Description of an individual patient methodology for calculating the cost-effectiveness of treatments for osteoporosis in women

MD Stevenson[1]*, JE Brazier[1], NW Calvert[1], M Lloyd-Jones[1], JE Oakley[2] and JA Kanis[3]

[1]School of Health and Related Research, University of Sheffield, Sheffield, UK; [2]Department of Probability and Statistics, University of Sheffield, Sheffield, UK; and [3]WHO Collaborating Centre for Metabolic Bone Diseases, University of Sheffield Medical School, Sheffield, UK

Models of the cost-effectiveness of pharmaceutical interventions for the treatment of osteoporosis have traditionally adopted cohort-based approaches. We present a transition-state model to simulate the experience of individual patients, allowing the full patient history and residential status to influence the probabilities of future fractures at the hip, spine, wrist or proximal humerus. Alongside epidemiological data, we used systematic literature reviews of costs, utilities and efficacy to populate the model for a UK setting. We established a statistical relationship between the inputs and outputs of the individual patient model creating a near instantaneous emulation of the individual patient model. We undertook extensive sensitivity analyses to analyse the uncertainty in the estimated incremental cost per quality-adjusted life year due to uncertainty in the efficacy of the drugs. We provide illustrative results accompanied by individual and multi-interventional cost-effectiveness acceptability curves.

Introduction

Osteoporosis is a systemic skeletal disease characterized by low bone mass and micro-architectural deterioration of bone tissue with a subsequent increase in bone fragility and susceptibility to fracture.[1] It is operationally defined by bone mineral density at the hip measured by dual energy X-ray absorptiometry and diagnosed in women by the finding of a t-score of -2.5 standard deviations or lower.[2] A t-score is defined as the number of standard deviations from the average bone mineral density of young healthy women. Women with osteoporosis and a previous fragility fracture are denoted as suffering from established osteoporosis. The most serious clinical consequences of osteoporosis are hip fractures that increase in incidence exponentially with age and incur high morbidity, mortality and health-care expenditure. Other common osteoporotic fractures occur at the spine, forearm, and shoulder.

In the UK, the recent establishment of the National Institute for Clinical Excellence (NICE) reflects the Government's commitment to investigating cost-effectiveness alongside clinical effectiveness for drugs and other interventions in order to make efficient use of scarce health-care resources.[3]

*Correspondence: MD Stevenson, Operational Research, School of Health and Related Research, University of Sheffield, 30 Regent St, Sheffield S1 4DA, UK.
E-mail: m.d.stevenson@shef.ac.uk

Many agents of widely different efficacy and cost are now available for the treatment of osteoporosis with an estimated £1030 million spent annually in the UK in treating osteoporotic fractures in women aged 50 years or over.[4] Accordingly NICE commissioned us to review treatments for osteoporosis in post-menopausal women and place them in a health economic perspective. We calculated results for women aged 50, 60, 70 and 80 years, with and without an assumed prior fracture. Our results formed the basis of the NICE final appraisal document.[5]

Methods

Standard measures of cost-effectiveness in health care modelling

An influential approach to the assessment of cost-effectiveness has been to compare interventions in terms of their incremental cost per 'quality-adjusted life year' (QALYs).[6] The QALY combines increased life expectancy and improvements in health status by assigning to each period of time a utility ranging from 0 to 1, corresponding to the health-related quality during that period, where a utility of 1 corresponds to optimal health, and a weight of 0 corresponds to a health state judged to be equivalent to death.[7]

The QALY approach thus 'quality adjusts' survival. A person expected to survive 10 years at a quality of 0.8 has

eight QALYs. The benefits of a treatment that increases survival at a utility of 0.8 (from 10 to 20 years) or improves the quality of the 10 years (from 0.8 to 0.9) can be valued in terms of the QALY gain (ie gains of eight and one, respectively).

Theoretically, the cost per QALY of all treatments in all disease areas for all potential patient groups would be calculated, and the optimal allocation of resources selected in order to maximize societal QALYs given a fixed budget (ie a constrained optimization problem).[8] However, there is insufficient data to do this and decision makers instead use an arbitrary cost per QALY threshold that represents value for money. It has been hypothesized that historically NICE has set this value at around £30 000 per QALY gained.[9]

Cost per QALY output from sensitivity analyses can be expressed as cost-effectiveness acceptability curves (CEAC).[10] The simplest form of a cost-effectiveness acceptability curve compares one intervention to another (typically a no treatment option), where a line denotes the proportion of times within the simulations that an intervention is estimated to have a cost per QALY ratio better than a given threshold value. This allows decision makers to visualize the likely range in the cost per QALY estimates and predict the likelihood of achieving a selection of cost per QALY thresholds. Where the line crosses the *y*-axis, the intervention dominates the comparator, as it produces a cost saving and increases societal health. The proportion of times that the intervention is dominated by the comparator, where less health is achieved at a cost, is signified by the difference between 1 and the value of the line when the cost per QALY threshold is equal to infinity.

