Dear Committee members,

Attached are Servier's comments on the DSU report and a document additional to Servier's comments on the DSU report.

The attached document includes the description of an alternative economic model developed on behalf of Servier in response to the consultation on the Assessment Group model used in the development of TA160 and TA161. Servier's document compares the Assessment Group model with the alternative model.

As you are aware, NICE is only considering representations on the Assessment Group model, in line with NICE's undertakings to the Court following the Judicial Review. A new economic model represents new evidence and NICE therefore cannot review a new economic model at this stage of the process.

However, Servier has requested for this additional document to be sent to the Appraisal Committee because it provides the mathematical foundations of Servier's comments on the Assessment Group model. NICE has agreed to this request and is asking the Appraisal Committee to review the document in this context.

Best wishes,

Jeremy

Jeremy Powell
Technology Appraisal Project Manager
National Institute for Health and Clinical Excellence
MidCity Place | 71 High Holborn | London WC1V 6NA | United Kingdom
Tel: 44 (0)20 7045 2248 | Fax: 44 (0)20 7061 9830

ALENDRONATE, ETIDRONATE, RISEDRONATE, RALOXIFENE AND STRONTIUM RANELATE FOR THE PRIMARY PREVENTION OF OSTEOPOROTIC FRAGILITY FRACTURES IN POSTMENOPAUSAL WOMEN

ALENDRONATE, ETIDRONATE, RISEDRONATE, RALOXIFENE, STRONTIUM RANELATE AND TERIPARATIDE FOR THE SECONDARY PREVENTION OF OSTEOPOROTIC FRAGILITY FRACTURES IN POSTMENOPAUSAL WOMEN

FURTHER RESPONSE BY SERVIER LABS LTD TO CONSULTATION

Due to the nature of the consultation process and the need for NICE to break down issues into small manageable parts we believe there is a great danger that the Appraisal Committee dismiss the issues found to be relatively minor differences of opinion and that they miss the bigger picture. This would lead to distorted conclusions and patient care suffering.

Any evidence based, peer review process requires an oversight on all elements which leads to the outputs upon which a decision is formulated. In this instance all of the model parameters need to be considered in their entirety regardless of whether they have been previously discussed or considered. It is often how and where they are used is the issue and this can only have been seen within the full model.

Additionally, the DSU report deflects back to inputs, simplifications and assumptions that they state were made by the Appraisal Committee and with which we have major concerns. For example and with respect, it is difficult to imagine that the Appraisal Committee can determine the most appropriate way of identifying patients and hence the cost of identification in a manner which differs with published guidelines and contradicts the positions of the academic societies and the majority of experts in the field of osteoporosis.

To ensure that we deal with the DSU responses the authors have incorporated these and made further comment set against the background of the original report. This allows the reader to understand the issue, make a judgement about the depth to which DSU has gone to reply and understand the context of the further reply.

The overriding fact is that the original model is not fit for purpose as evidenced by the number and type of errors contained within its framework. As other consultees have also noted it has not be possible to reproduce their work by any other group to date.

In our report we detail 20 errors of fact, misinterpretation or misuse of information. These have not adequately been addressed by the response of the DSU. Additionally

we are concerned about the direction that the Appraisal Committee has given to the developers of the model.

The following are areas where the Appraisal Committee has set parameters for them to work within and our concerns:-

- Weightings applied to clinical risk factors (CRFs) – these are all assigned the same impact which is inaccurate e.g. predictive risk of future fracture is greater with family history than alcohol intake. (see page 28)
- Providing guidance by number of CRFs not by actual fracture risk. The predictive risk should depend upon the combination of risk factors and the interaction risk coefficients which have been supplied by WHO.not on the number of risk factors (page 28)
- Altering the efficacy of interventions (as compared to trial evidence) when entering into model. Efficacy set higher for women without CRFs and set lower for women with CRFs in model as compared to trial evidence. (see page 25)
- Vertebral fracture utility value – Why did the Appraisal Committee ask for the utility score to be entered 27% lower than the evidence ? (page 28)
- Side effect disutility sensitivity analysis – why was Strontium ranelate set at the same disutility when it is agreed that it doesn't have the same side effect profile ? (page 26)
- Costs of fracture – goes against recent evidence (page 27)
- Identification of women at high risk – the Appraisal Committee does not follow the guidelines of RCP or other experts. The report highlights the 10 fold difference in the requirement for BMD scans per patient identified for treatment (page 40).
- Compliance – inadequacy of model to calculate accurately or mirror real life accurately (page 25)

These comments are meant in a manner to allow us to move forward together constructively and produce improved guidance for patients with osteoporosis.

