**QUEEN'S
UNIVERSITY
BELFAST**

Predicting survival after treatment for fracture of the proximal femur and the effect of delays to surgery

Elliott, J., Beringer, T., Kee, F., Marsh, D., Willis, C., & Stevenson, M. (2003). Predicting survival after treatment for fracture of the proximal femur and the effect of delays to surgery. *Journal of Clinical Epidemiology*, 56(8), 788-795.

Published in:
Journal of Clinical Epidemiology

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Predicting survival after treatment for fracture of the proximal femur and the effect of delays to surgery

J. Elliott^a, T. Beringer^b, F. Kee^{c,*}, D. Marsh^d, C. Willis^e, M. Stevenson^c

^aDepartment of Trauma and Orthopaedics, Royal Victoria Hospital, Grosvenor Road, Belfast BT 12 6BJ, Ireland

^bThe Elliot Dynes Rehabilitation Unit, Royal Victoria Hospital, Grosvenor Road, Belfast BT 12 6BJ, Ireland

^cDepartment of Epidemiology and Public Health, Queen's University of Belfast, Mulhouse Building, Grosvenor Road, Belfast BT 12 6BJ, Ireland

^dDepartment of Trauma and Orthopaedics, Queens University, Musgrave Park Hospital, Stockmans Lane, Belfast BT 9 7JB, Ireland

^eEastern Health, Social Services Board, Champion House, Linenhall St, Belfast BT 2 8BS, Ireland

Accepted 2 October 2002

Abstract

Background and Objectives: The aim of this study was to evaluate the prognosis after treatment for femoral neck fracture, to assess the impact of delay to surgery, and to devise a clinical prediction rule and score.

Methods: A prospective observational study was conducted in which 1780 patients treated surgically in two teaching hospitals between 1 November 1997 and 31 October 1999 were followed over 12 months. Logistic regression was used to distinguish the effects of predictor variables on survival. Using a probit transformation of the predicted posterior probabilities of death, a prognostic score was devised with scores constrained so that a nominal score of approximately 90 represented a 50:50 chance of survival over 12 months.

Results: Mortality was 30.1% in men and 19.5% in women. Increasing age, male gender, longer pre-operative delay, a higher American Society of Anesthesiology score, a lower Mental Test score, and a lower activities of daily living (Barthel) score were associated with increased risks of death. Of those waiting between 1 and 5 days for surgery, approximately 8 medium-risk and 17 high-risk patients (with prognostic scores of 90 and 120, respectively) would have to have their delay reduced to <24 hours to yield one additional survivor.

Conclusion: The application of prediction rules must be guided by ethical, social, and scientific concerns. © 2003 Elsevier Inc. All rights reserved.

Keywords: Femoral neck fracture; Prognosis; Operative delay

1. Introduction

Although we hope that risk factors such as poor diet and inactivity may be improved by lifestyle changes, current demographic trends guarantee that the health service will have to deal with growing numbers of patients with hip fractures for the foreseeable future. For example, over the last decade, caseloads have risen by between 5% and 10% per year, and it is estimated that by 2011 nearly 70,000 patients will require admission for this condition in England alone [1].

A concerted effort has been made to enhance services for these patients, with guidelines being produced by different bodies to improve their organization, delivery, and efficiency [2–4]. Although the objective of these services is to restore functional independence as much as possible, the consequence of the fracture can be life threatening for a

substantial minority of patients. Although evidence from randomized trials is scarce, if patients whose mortality risk was highest could be identified earlier, it might be possible to marshal resources and target preventive interventions to reduce the risk. Knowing the magnitude of the risk attributable to the main prognostic factors is essential for the design of appropriate intervention trials. The use of risk scores, however, can also help us understand trends in outcomes over time and between treatment units. Consequently, the longer-term outcomes of hip fracture have been studied in a number of prospective studies with a view to identifying the key prognostic variables. Some of these studies have been modest in size, and few have attempted to validate the predictive accuracy of the resulting prognostic score.