More complex cost-effectiveness acceptability curves incorporate a number of interventions simultaneously and display the probability that each treatment is the optimal intervention at a given cost per QALY threshold.[11] Optimal interventions are calculated by ranking them in order of ascending health gain and initially comparing the two least effective treatments. If the incremental cost per QALY between the more effective treatment and the lesser one is below the cost per QALY threshold, the more effective treatment is selected as optimal. Similar comparisons are then iteratively conducted between the current optimal treatment and the next most efficacious treatment, until the list is exhausted, and the optimal treatment found.

The structure of the model

We constructed a patient-based transition-state model, with time slices of 1 year, in Microsoft Excel ([©]Microsoft Corporation). Four fracture sites were incorporated: hip, wrist, spine and proximal humerus and two extra-skeletal conditions, breast cancer and coronary heart disease, due to the effects of some osteoporosis interventions on the incidences of these diseases. These transition states will be collectively referred to as events. We also added a further transition state denoting no event occurring in the year. Absorption states, which cannot be exited, were included for death through natural causes and for death following hip fracture, breast cancer or coronary heart disease.

The modelling methodology is similar to that described by Eastman *et al*[12] with individual patients being simulated. The model is updated annually, with the probability of each event occurring in the next time period calculated at the start of the year, based on epidemiological data, patient fracture history and residential status. The risks of events were further adjusted for the assumed efficacy of any treatment for osteoporosis. For example, where a treatment was shown to reduce the rate of hip fracture by 50%, the risks of suffering a hip fracture would be halved during the period that treatment was taken.

Having calculated transition probabilities for each event, these events were cumulatively proportioned across a 0–1 interval. We then drew a random number using the standard random number generator in Excel to determine which event, if any, would occur in the forthcoming year. The simulation of events occurring in the next annum is repeated until a time horizon of 10 years was reached or the patient had died. Due to problems in Excel, the random numbers were unseeded in all simulations.

The model structure is depicted in Figure 1. The outputs were costs due to events, QALYs accrued, number of years of pharmaceutical treatment, number of GP consultations and number of bone mineral density scans performed. Based on UK guidance at the time, we discounted costs and QALYs at 6 and 1.5%, respectively.[13]

An event-specific cost and a utility multiplier were incurred on entering each transition state. However, costs and utility detriments persist beyond the 1-year time period. We used Boolean variables within the model to monitor the presence of prior events and residential status allowing ongoing costs and quality of life decrements related to previous events to be considered regardless of the current transition state. We assumed utility was multiplicative; thus a woman who had suffered two separate events with a multiplier of 0.8 would be expected to be at 0.64 (the average utility for that age). This approach does not allow utility to fall below zero and assumes that a fracture would produce more disutility in women who were healthier at the time of the event. Costs were assumed to be additive.

The simulated costs incurred and QALYs gained by the patient in the current time period were calculated and an aggregated total produced at the end of the modelling horizon. Although the model simulates individual patients, decision makers are more interested in the results for a cohort of patients with similar characteristics. Aggregating data relating to a large number of individual patients produces such results.

We used a time horizon of 10 years for the model as health-care interventions may alter significantly in the future. However, the QALYs accrued beyond the 10-year period
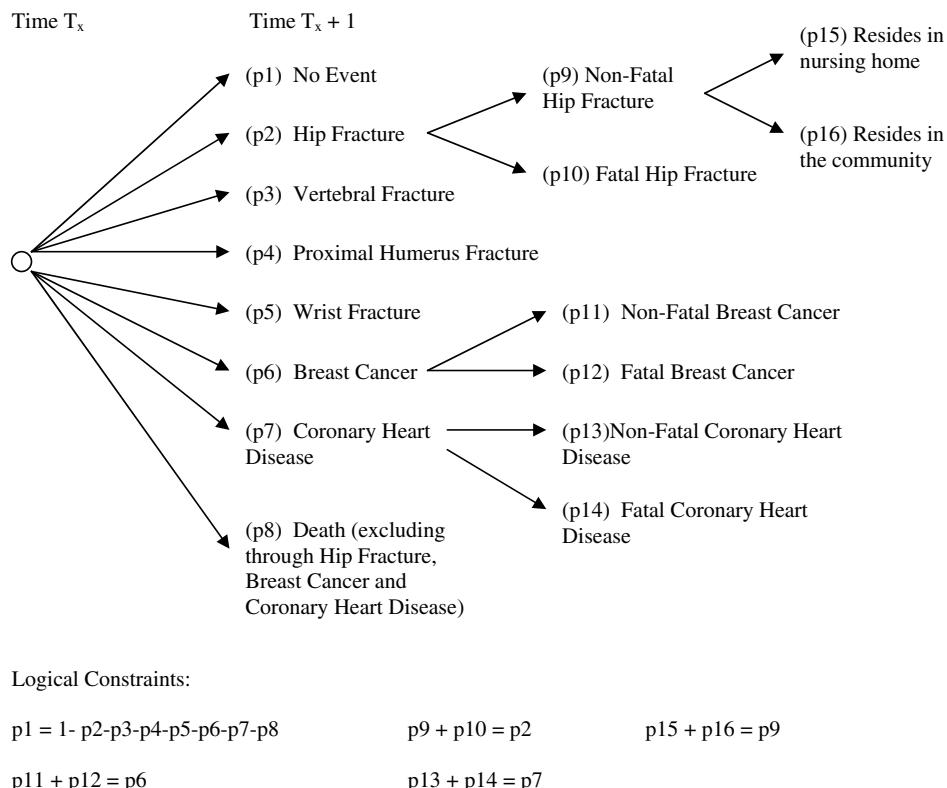
Time T_x Time $T_x + 1$

