1.11	(0.72–1.72)	67	1.17	(0.62–2.19)	61
53011	Lidköping	351	1.12	(0.67–1.88)	68	1.38	(0.65–2.93)	73
28013	Simrishamn	213	1.13	(0.60–2.13)	69	1.35	(0.47–3.85)	70
28010	Kristianstad	53	1.13	(0.58–2.22)	70	1.31	(0.40–4.32)	68
12481	Elisabeth Hospital	12	1.15	(0.53–2.48)	71	5.85	(0.79–43.14)	93
21013	Norrköping	763	1.15	(0.81–1.64)	72	1.18	(0.69–2.01)	63
52016	Vänersborg–NÄL	249	1.19	(0.75–1.91)	73	1.20	(0.61–2.36)	64
65012	Gällivare	338	1.23	(0.79–1.92)	74	1.37	(0.72–2.61)	72
11013	Löwenströmska	390	1.25	(0.81–1.93)	75	1.30	(0.69–2.45)	67
12010	Enköping	325	1.25	(0.72–2.19)	76	1.79	(0.76–4.19)	81
22011	Eksjö–Nässjö	471	1.28	(0.84–1.94)	77	1.49	(0.81–2.72)	75
61014	Söderhamn	53	1.29	(0.68–2.43)	78	2.10	(0.74–6.00)	85
12001	Akademiska sjukhuset	819	1.29	(0.93–1.78)	79	1.33	(0.80–2.21)	69
54011	Kristinehamn	167	1.33	(0.78–2.27)	80	1.79	(0.82–3.94)	82
30001	Malmö	355	1.34	(0.86–2.09)	81	1.50	(0.79–2.85)	76
61013	Sandviken	24	1.34	(0.69–2.63)	82	3.36	(1.02–11.06)	90
57011	Mora	626	1.37	(0.96–1.96)	83	1.56	(0.91–2.67)	77
54012	Arvika	175	1.43	(0.85–2.40)	84	1.95	(0.92–4.13)	84
23011	Ljungby	447	1.49	(1.02–2.19)	85	1.74	(0.99–3.06)	78
56011	Örebro	24	1.49	(0.79–2.82)	86	4.34	(1.52–12.37)	92
51011	Mölndal	348	1.51	(0.99–2.31)	87	1.94	(1.05–3.58)	83
10011	St. Göran	1693	1.56	(1.22–1.99)	88	1.75	(1.12–2.73)	79
41013	Ystad	244	1.59	(1.01–2.51)	89	2.20	(1.14–4.26)	87
57010	Falun	1115	1.61	(1.24–2.09)	90	1.78	(1.13–2.82)	80
41011	Trelleborg	688	1.62	(1.13–2.32)	91	2.11	(1.23–3.64)	86
61010	Gävle	317	1.89	(1.28–2.80)	92	2.60	(1.47–4.63)	88
62013	Sollefteå	238	1.94	(1.23–3.06)	93	3.56	(1.83–6.91)	91

Discussion

Our material spanned a 10-year period. More recent data, e.g. for the past 5 years, would have provided more current information for comparison of hospitals. However, with the low revision risk observed, any reduction in the number of patients analyzed would make the results of analyses more difficult to interpret.

Random-effects models are being increasingly used to study variation in hospital outcome and for profiling providers of medical care (Gatsonis et al. 1995, Goldstein and Spiegelhalter 1996, Christiansen and Morris 1997, Normand et al. 1997). As previous stated, it is known that fixed-effects

models overestimate heterogeneity of effects. The differences between effect estimates from the two models have been clearly shown here. Overestimation of heterogeneity may in itself be misleading, and is an issue that should be addressed when planning and performing investigations on quality improvement.

Results from random-effects models may, however, require different interpretation than those from fixed-effects models; explanatory variables may not be interpretable globally—only at the hospital level. One disadvantage of the random-effects model is also its greater complexity, which makes computation and reporting of analyses more difficult. Then again, these difficulties are diminish-

ing as modern statistical software packages include random effects models and as they are increasingly used in medical research.

From a methodological point of view, random-effects models have several other advantages when investigating geographically aggregated health data (Langford et al. 1998). For instance, fixed-effects models do not produce meaningful revision risk estimates for hospitals with few patients because in such cases random errors tend to have too much influence on the outcome.