In 1997, the fracture surgeons and the ortho-geriatric service in the Royal Victoria Hospital, Belfast, set up a large prospective Audit of Fractures of the Proximal Femur, with the express intention of determining the early and late consequences of hip fracture in their catchment area and of deriving an improved prognostic scoring system. Our objectives were twofold: (1) to determine what were the most important

* Corresponding author. Tel.: +44-2890-894614; fax: +44-2890-231907.

E-mail address: f.kee@qub.ac.uk (F. Kee).

prognostic variables in our population for outcome 12 months after hip fracture repair and (2) to devise a clinical prediction rule. We report here on the results for patients ascertained during the first 2 years of the audit's operation. Given the local health Minister's Priorities for Action (to reduce the maximum delay to treatment to no more than 48 hours), our null hypothesis was that delay was not associated with outcome at 12 months.

2. Methods

The fracture units based at the Belfast City and the Royal Victoria hospital co-operated in the establishment of the Audit before their merger and amalgamation on the Royal Victoria site in November 1999. Together they serve a population of approximately 800,000. A range of socio-demographic and medical information is routinely collected on admission and at 4, 6, and 12 months after injury. The main variables used in this study were age, sex, marital status, a proxy measure of material affluence based on the Townsend deprivation score of the electoral ward [5], an abbreviated Mini-Mental State score on admission [6], pre-injury Barthel (activities of daily living) score [7], and the American Society of Anesthesiology (ASA) Physical Status grading [8] (categories 1 [least impaired physical status] to 4 [most impaired physical status]). In addition, the delay from injury to operation was known. The Townsend Deprivation score is derived for each census ward and has four components: (1) the proportion of the labor force who were unemployed, (2) the proportion of households with no access to a car, (3) the proportion of households not owner occupied, and (4) the proportion of employed men and women in social classes IV and V.

This article is based only on patients who came to surgery and whose date of injury was known. Of the 1876 patients on the database for the years 1998 and 1999, 78 had no surgery, and 18 had injury dates that were clearly estimates (their admission dates were approximately 3 weeks or more from the inferred time of injury). Vital status 1 year after injury was ascertained by regular mailing and telephoning of the patients and their general practitioners. When patients did not reply (by mail or phone), their named caretaker was contacted. When no contact could be established through this route, the practice register was checked. The Central Services Agency maintains the practice registers and, by using the Registrar General's data on death registrations, estimates that 99% of deceased patients' records are removed within 100 days of the death (Dr. David Marshall, CSA, personal communication, 2003).

Univariate contingency tables using chi-squared statistics were calculated to initially assess the effect of possible prognostic factors on vital status (live/dead) 12 months after surgery. Multiple logistic regression was then used to distinguish independent effects on outcome. The odds ratios represent the multiplicative effect of a one-unit change in

the independent variable (operation delay is on a square root scale). A backward procedure was used for variable selection with an initial F test value of $P > .05$ for exclusion. We also sought bi-variate interaction terms for main effects.

The performance of a prognostic model can be assessed in a number of ways. One commonly used method is to derive the c-statistic, which is identical to the area under the receiver operator characteristic curve. Essentially, all possible pairs of individuals where one is dead after 12 months and one is alive after 12 months are considered. Then, the number of such pairs where the posterior probability for the dead person is higher than the posterior probability for the person alive is counted. The latter as a ratio of the former is defined as the c-statistic [9].

We have assessed the performance of our models by randomly splitting the data set into two, deriving independent logistic regression models for each separately, and comparing the predictive performance of each (deriving both a c-statistic and a kappa value) using first their own group's and then the alternate group's data set. Finally, we compared the outcome of these models with a single model using the whole data set.

The logistic regression model provides posterior probabilities of death within 12 months as final output. Although these probabilities are useful to epidemiologists and statisticians, they are not user friendly to clinicians. We have further explored the utility of our model by deriving a simpler prognostic score, based on the magnitude of the coefficients in the finally selected logistic model. To check the validity of the scoring system, the scores were regressed against the probit transformation of the posterior probabilities. The probit transformation also normalizes the posterior probabilities (Figs. 1, 2), which is advantageous when there is some skew in the data. For instance, a model that is asked to discriminate between a large number of cases in the middle of the risk distribution has a much harder task than one asked to discriminate between a distribution of cases in which posterior probabilities close to 0 or 1 occur frequently.