○
- (p1) No Event
- (p2) Hip Fracture
  - (p9) Non-Fatal Hip Fracture
    - (p15) Resides in nursing home
    - (p16) Resides in the community
  - (p10) Fatal Hip Fracture
- (p3) Vertebral Fracture
- (p4) Proximal Humerus Fracture
- (p5) Wrist Fracture
- (p6) Breast Cancer
  - (p11) Non-Fatal Breast Cancer
  - (p12) Fatal Breast Cancer
- (p7) Coronary Heart Disease
  - (p13) Non-Fatal Coronary Heart Disease
  - (p14) Fatal Coronary Heart Disease
- (p8) Death (excluding through Hip Fracture, Breast Cancer and Coronary Heart Disease)

Logical Constraints:

$p1 = 1 - p2 - p3 - p4 - p5 - p6 - p7 - p8$ $p9 + p10 = p2$ $p15 + p16 = p9$

$p11 + p12 = p6$ $p13 + p14 = p7$

Figure 1 The transition states included in the model. The exact values of p2:p14 will be determined by the patient age, patient history regarding the presence of previous fracture at each site, and the residential status of the patient. These probabilities are calculated for each individual at the beginning of each year. The cycle is repeated for all non-absorbing states until the time horizon is reached.

must be considered for the additional survivors associated with treatment. We assumed that patients alive at the end of the modelling horizon would have a mortality rate equal to that of an age- and sex-matched population.[14] This resulted in an additional 10.5, 5.8, 2.3 and 0.3 QALYs for each additional patient alive that entered the model at 50, 60, 70 and 80 years of age, respectively. We added the expected QALYs gained due to lower mortality rates to the QALYs produced by the individual patient model to produce the total QALY gain.

Data used within the model

A detailed account of the epidemiological, cost and utility data used within the model is provided elsewhere;[15] however, we provide a summary of cost and QALY data in Tables 1 and 2. All prices have been inflated to 2001/2002 prices.[16,17] We undertook a systematic review of all randomized controlled trials (RCTs) of drugs that used fracture data to measure the clinical efficacy of interventions in patients with osteoporosis. Full details of the methodology used, including the search terms, quality scores, and databases interrogated, are provided elsewhere.[15] There is a paucity of RCTs of oestrogen in osteoporotic women as it was prescribed before

the need to prove efficacy in such a manner. We therefore allowed fracture evidence for oestrogen from women without osteoporosis to be considered in a sensitivity analysis.

We undertook meta-analyses and the relative risks associated with a selection of interventions are given in Table 3. These risks represent the likelihood of a fracture relative to receiving no treatment, that is, a relative risk of 0.8 would mean that a patient on treatment would be 20% less likely to suffer a fracture. We assumed an offset time, defined as the length of time following cessation of treatment until the relative risk returns to 1, of 5 years for all treatments analysed. We assumed a linear decay in efficacy during this period.