We look forward to the views of the Appraisal Committee subsequent to their meeting and would welcome a meeting with you to discuss any of the issues raised.

Yours sincerely,

[redacted]

On Behalf of
Servier Laboratories Ltd

6th September 2009

THE COST-EFFECTIVENESS OF STRONTIUM RANELATE IN THE UK FOR THE MANAGEMENT OF OSTEOPOROSIS

An evaluation of the NICE model

A supplementary report prepared for Servier Laboratories, UK

John A Kanis [a] Oskar Ström [b] and Fred Borgström [b]

[a]WHO Collaborating Centre, University of Sheffield, Sheffield, UK
[b] i3 Innovus, Stockholm, Sweden and Medical Management Centre, Karolinska Institute, Stockholm, Sweden

*Author for Correspondence: Prof J A Kanis, WHO Collaborating Centre for Metabolic Bone Diseases, University of Sheffield Medical School, Beech Hill Road, Sheffield S10 2RX, UK; Tel: [redacted], Fax: [redacted]; [redacted]

Abbreviations

BMI	Body mass index (computed as kg/m^2)
BMD	Bone mineral density (in this report at the femoral neck measured by dual energy x-ray absorptiometry)
CRFs	Clinical risk factors
CHMP	Committee for Medicinal Products for Human Use
DSU	Decision Support Unit
EC	European Community
FAD	Final appraisal document
FRAX®	Algorithms that assess the probability of fracture related to any combination of clinical risk factors with or without BMD.
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
NICE	National Institute for Health and Clinical Excellence
NOF	National Osteoporosis Foundation (US)
NOGG	National Osteoporosis Guideline Group
OP	Osteoporosis
QALY	Quality adjusted life years
QoL	Quality of life
RCP	Royal College of Physicians, London
RR	Relative risk
ScHARR	School of Health and Related Research, University of Sheffield, UK
T-score	The deviation in SD units that BMD differs from that of the young adult female reference range
WHO	World Heath Organization
WTP	Willingness to pay. The threshold at which intervention can be considered to be cost-effective

Short Summary

1. The National Institute for Health and Clinical Excellence (NICE) in the UK has recently issued final appraisal documents (FADs) on the health economic assessment of interventions for the primary and secondary prevention of osteoporotic fracture that included an appraisal of strontium ranelate. Strontium ranelate was considered to be cost-ineffective except at very low T-scores for bone mineral density (BMD).

2. There have been a number of concerns raised with respect to the construct and assumptions that populate the model used by NICE, compounded by a lack of access to the model. This prompted a judicial review, the outcome of which required NICE to release to the consultees in the appraisal process a fully executable and non-redacted model for evaluation. A model was released by NICE for comment and responses lodged by several consultees, including Servier. The DSU responded to these comments on 14 August 2009.

3. The primary focus of this report is to evaluate the construct and assumptions that populate the model used by NICE and the adequacy of the DSU response to a previous report with particular regard to the consequences for the cost-effectiveness of strontium ranelate.

Transparency and validation.

4. The excel model supplied by NICE estimates the cost-effectiveness based on Gaussian regression functions which are derived from an individual state transition model. The source individual state transition model was not supplied until late in the consultation period so that the Gaussian functions could not be evaluated. Thus, it is not possible to fully evaluate the model and it cannot be considered, therefore, to be fully executable.

5. The DSU claim that the model released was the only model that was the subject of this consultation exercise. The statement implies that the two models are to be viewed as separate entities and have no direct connection. In reality, they are very much linked since the current model relies on the Gaussian functions that are estimated by using the previous model. The current model would incapable of producing any ICERs without the existence of the previous model.

6. The validity of the model cannot be assessed from the data supplied. Nor is there any previous publication available to demonstrate its validity. The DSU report is unhelpful. It is not possible to test the manner by which mortality, fracture risks are accommodated in the model supplied.