Because probits are standard normal variates (z values), a value of 0 equates to a cumulative probability of 0.5.

3. Results

A total of 415 men and 1365 women sustained hip fracture that was surgically treated in the Belfast City and Royal Victoria hospitals between 1 November 1997 and 31 October 1999. Twelve-month follow-up on all patients showed 30.1% mortality in men and 19.5% in women. The characteristics of these patients and their effects on 12-month mortality are shown in Table 1. The most obvious effects are the higher mortality with increasing age, male gender, ASA status, and lower mini-mental state and Barthel scores. Longer operation delays were also associated with poorer outcome, but there was a better outcome among single and divorced subjects.

Number of patients

Predicted probability using all data

Fig. 1. Predicted probability of death at 12 months (using all data)

In the modeling process, we found a linear relationship with the square root of delay. Although the effect is illustrated in Table 1 using five categories (ranging from <1 day to >10 days), our most parsimonious model was based on

a single parameter SQRD-delay as a continuous variable. Although post-coding and Townsend scoring was unsuccessful in 7% of cases, there was no apparent effect of deprivation on outcome.

Number of patients

Fig. 2. Probit transformation of predicted probabilities.

Is your patient at risk?

Following fractured neck of femur some patients have much reduced life span. The following ready reckoner is designed as a guide to indicate those patients who would most benefit from prioritisation for surgery.

Is the patient male? If so award 15 points

For ASA stage 2 award 20 points; at stage 3 award 25 points and at stage 4 award 35 points

For every completed two years of life award a point.

For every point lost on the Mental Test Score and Barthel scales, award a point.

Is the patient single or divorced? If so deduct 8 points.

How did your patient score? If the total is below 50, the patient will almost certainly survive and a small delay in time to surgery will be of small consequence.

If the total is between 50 and 120 your patient is in the group that would most benefit from priority – particularly so if the score falls between 70 and 100.

If the score is over 120 the odds are now stacked against survival and the effect of speedy surgery, while obviously still the goal, is nevertheless diminishing.

Fig. 3. Illustrative derivation of “at-risk” score.

Tables 2, 3, and 4 give the results of logistic regression for the data set overall and when divided randomly into two groups (group 1 and group 2). Because there were no deaths in ASA category 1 in group 2, the comparison category is ASA group 4. Largely, the same variables were selected by each data set. In group 1, the effect of marital status was not significant, and the confidence intervals for the Barthel score effect were 0.90 to 1.00. In group 2, the main (and interaction) effect of delay was not significant.

Table 5 summarizes the results of our data-splitting validation procedure. All of the c-statistics were in the range 0.73 to 0.80. Group 2 had a slightly wider range of explanatory variables than group 1, and predictions of the status of its members tended to be a little better. There was little deterioration in predicting individual status from an independently derived set of data when compared with predicting status directly from the data used to derive the model.

The distribution of posterior probabilities (Fig. 1) is considerably skewed to the right. The typical person admitted for neck of femur fracture had a small probability of not surviving 12 months, and hence the mode of the distribution was approximately 0.1. A minority of patients had a large likelihood of nonsurvival, with posterior probabilities in excess of 0.8. The distribution can be normalized by

applying a probit transformation as shown in Fig. 2. Because probits are standard normal variates (z values), a value of 0 equates to a cumulative probability of 0.5, -1 equates to a cumulative probability of 0.16, and -2 equates to a cumulative probability of approximately 0.025. A value of $+1$ equates to a probability of 0.84.

Because it is rarely desirable for a clinician to feed data into a logistic regression “black box” to produce a posterior probability, we have devised a moderately simple scoring system that has a strong relationship with the probit scale described above. We have constrained these scores in such a way so that a nominal score of approximately 90 represents roughly a 50:50 chance of survival over 12 months.