First- and second-order uncertainty

A stochastic individual patient approach introduces first-order uncertainty, that is, variation in output from the same input parameters solely due to the random numbers sampled. Second-order uncertainty is the uncertainty around the true parameter values, for example, the efficacy of alendronate in preventing hip fractures. First-order uncertainty can be virtually eliminated by simulating an arbitrarily large number of patients.[10] For patients with a prior

Table 1 Utility data used in the model

	Utility multiplier		Source data
State	1st Year	Subsequent years	State
Hip fracture	0.830	0.925	Murray *et al*[18]
Hip fracture leading to nursing home entry	0.400	0.400	National osteoporosis foundation[19]
Death due to hip fracture	*	0.000	—
Vertebral fracture	0.830	0.930	Oleksik *et al*[20]
Wrist fracture	0.981	1.000	Dolan *et al*[21]
Proximal humerus fracture	0.794	0.973	Kanis *et al*[22]
Non-fatal breast cancer	0.620	0.620	Hutton *et al*[23]
Fatal breast cancer	†	0.000	—
Non-fatal CHD	0.850	0.850	Kanis *et al*[24]
Fatal CHD	*	0.000	—
Death by other causes	*	0.000	—

Baseline population utility: age 50 years = 0.850, age 60 years = 0.829, age 70 years = 0.747, age 80 years = 0.699.
*This value has been set to half the utility in the previous year to simulate the death occurring midway through the time period.
†Data on breast cancer mortality showed that the median time before death was approximately 2 years. We incorporated this by doubling the QALYs value in the previous year for the patient in the year of death.

Table 2 Cost data used in the model

	Age 50 costs (£)		Age 60 costs (£)		Age 70 costs (£)		Age 80 costs (£)	
State	1st year costs	Subsequent annual costs	1st year costs	Subsequent annual costs	1st year costs	Subsequent annual costs	1st year costs	Subsequent annual costs
Hip fracture	4880	—	4880	—	6139	—	8080	—
Hip fracture leading to nursing home entry	29 620	22 298	29 620	22 298	30 857	22 940	32 795	23 997
Death due to hip fracture	8666	*	8666	*	8666	*	8666	*
Vertebral fracture	451	210	451	210	510	210	550	210
Wrist fracture	340	—	340	—	340	—	554	—
Proximal humerus fracture	969	—	969	—	969	—	1584	—
Non-fatal breast cancer	8541	—	8541	—	8541	—	8541	—
Fatal breast cancer	10 981	†	10 981	†	10 981	†	10 981	†
Non-fatal CHD	2058	665	2058	665	2058	665	2058	665
Fatal CHD	2160	*	2160	*	2160	*	2160	*
Death by other causes	0	*	0	*	0	*	0	*

The methodology and sources for calculating costs in the initial and subsequent years for each age group is presented in Kanis *et al*.[24]
*This value has been set to half the ongoing costs in the previous year to simulate the death occurring midway through the time period.
†Data on breast cancer mortality showed that the median time before death was approximately 2 years. We incorporated this by doubling the on going cost in the previous year for the patient in the year of death.

Table 3 The assumed efficacy of the interventions on fracture risk, coronary heart disease event and breast cancer events

	Cost of intervention (£)	Grade of evidence*	Relative risk (95% CI)					
			Spine	Hip	Wrist	Proximal humerus	Breast cancer	Coronary heart disease
Alendronate	301	A2	0.53 (0.42–0.67)	0.46 (0.23–0.91)	0.48 (0.31–0.75)			
Etidronate	163	A1	0.40 (0.20–0.83)					
Risedronate	284	A1	0.63 (0.51–0.78)	0.60 (0.42–0.88)	0.68 (0.43–1.08)	0.46 (0.23–0.94)		
Raloxifene	257	A2	0.65 (0.53–0.79)				0.38 (0.24–0.58)	
Oestrogen A	58	A1	0.58 (0.26–1.30)				1.27 (1.02–1.56)	
Oestrogen B	58	A3	0.58 (0.26–1.30)	0.86 (0.42–1.75)	0.32 (0.13–0.78)	0.63 (0.45–0.89)	1.27 (1.02–1.56)	

*Grading system: A1, Evidence from RCTs in women with established osteoporosis; A2, Evidence from RCTs in women with either established osteoporosis or osteoporosis; A3, Evidence from RCTs in women unselected for bone density.
Blank cells denote that no effect is assumed.

fracture, we simulated 8000 patients, which reduced the standard error of the mean to around to approximately 0.3% for QALYs and 4% for costs of the respective mean values. For patients without a prior fracture, the number of patients required to maintain similar ratios rose to 15 000 due to the lower absolute risk of fracture. These numbers required an approximate processing time per simulation of 50 min, and 90 min, respectively, on a machine containing a 2.0 GHz Athlon Thunderbird processor with 256 mB RAM running Windows 2000 (©Microsoft Corporation).