The hospitals with the 7 lowest ranks using the fixed-effects model (Table 2) all had results based on very few operations, which of course reduces the credibility of these findings. The random-effects model handles the information better; intuitively, the ranking of the random-effects model appears more adequate.

It has been suggested that the shrinkage of effect estimates in random-effects models can mask important observations from hospitals with high risk and should therefore—as a safety precaution—be avoided. In our opinion, the random-effects model provides more interesting information on safety issues than the fixed-effects model; shrinkage clarifies a picture blurred by random variation. For example, the fixed-effects model could be interpreted as suggesting safety issues in the hospitals having the highest relative risks (Figure). This interpretation could, however, be misleading; other hospitals may be more important to concentrate on. Of the 5 hospitals with the highest (fixed-effect) risk, only 1 is among the 5 with highest risk when the amount of information on which risk estimates are based has been accounted for using the random-effects model (Table 2).

We conclude that the advantages of random-effects models make them more suitable for evaluation of hospital-specific effects in SKAR data. This is not a controversial conclusion. On the contrary, we believe that it is in line with a large number of publications on the methodology of random-effects models, some presenting them under other names such as mixed-effects models, hierarchical models, multi-level models, etc.

Both the fixed- and the random-effects models used in this report included adjustment for differences in sex, age and use of patellar button, but not brand of implant. The reason for not including the brand was mainly that each brand of implant is generally limited to few hospitals and that no brand is used at all participating hospitals. In addition, selection of the correct implant for a patient can be considered a quality issue and part of the outcome.

We found statistically significant variation in revision risk after knee arthroplasty related to the location where the primary operation took place. However, because of the lag time when gathering data and comparing databases, the procedures had been performed 3–13 years prior to the analysis. We wish to stress that this information may not provide rational guidance for patients seeking healthcare. The reasons for this are that risks are generally low, that a considerable part of the variability described may reflect a variation in patient characteristics (so-called case-mix) rather than being the result of surgery, and that the risks presented display historic events, which are not necessarily relevant to current and future patients having the operation.

The main purpose of presenting differences in revision risk between hospitals can mainly be considered to be part of a quality assessment, stimulating participating units to continuously assess their performance and to find ways to improve current practice.

Bauer G, Knutson K, Lindstrand A. (Swedish knee arthroplasties) Läkartidningen 1980; 77: 2088-91.

Christiansen C L, Morris C N. Improving the statistical approach to health care provider profiling. Ann Intern Med 1997; 127: 764-8.

Efron B, Morris C. Stein's paradox in statistics. Sci Am 1977; 236: 119-27.

Eisenhart C. The assumptions underlying the analysis of variance. Biometrics 1947; 3: 1-21.

Gatsonis C A, Epstein A M, Newhouse J P, [alla namn]. Variations in the utilization of coronary angiography for elderly patients with an acute myocardial infarction: an analysis using hierarchical logistic regression. Med Care 1995; 33: 625-42.

Goldstein H, Spiegelhalter D J. League tables and their limitations: statistical issues in comparisons of institutional performance (with discussion). J R Stat Soc A 1996; 159: 285-443.

Langford I H, Bentham G, McDonald A L. Multi-level modelling of geographically aggregated health data: a case study on malignant melanoma mortality and uv exposure in the European community. Stat Med 1998; 17: 41-57.

Normand S-LT, Glickman M E, Gatsonis C A. Statistical methods for profiling providers of medical care: issues and applications. J Am Stat Assoc 1997; 92: 803-14.

Robertsson O, Ranstam J. No bias of ignored bilaterality when analysing the revision risk of knee prostheses: Analysis of a population based sample of 44,590 patients with 55,298 knee prostheses from the national Swedish Knee Arthroplasty Register. BMC Musculoskeletal Disorders 2003; 4: 1-5.

Robertsson O, Dunbar M, Knutson K, Lewold S, Lidgren L. Validation of the Swedish Knee Arthroplasty Register. Acta Orthop Scand 1999; 70: 467-72.

Therneau T M, Grambsch P M. Modelling survival data: extending the Cox model. Springer: New York 2000.