The basis of the score is illustrated in Panel 1. There is a virtually linear relationship between the derived scores and the probits. The calibration of our scoring system is illustrated in Appendix 1. Even though we have excluded terms that are delay dependent (we considered that practitioners would form a view about the patient prognosis on admission, before any elapsed delay), the calibration is still excellent. If investments are made to shorten delay time to surgery, for example, this scoring system can illustrate the implications. It enables us to make calculations of how the probability of death within 12 months changes as a result

Table 1
Univariate relationship of key variables with vital status at follow-up

	Alive after 1 yr		Dead after 1 yr		Chi-squared (df)	P value
	n	%	n	%		
Sex						
Males	290	69.9	125	30.1	$\chi^2_1 = 20.4$	<.001
Females	1099	80.5	266	19.5		
Missing = 0						
Age group						
Under 65 yr	198	90.8	20	9.2	$\chi^2_3 = 67.3$	<.001
65–74 yr	268	85.9	44	14.1		
75–84 yr	552	78.7	149	21.3		
Over 85 yr	371	67.6	178	32.4		
Missing = 0						
Marital status						
Married	286	79.9	72	20.1	$\chi^2_4 = 16.3$	.003
Widowed	552	74.3	191	25.7		
Single	246	84.2	46	15.8		
Divorced	29	90.6	3	9.4		
Not stated	276	77.7	79	22.3		
Missing = 0						
Townsend						
Most deprived	278	80.3	68	19.7	$\chi^2_3 = 5.56$	.135
Deprived	259	75.7	83	24.3		
Affluent	410	76.6	125	23.4		
Most affluent	357	81.3	82	18.7		
Missing = 120						
Barthel score						
0–7	83	55.0	68	45.0	$\chi^2_4 = 164.3$	<.001
8–11	152	59.6	103	40.4		
12–16	331	74.5	113	25.5		
17–19	377	85.5	64	14.5		
20	446	91.2	43	8.8		
Missing = 0						
Mental test score						
0–2	227	60.5	148	39.5	$\chi^2_3 = 141.1$	<.001
3–6	203	67.7	97	32.3		
7–9	307	82.5	65	17.5		
10	652	88.9	81	11.1		
Missing = 0						
ASA categories						
1	143	97.9	3	2.1	$\chi^2_3 = 102.4$	<.001
2	408	86.1	66	13.9		
3	756	74.7	256	25.3		
4	82	55.4	66	44.6		
Missing = 0						
Delay categories						
<1 d	155	91.7	14	8.3	$\chi^2_4 = 53.8$	<.001
1 <3 d	540	79.5	139	20.5		
3 <5 d	357	81.5	81	18.5		
5 <10 d	254	70.4	107	29.6		
>10 d	83	62.4	50	37.6		
Missing = 0						
Total	1389	78.0	391	21.9		

of delay and hence to calculate the number of additional patients who need to have their surgery within 24 hours to yield one additional survivor at 1 year (a sort of “number needed to treat”). Table 6 illustrates the results. The dividend is greatest for patients with scores in the middle range.

Table 2
Logistic regression model^a for outcome at 12 months in data set 1

	Odds ratio	Confidence intervals	Wald	P value
Sex (female versus male)	0.44	0.29–0.66	15.45	<.001
ASA (1) versus (4)	0.15	0.04–0.54	8.26	.004
ASA (2) versus (4)	0.34	0.18–0.64	11.22	<.001
ASA (3) versus (4)	0.35	0.20–0.61	13.80	<.001
Mental score (0–10)	0.94	0.88–1.00	4.14	.042
Barthel score (0–20)	0.95	0.90–1.00	3.39	.065
Age (per yr)	1.09	1.04–1.15	10.62	.001
Operation delay ^b	11.5	1.80–73.6	6.65	.010
Interaction (delay × age)	0.97	0.95–1.00	5.02	.025
Marital status ^c	0.71	0.44–1.14	2.00	.157

Abbreviation: ASA, American Society of Anesthesiology.

^a Adjusted for all co-variables.

^b Delay in days, with square root transformation.

^c (Single or divorced) versus (all other categories).