Once first-order uncertainty is minimized, analyses of second-order uncertainty can be undertaken. These results are most appropriate for NICE as they describe the range of cost per QALY values produced by the current uncertainty in the efficacy of the drugs. Simulating a large number of patients cannot eliminate second-order uncertainty, however, as it reflects inherent gaps in the knowledge base.[25]

We confined analyses of second-order uncertainty to the efficacy values (the six relative risks presented in Table 3) for each treatment. The values of the remaining parameters such as utility and costs were held constant at their mean values. Uncertainty in the efficacy data was large in this case study with wide confidence intervals around the relative risk of fracture for each intervention. Such variations around the central estimate, particularly when modelling up to six independent relative risks, necessitate that extensive sensitivity analyses be undertaken. We assumed that 1000 runs of the individual patient model would be sufficient to adequately describe the distribution in the cost per QALY ratios caused by uncertainty in the efficacy data for each intervention.

In all, 1000 sets of six random numbers were drawn and these were used to sample from the distribution of relative risks for each intervention (Table 3). This approach reduces bias by removing the possibility that a treatment produces better results as an artefact of the random numbers selected. We could not use the individual patient model to calculate directly the results due to the time constraints of the project. This problem is common when using complex individual patient simulation models and, typically, lengthy running times have prohibited the conduct of full and adequate sensitivity analyses.

To resolve this problem we used a methodology based on Gaussian process regression, which is described in detail elsewhere.[26] Gaussian process regression is a non-parametric regression technique. It is assumed that the input–output relationship in the patient simulation model is described by a smooth, continuous function, but no other assumptions about the nature of the function (eg regarding linearity, the presence/absence of interactions) are made. As values of the model output at different input values are obtained, the Gaussian process model 'learns' the input–output relationship and can then be used as a fast approximation. We derived the approximation using a relatively small number (200) of individual patient model runs. It was then possible
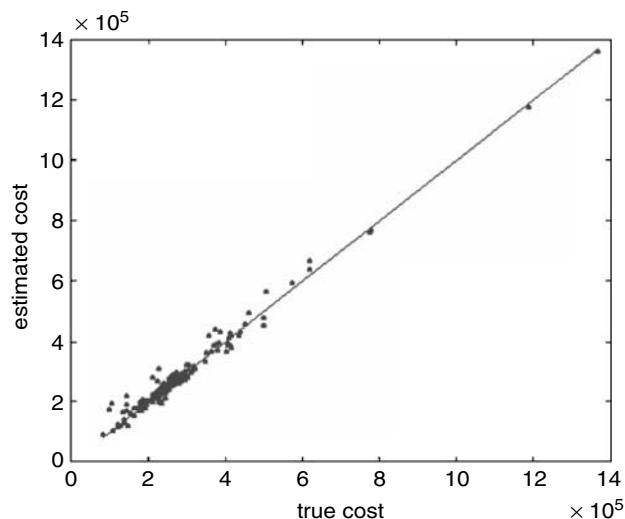

Figure 2   Demonstrating the accuracy of the Gaussian process emulator in predicting the output from the individual patient model.

to predict the output of the model at any set of input values almost instantaneously. By withholding each data point in turn and fitting a Gaussian model to the remaining 199 values, we estimated the accuracy of the statistical relationship. We then compared the output values predicted by each Gaussian model, with the 'true' output from the individual patient model. The approximation was seen to be good (Figure 2).

We produced cost per QALY ratios compared with no treatment (where all relative risks were equal to 1) for each treatment for each of the 1000 sampled sets of relative risks. The mean cost per QALY was provided by dividing the aggregated incremental costs by the aggregated incremental QALYs. We calculated a 95% credible interval by ranking the 1000 samples in order of cost-effectiveness and noting the 25th and 975th values.

Results

We provide illustrative results for those treatments most commonly prescribed: bisphosphonates (alendronate, etidronate and risedronate), selective oestrogen-receptor modulators (raloxifene) and oestrogen (assumed generic) for women aged 70 years with established osteoporosis (Table 4). The full results are presented elsewhere.[15] It is seen that both alendronate and risedronate had mean cost per QALY values below £25 000. The cost-effectiveness ratios for all interventions have wide confidence intervals due to the large uncertainty in the efficacy data. Due to the adverse effects on breast cancer, the confidence interval for oestrogen includes the possibility of incurring costs while reducing health.

Table 4   The incremental costs and QALYs of each intervention compared to no treatment for 100 women aged 70 years with established osteoporosis and *t*-scores of −2.5 standard deviations

	Increase in cost compared to no treatment per 100 patients (£000)	Increase in QALY compared to no treatment per 100 patients	Mean cost per QALY (£000)
Alendronate	95	5.61	17 (10–44)
Etidronate	89	3.01	30 (23–48)
Risedronate	104	4.71	22 (14–43)
Raloxifene	128	4.26	30 (24–41)
Oestrogen A	51	0.74	70 (15–D-ed)
Oestrogen B	38	1.56	25 (D-ing–D-ed)