4. Discussion

Our main objectives were to assess the determinants of survival 12 months after repair of hip fracture to assess the independent effects of delay to surgery and to devise a clinical prediction score. Although a number of studies from within and outside the UK have examined mortality after treatment for hip fracture, it is useful to put our own results in context and to describe how they affirm or extend previous findings.

First, although we share the rigor and discipline of prospective data collection to which studies like the East Anglian audit have subscribed [10], our sample size was substantially larger, and our follow-up was longer than in many others in the field [11–15]. Studies with different durations of follow-up are not strictly comparable because the prognostic factors for early versus late mortality may be inherently different. Studies may also differ according to the proportion of all patients presenting who end up having surgery. In our study, the proportion not receiving surgery was 4.2%, and we confined our prognostic model to the

Table 3
Final logistic regression model^a of outcome at 12 months for Data Set 2

	Odds ratio	Confidence intervals	Wald	P value
Sex (female versus male)	0.34	0.22–0.51	25.69	<.001
ASA (1) versus (4)	0.00	Indeterminate		
ASA (2) versus (4)	0.45	0.22–0.90	5.17	.023
ASA (3) versus (4)	0.66	0.37–1.16	2.08	.149
Mental score (0–10)	0.90	0.84–0.96	10.96	.001
Barthel score (0–20)	0.93	0.88–0.98	7.49	.006
Age (per yr)	1.06	1.01–1.12	5.22	.022
Operation delay ^b	5.08	0.76–34.0	2.80	.094
Interaction (delay × age)	0.98	0.96–1.01	2.07	.150
Marital status ^c	0.43	0.25–0.74	9.38	.002

Abbreviation: ASA, American Society of Anesthesiology.

^a Fully adjusted model.

^b Delay in days with square root transformation.

^c (Single or divorced) versus (all other categories).

Table 4
Logistic regression model^a for outcome at 12 months
for combined data set

	Odds ratio	Confidence intervals	Wald	P value
Sex (female versus male)	0.39	0.29–0.52	39.96	<.001
ASA (1) versus (4)	0.10	0.03–0.35	13.06	<.001
ASA (2) versus (4)	0.39	0.25–0.63	15.42	<.001
ASA (3) versus (4)	0.47	0.32–0.70	14.01	<.001
Mental score (0–10)	0.92	0.88–0.96	14.22	<.001
Barthel score (0–20)	0.94	0.91–0.98	10.20	.001
Age (per yr)	1.08	1.04–1.12	16.47	<.001
Operation delay ^b	7.98	2.17–29.3	9.79	.002
Interaction (delay × age)	0.98	0.96–0.99	7.24	.007
Marital status ^c	0.56	0.39–0.80	10.23	.001

Abbreviation: ASA, American Society of Anesthesiology.

^a Fully adjusted model.

^b Delay in days, with square root transformation.

^c (Single or divorced) versus (all other categories).

remainder. In the studies cited above, the proportion of their cohorts not receiving surgery varied from 0% to 13.9%. Patients not operated upon represent a high-risk group (in our study, the mortality exceeded 60% at 12 months).

Second, we have gone beyond subjective assessments by using, where possible, validated measures of pre-operative function and by excluding patients for whom surgery was not offered or with unknown dates of injury. Fifteen years ago, orthopedic surgeons from our region derived a simple grading system to guide prognosis after hip fracture, but that was based on the surgeon's pre-operative subjective judgment of the importance of factors such as the patient's medical state and social circumstances [16].

Third, we have attempted some validation of our prognostic model and its translation into a more clinically friendly "score."

The poorer outcome among men has been reported previously [10,12], as has the worse outcome with declining mental function [17,18]. Similarly, the higher mortality among those who, pre-injury, were less able to perform activities of daily living (measured here using the Barthel score) was found in the East Anglian audit [10]. The better survival among the single and divorced (compared with all other categories) runs counter to other studies and will be further explored in future data from this ongoing cohort study. Had this observation been attributable to bias as a

consequence of how patients were selected for surgery, we would have expected to see relatively more high-mortality single and divorced subjects among those excluded from surgery. This was not the case (data not shown). Residual confounding remains a possible explanation. When comparing disparate findings in the literature for the effects of any specific variable, it is important to bear in mind how the choice of the statistical method affects the magnitude of the observed effect. Employing a Cox's proportional hazards function, the method of choice when there is variable follow-up would permit the estimation of (instantaneous) relative risks, and these would tend to be somewhat lower than the equivalent relative odds for a reasonably common outcome (as in our study) that occurs in 20% of the cohort. Logistic regression has frequently been the approach in previous studies [10,15,19,20], probably because the chances of survival or death at 12 months (or a given period) are more readily interpretable than instantaneous hazards.

The real impact of operative delay is rather harder to determine, and findings in previous studies have been mixed [14,21]. Although we attempted multivariate adjustment in the logistic regression analysis, longer delays were more common among patients with higher ASA scores (data not shown). The real impact of delay is confounded if the sicker patients, with an inherently higher risk of mortality, take longer to prepare for the shorter-term risks of surgery. If this were the entire explanation for our findings, then what we observed as the effect of delay might have been different (i.e., greater) for those dying early in the follow-up period (the sickest of the sick) compared with those dying later. However, the odds ratio for the effect of delay increased when our analysis was restricted to deaths after 30 days (data not shown), and the average delay to surgical intervention for those dying within 30 days was 4.8 days versus 5.7 days for those dying later. These observations are less compatible with a "confounding by intention" explanation for the delay effect.

Our analysis showed evidence of a significant interaction between the effects of delay and age. The degree of moderation of the multiplicative effects of age and delay was 0.98. However, we have excluded terms that are delay dependent from our prognostic score because we considered that practitioners would form a view about the patient's prognosis on admission, before any elapsed delay. The calibration of the model is still excellent. By the same token, the NNTs in Table 6 are based on the main effect of delay without interaction, and these estimates are therefore more reliable in the middle of the age distribution. Because there are seven main effects in the model, there are 21 possible bi-variate interaction terms. Thus, given the chance of a type I error, observing a significant interaction should be viewed as *hypothesis generating* rather than *hypothesis testing*.

Given the magnitude of the possible effect of delay, however, any attempt by a fracture unit to improve efficiency and reduce delays needs to be targeted to those most likely to benefit. The derivation of our prognostic score places the

Table 5
Validation and calibration of prognostic models

Data used for prediction	Data predicted	c statistic
All	All	0.769
Set 1	Set 1	0.747
Set 1	Set 2	0.791
Set 2	Set 1	0.728
Set 2	Set 2	0.800
Both sets	Same sets	0.773
Both sets	Opposite sets	0.759

Table 6

The number of patients whose delay needs to be shortened to <24 hr to yield an additional survivor (NNT)

Score	Probability of survival if treated within 24 hr	With delay 1 to 5 d (NNT)	With delay > 5 days (NNT)
50	0.9634	0.9276 (28)	0.8783 (12)
60	0.9154	0.8512 (16)	0.7734 (7)
70	0.8311	0.7340 (10)	0.6306 (5)
80	0.7060	0.5825 (8)	0.4668 (4)
90	0.5497	0.4175 (8)	0.3085 (4)
100	0.3853	0.2660 (8)	0.1797 (5)
110	0.2394	0.1488 (11)	0.0912 (7)
120	0.1303	0.0724 (17)	0.0401 (11)

delay effect, if real, into context. Using the probit transformation, for example, it is straightforward to calculate, at different positions on the score's distribution, the "number-needed-to-treat," or the number of patients whose delay would need to be curtailed to yield an additional survivor. From this perspective, when pressure on theater time and beds is at its highest, the aggregate risk of death for patients awaiting surgery can be lowered more efficiently by attempting to curtail the delay more for the patients in the middle of the score distribution, rather than for those at the "right hand" tail of the distribution (whose risk is high almost irrespective of the delay). The acid test would be to compare different strategies in a trial in which the prognostic score might be incorporated into some sort of prioritization system.