Numbers in parentheses indicate 95% confidence intervals estimated using a percentile method from 1000 samples.
D-ing represents dominating (costing less and increasing health).
D-ed represents dominated (costing more and reducing health).
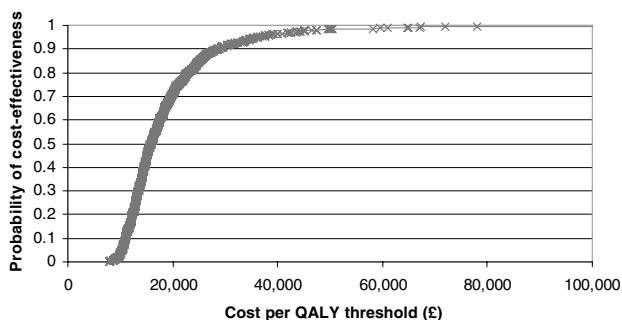
Figure 3   The CEAC for alendronate in women aged 70 years with established osteoporosis and a *t*-score of −2.5 standard deviations.
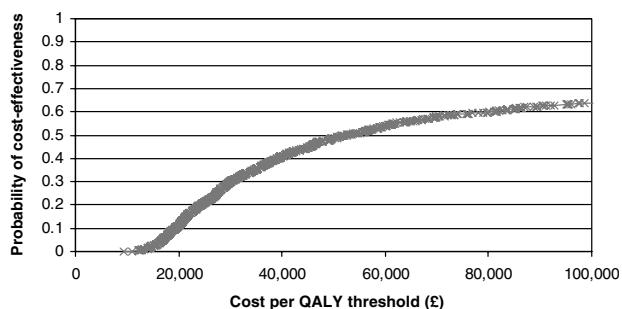

Figure 4   The CEAC for oestrogen in women aged 70 years with established osteoporosis and a *t*-score of −2.5 standard deviations. RCT evidence only used.

The uncertainties in the mean cost per QALY values for alendronate and oestrogen (incorporating observational data) are displayed in the cost-effectiveness acceptability curves shown in Figures 3 and 4. It is seen that, in patients with established osteoporosis, alendronate is 90% likely to
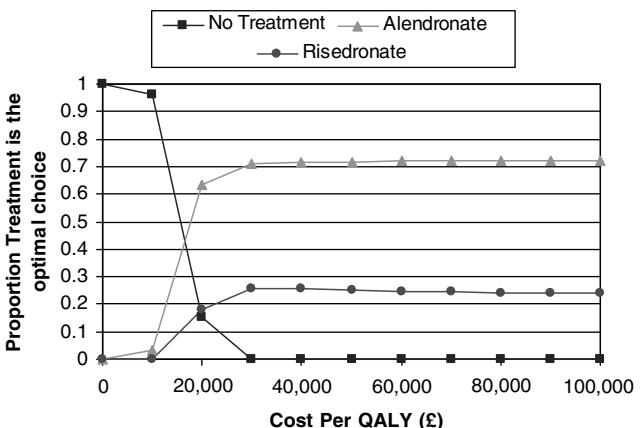

Figure 5   The multi-way CEAC for all treatments for women aged 70 years with a prior fracture and a *t*-score of −2.5 standard deviations.

have a cost per QALY value below £30 000 and is very unlikely to have a cost per QALY above £60 000. Oestrogen has a 30% probability of having a cost per QALY below £30 000, but has a 35% chance of having a cost-effectiveness ratio above £100 000 or worse.

A multi-interventional CEAC is presented in Figure 5, with the omission of etidronate, raloxifene and oestrogen, as the probability of these interventions being optimal was less than 5% at all cost per QALY thresholds. It is seen that at all cost per QALY thresholds above £30 000 it was always optimal to treat with alendronate or risedronate.

Although the results from this analysis appear to favour alendronate over risedronate and etidronate the NICE committee were reluctant to differentiate between these interventions in their guidance (section 4.3.5).[5] Their reasons were based on the fact that no head-to-head trials had been conducted, that the trials on which these data are based differ in some respects including the type of patients involved, that the interventions had different tolerability profiles and that the etidronate trials were not powered to detect a significant reduction in hip fracture rates.

Discussion

An overview of the differences between Markov and individual patient-based models is given in Briggs[27] with an example provided where both methodologies give similar mean answers. However, in this case study, we had reasons for believing the answers may be substantially different.

(1) It has been proven that suffering a prior fracture greatly increases the risk of future fractures, even when adjusted for bone mineral density.[28] For example, the risk of a subsequent vertebral fracture following an initial vertebral fracture is 4.4 times greater than that without a prior fracture. This value is 2.5 were the first fracture at the hip.