The value of any prognostic scoring system depends on three main issues: (1) its statistical validity and calibration, (2) whether it has an effect on clinical decision-making, and (3) its ethical validity. The statistical validity depends on, among other things, how robust the procedures are for variable selection. This can depend on the shape and distribution of independent variables [22]. Notwithstanding the procedure used, all models should have their calibration or discriminative ability tested, to guard against overfitting, by some sort of cross-validation procedure [23,24]. We have attempted this by data splitting. The derived prognostic model was reasonably robust. However, the best way to test a prediction rule is to measure the misclassification rate in a new clinical setting or in a new cohort of patients.

Determining the clinical utility of a prognostic model is often more difficult. The ability to be more explicit about prognosis can, if expressed in the right way, foster better communication with and realism among patients' families. In our experience, many of them believe their relative's injury to be non-life threatening and would be surprised to learn that the outcome at 1 year was little better than for many cancers in the same age group.

Although some clinicians may appreciate that the association between the prognostic strength and statistical significance of a relationship between a variable and outcome may be weak, fewer appreciate the distinction between a

prognostic factor and a predictive factor [25]. Observational cohorts like ours (and all of the others we have cited) cannot distinguish a purely prognostic factor from a predictive factor, the latter telling us something more about how a variable affects the response to treatment. The proper way to distinguish the two is in a clinical trial that has a randomly selected untreated cohort.

The clinical application of prediction rules needs to be guided by ethical, social, and scientific concerns. We can conceive of two sorts of trials that help clarify the utility of our prognostic model. One might try to affect the timing of surgery for some patients (if the total volume of daily procedures was limited) by disclosing scores to one group of clinicians or during one period, but then an ethical judgment must be made about the "right" distribution of benefits (if risk of mortality for some was reduced). We do not favor the application of this prognostic score in its current form, unless explicit account was taken of the elapsed time waiting for surgery. Otherwise, we could foresee a scenario in which patients with a poor outlook might continually sink to the end of the queue. That would be unacceptable and in our view ethically indefensible.

In another hypothetical trial, scores might be disclosed (or not) to rehabilitation and community teams to determine whether targeted rehabilitation interventions were facilitated. In the latter instance, however, there would be a necessary debate about the value of the length of life versus the patient's ability to function independently and perceptions of its quality. In that case, it is likely that a different prediction tool would be required. We are currently examining our data set to discover which factors can predict restoration of function.

Acknowledgments

We thank the patients who took part and their respective clinicians. We are grateful to Angela Dolan, Martin McAnespie, and Patricia Durkin for supervising the follow-up procedures and data entry. This study was supported by the Eastern, Northern and Southern Health and Social Services Boards; the Regional Multi-Disciplinary Audit Committee; Johnson and Johnson; Howmedica UK Ltd; and the BCH and RGH Trusts.

Appendix 1 Calibration of the scoring method

This is illustrated by producing an unconstrained goodness-of-fit test for all patients whose scores rounded to the nearest 10 were in the range 50 to 120. The probabilities from the probit transformed model are applied, and expectations for each of the eight groups formed can be calculated. An approximate chi-squared test is then constructed to test the goodness of fit.

Score (Group)	Dead	Not dead	Expected dead	Expected not dead	Chi square(d)	Chi square (not d)
50	12	157	8.08	160.92	1.902	0.095
60	22	219	21.98	219.02	0.000	0.000
70	49	302	54.86	296.14	0.626	0.116
80	90	270	90.90	269.10	0.009	0.003
90	96	190	105.65	180.35	0.881	0.516
100	70	75	72.50	72.50	0.086	0.086
110	42	21	39.63	23.37	0.142	0.240
120	9	7	11.96	4.04	0.733	2.169
Total	390	1241	405.56	1225.44	4.379	3.226

Total 7.605 on 8 degrees of freedom because totals were not constrained
P value .473

This illustrates that the scoring system appears to be well calibrated
for this group of patients.