(2) The costs and health detriment following a fracture persist and these relationships need to be incorporated into the model.[15]

(3) The residential status of the patient needs to be modelled, as there are high costs and utility detriments associated with nursing home care. Also the risk of mortality following a hip fracture has been shown to differ between those living in the community and those in nursing homes.[29]

(4) We anticipated (incorrectly) that data on the interactions between variables would be found. For example, the subsequent cost or health detriment of a hip fracture may depend on whether a prior vertebral fracture had been suffered, or the elevated risks for subsequent fractures may depend on the time since a previous fracture. It is expected that these relationships will become known in future years. For the above reasons, we must record for each patient, whether any prior fractures have occurred, and if so, when and at which site. These are routinely recorded in an individual patient model.

Such a methodology could theoretically be replicated in cohort models by employing a very large number of transition states. Assuming no deaths, were six conditions and a no-event state to be modelled, it would require 7^{10} (282 million) transition states by year 10. This number would be doubled were the residential status of the patient also recorded. Accordingly, the cohort model introduces simplifications that can affect the accuracy of the results. The potential gain in accuracy through applying an individual patient methodology results is not known as results produced by an individual patient methodology and those produced by a standard cohort methodology have not been compared. It is acknowledged that all models are dependent on the quality of the available data and, while the individual patient approach should theoretically produce more accurate results, the inaccuracy of the cohort methodology may be relatively small compared with the sensitivity of both sets of results to variables where there is large uncertainty. Further research is needed to calibrate the differences in cost-effectiveness ratios expected via an individual patient methodology compared with a cohort model, given a known data set.

We chose a time-slice methodology rather than a discrete event simulation approach due to fewer random numbers being required. The time slice approach requires a maximum of 10 random numbers to be sampled per patient. Discrete event simulation would require a number from 7 random number streams to calculate the next event to occur from the four fractures, breast cancer, coronary heart disease and death through other causes), and then further samples from the seven streams each time an event occurs.

The annual time slice prohibits more than one event occurring each year. However, due to the relative small chance of a fracture ($<4\%$ per annum at age 70) the period of 1 year seemed appropriate, and the results are unlikely to change with a move to shorter time slices.

We conducted analyses using data for costs and utilities fixed at their means. Scarce data were available on costs; however, the uncertainty in the utility values was reasonably well described. We conducted extreme value sensitivity analyses for hip, vertebral fracture and proximal humerus assuming that the upper and lower 95% confidence intervals reported by Murray *et al*[18] for hip fracture were applicable to all three fractures. These were 0.72 in the initial year and 0.81 in subsequent years assuming high detriments, and 0.96 and 1.00, respectively, assuming low detriments. This changed the mean cost per QALY for alendronate at age 70 to £11 424 assuming high detriments and £30 254 assuming low detriments. Probabilistic sensitivity analyses over the full range of the utility distribution rarely changed the cost-effectiveness rankings of the interventions or affected whether the cost per QALY rose above the assumed £30 000 threshold for cost-effectiveness.

While some parameters, such as *t*-score and epidemiological data, are assumed to be constant within each scenario, the effects of changing these variables can be estimated by adjusting the relative risks for both no-treatment and a treatment scenario. For example, an estimate of cost-effectiveness assuming double the risk of hip fracture (either through a lower *t*-score or due to new epidemiological data becoming available) can be obtained by using a relative risk of 2 for hip for no treatment, and doubling the sampled relative risk of hip fracture for treatment. This approach would however introduce slight inaccuracies within the offset period, whereas the relative risk would be assumed to be returning to 1, with both the intervention and no treatment underestimating the number of fractures suffered. Similarly, near the conclusion of the project, data became available on the risk of mortality due to vertebral or proximal humerus fracture. Approximations of these effects were included using a cohort methodology as there was insufficient time to construct new Gaussian models. Our research showed that these approximations do not significantly change the cost-effectiveness results produced. Changing some variables, however, such as the discount rate would require a new batch of individual patient models to be run and new Gaussian models formulated using this data.

Due to resource restrictions, we modelled only four age groups (50, 60, 70 and 80 years old) and female patients. Despite these limitations, this initial work was able to address important policy questions about which treatments are cost-effective for scenarios of interest. The meta-model can also quickly calculate cost-effectiveness ratios when new efficacy data become available for existing drugs, when new treatments for osteoporosis are developed, or where new estimates of the costs and utility values become available.

Conclusion

Previous models to determine the cost-effectiveness of treatments for osteoporosis have used cohort-based approaches. They have inherent weaknesses that can be overcome in an individual patient structure such as that presented. Our approach outlines a methodology for calculating the cost-effectiveness of osteoporosis treatments that allows the accuracy of results produced by an individual patient-based simulation to be maintained, while facilitating extensive sensitivity analyses. The modelling approach has had applications in the real world, forming the basis of an academic review for NICE and has been highly influential in developing the guidance issued.