References

- [1] Hollingworth W, Todd C, Parker M. The cost of treating hip fractures in the 21st century. *J Pub Health Med* 1995;17:269–76.
- [2] Fractured neck of femur. Prevention and management. A Report of the Royal College of Physicians. London: The Royal College of Physicians of London; 1989.
- [3] Management of elderly people with fractured hip. A national clinical guideline for use in Scotland. Edinburgh (Scotland): Scottish Inter-collegiate Guidelines Network; 1997.
- [4] Fractured proximal femur: health outcome indicators. Report of a working party to the Department of Health. Oxford (England): Oxford National Centre for Health Outcomes Development; 1999.
- [5] Townsend P, Phillimore P. Health and deprivation. London: Routledge; 1999.
- [6] Qurshi KN, Hodkinson HM. Evaluation of a ten question mental test in the institutionalised elderly. *Age Ageing* 1974;3:152–7.
- [7] Mahoney FI, Barthel OW. Functional evaluation: the Barthel Score. *Maryland State Med J* 1965;14:61–5.
- [8] Owens WD, Felts JA, Spitznagel EL. ASA Physical Status Classifications: a study of consistency ratings. *Anesthesiology* 1978;49:239–43.
- [9] Harrell FE, Califf RM, Pryor DB, et al. Evaluating the yield of medical tests. *JAMA* 1982;247:2543–6.
- [10] Todd CJ, Freeman CJ, Camilleri-Ferrante C, et al. Differences in mortality after fracture of the hip: the East Anglian audit. *BMJ* 1995;310:90–8.
- [11] Lyons AR. Clinical outcomes and treatment of hip fracture. *Am J Med* 1997;103:51S–64S.
- [12] Diamond TH, Thornley SW, Sekal R, et al. Hip fracture in elderly men: prognostic factors and outcomes. *Med J Aust* 1997;167:412–5.
- [13] Baudoin C, Fardellone P, Bean K, et al. Clinical outcomes and mortality after hip fracture: a two year follow-up study. *Bone* 1996;18:149S–57S.
- [14] Zuckerman J, Skovron ML, Koval KJ, et al. Post operative complications and mortality associated with operative delay in older patients who have a fracture of the hip. *J Bone Joint Surg* 1995;77:1551–6.
- [15] Hannan E, Magaziner J, Wang JJ, et al. Mortality and locomotion 6 months after hospitalisation for hip fracture: risk factors and risk adjusted hospital outcomes. *JAMA* 2001;285:2736–42.
- [16] Wallace RGH, Lowry JH, McLeod NW, et al. A simple grading system to guide the prognosis after hip fracture in the elderly. *BMJ* 1986;293:665.
- [17] Wood DJ, Ions GK, Quinby JM, et al. Factors which influence mortality after hip subcapital fracture. *J Bone Joint Surg Br* 1996;74B:199–202.
- [18] Ions GK, Stevens J. Prediction of survival in patients with femoral neck fracture. *J Bone Joint Surg* 1987;69B:384–7.
- [19] Clague JE, Craddock E, Andrew G, et al. Predictor of outcome following hip fracture: admission time predicts length of stay and in hospital mortality. *Injury* 2002; 1–6.
- [20] Kramer A, Steiner JF, Schlenker RE, et al. Outcomes and costs after hip fracture and stroke. *JAMA* 1997;277:396–404.
- [21] Hoenig H, Rubenstein L, Sloane R, et al. What is the role of timing in the surgical and rehabilitative care of community dwelling older persons with hip fracture? *Arch Intern Med* 1997;157:513–20.
- [22] Walter S, Feinstein AR, Wells CK. A comparison of multi-variate methods for predicting survival. II. Statistical selection of prognostic variables. *J Clin Epidemiol* 1990;43:349–59.
- [23] Concato J, Feinstein AR, Holford TR. The risk of determining risk with multi-variable models. *Ann Intern Med* 1993;118:201–10.
- [24] Wasson J, Sox H, Neff R, Goldman L. Clinical prediction rules: applications and standards. *N Engl J Med* 1985;313:793–9.
- [25] Hayes DF, Trock B, Harris AL. Assessing the clinical impact of prognostic factors: when is statistically significant clinically useful? *Breast Cancer Res Treat* 1998;52:305–19.