References

1 Consensus development conference (1991). Diagnosis, prophylaxis and treatment of osteoporosis. *Am J Med* **90**: 107–110.

2 Kanis JA *et al* (2000). An update on the diagnosis and assessment of osteoporosis with densitometry. *Osteoporosis International* **11**: 192–202.

3 Department of Health and the Association of the British Pharmaceutical Industry (1994). *Guidelines for the Economic Evaluation of Pharmaceuticals* (Press release). Department of Health: London.

4 Dolan P and Torgerson DJ (1998). The cost of treating osteoporotic fractures in the United Kingdom female population. *Osteoporosis Int* **8**: 611–617.

5 NICE FAD Reference National Institute for Clinical Excellence (2004). *Appraisal Consultation Document: The Clinical Effectiveness and Cost Effectiveness of Technologies for the Secondary Prevention of Osteoporotic Fractures in Postmenopausal Women*. National Institute for Clinical Excellence: London, Available from: www.nice.org.uk.

6 Drummond MF *et al* (1997). *Methods for the Economic Evaluation of Health Care Programmes*, 2nd edn. Oxford Medical Publications: Oxford.

7 Weinstein MC and Statson WB (1977). Foundations of cost-effectiveness analysis for health and medical practices. *N Eng J Med* **296**: 716.

8 Birch S and Gafni A (1992). Cost-effectiveness/utility analyses: do current decision rules lead us where we want to be? *J Health Econ* **11**: 279–296.

9 Raftery J (2001). NICE: faster access to modern treatments? Analysis of guidance on health technologies. *BMJ* **323**: 1300–1303.

10 Briggs AH and Gray AM (1999). Handling uncertainty when performing economic evaluations of healthcare interventions. *Health Technol Assess* **3**(2).

11 Fenwick E *et al* (2001). Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ* **10**: 779–787.

12 Eastman RC *et al* (1997). Model of complications of NIDDM. Model construction and assumptions. *Diabe Care* **20**: 725–727.

13 Department of Health (1995). *Policy Appraisal and Health*. Department of Health: London.

14 Government Actuary Department (1999). Expectation of life United Kingdom females. http://www.gad.gov.uk/Life_Tables/docs/whltukf.xls

15 Stevenson MD *et al* (2004). A systematic review and economic evaluation of interventions for the prevention and treatment of postmenopausal osteoporosis. *Health Technol Assessment*, in press.

16 Netten A *et al* (1998). Unit costs of health and social care. Canterbury PSSRU, University of Kent: Canterbury, UK.

17 HM Treasury. *GDP and GDP Deflators at Market Prices*, 2000. http://www.hm-treasury.gov.uk/economic_data_and_tools/gdp_deflators/data-gdp-fig.cfm

18 Murray C *et al* (2002). Utility following a fracture in a group of elderly women. *Qual Life Res* **11**: 642.

19 National Osteoporosis Foundation (1998). Osteoporosis: review of the evidence for prevention, diagnosis, and treatment and cost-effectiveness analysis. Introduction. *Osteoporosis Int* **8**: S7–S80.

20 Oleksik A *et al* (2000). Health-related quality of life in postmenopausal women with low BMD with or without prevalent vertebral fractures. *J Bone Miner Res* **15**: 1384–1392.

21 Dolan P *et al* (1999). Health-related quality of life of Colles' fracture patients. *Osteoporosis Int* **9**: 196–199.

22 Kanis JA *et al* (2004). The risk and burden of vertebral fractures in Sweden. *Osteoporosis Int* **15**: 20–26.

23 Hutton J *et al* (1996). A new decision model for cost-utility comparisons of chemotherapy in recurrent metastatic breast cancer. *Pharmacoeconomics* **9**(Suppl 2): 8–22.

24 Kanis JA *et al* (2002). Treatment of established osteoporosis: a systematic review and cost-utility analysis. *Health Technol Assess* **6**(29).

25 Stinnett AA and Paltiel AD (1997). Estimating CE ratios under second-order uncertainty. *Med Decision Making* **17**: 483–489.

26 Stevenson MD *et al* (2004). Gaussian Process modelling in conjunction with individual patient simulation modelling. A case study describing the calculation of cost-effectiveness ratios for the treatment of established osteoporosis. *Med Decision Making* **24**: 89–100.

27 Briggs A (2000). Uncertainty in cost-effectiveness models. *Pharmacoeconomics* **17**: 479–500.

28 Klotzbeucher CM *et al* (2000). Patients with prior fracture have a increased risk of future fractures: a summary of the literature and statistical synthesis. *J Bone Miner Res* **15**: 721–727.

29 Todd CJ *et al* (1999). *Anglian Audit of Hip Fracture 2*. Cambridge Health Services Research Group, University of Cambridge: Cambridge.