7. The model as supplied does not permit alterations to discount rates, body mass index, population mortality, mortality associated with clinical risk factors, time horizon and the estimation of the annual risk of fracture for CRF scenarios other than those pre specified, so that sensitivity analysis around the assumptions cannot be performed.

The use of FRAX®

8. The NICE model does not permit the calculation of 10-year fracture probabilities, so that the integrity of the NICE application of FRAX® cannot be directly addressed. The DSU considers this to be unnecessary.

9. There are discrepancies between the reviewers and NICE in the calculation of annual risks associated with clinical risk factors (CRFs), particularly a prior fracture.

10. The FRAX® algorithms also assess the probability of death related to any combination CRFs. That is, the FRAX® coefficients should be used to adjust the mortality for a specific patient group. The DSU report confirms that this part of FRAX® has not been implemented in the NICE model.

11. There are a number of significant interactions that are incorporated into FRAX® some of which may have been omitted from the NICE model

12. Body mass index (BMI) is set at a fixed value by NICE (26kg/m^2). The use of a fixed BMI is not consistent with the construct of FRAX®. The deficit decreases the accuracy of all risk estimates except at a BMI of 26kg/m^2. The effect is very marked when BMD is not used to estimate risk. A detailed response of the DSU avoids addressing this issue.

13. The risk associated with alcohol intake is incorrect for the exposure recommended by NICE and will adversely affect cost-effectiveness.

14. Whereas FRAX® provides the mechanism to compute the cost-effectiveness according to the specific risk factor, NICE weights all risk factors equally. The impact of this on fracture probability is marked. For example the average ten year probability for women aged 65 years with two risk factors and a T-score of -2.0 SD is 20%, but varies more than two-fold (13 to 29%) depending on the risk factor.

15. A similar inaccuracy results from the presentation of age and BMD in categories. Thus NICE present ICERs in age bands (e.g. 55-59 years) and T-score bands (e.g. T= -3.0 to -3.5 SD).

Time horizon

16. The NICE model uses predominantly a ten-year time horizon which has a large effect on apparent cost-effectiveness. In order to overcome this deficit, the NICE model preserved the time frame but 'bolted on' adjustments to overcome this flaw in the model construct.

17. The estimation of the 'bolt-on' cost consequences which are included in the NICE model were not transparent since they are not mentioned in the HTA report and there was no information on how they are derived. This is partly resolved in the DSU report, but errors noted by the DSU are not corrected. There are no data that

test the sensitivity of the NICE model to changes in the time horizon and no way to test the adequacy of the 'bolt-on' to overcome the intrinsic deficit in the model.

18. A comparator model developed by the reviewer revealed discrepancies in the coefficients to calculate both the long term costs and QALYs which adjust a 10-year time horizon to a lifetime horizon. These were consistently higher in the NICE model than that calculated by the comparator model.

Discount rates

19. Discount rates used are not those recommended by NICE. The model does not allow changes in the discount rates for costs or QALYs.

Compliance

20. Compliance is not modelled where all patients are simulated in the model but an adjustment is made on the cost side. The incremental costs and QALYs gained will be overestimated in the initial group of patients that start treatment but do not adhere.

Comparison of cost-effectiveness

21. The comparator model was populated with the assumptions used by NICE except for those for which the reviewers considered to be unsound or unsafe. Unsafe or unsound assumptions included the setting of the discount rate, the effect of intervention in women with clinical risk factors, the impact of side effects on quality of life, the cost of fractures, the mortality consequences associated with clinical risk factors and the utility weight given for vertebral fracture. These were not considered by the DSU.

22. In the evaluation of strontium ranelate the NICE model consistently gave higher ICERs than the comparator model, an effect more marked at higher values for BMD.

23. The rank order of cost-effectiveness for strontium ranelate in women with a clinical risk factor differed between the models assessed.

24. The difference in cost-effectiveness ratio between the NICE model and the comparator model was more marked at higher T-scores than at lower T-scores.

Intervention thresholds

25. The intervention threshold (the fracture probability at which treatment became cost-effective) varied little with age with a mean value of 37.8% at a willingness

to pay (WTP) of £20,000 and 21.6% and at a WTP of £30,000. Intervention thresholds could not be assessed using the NICE model.

26. The arguments offered by NICE and the DSU that intervention thresholds cannot be based on fracture probability are unsound.

Replication of NICE model

27. The comparator model was populated as closely as possible with the data and assumptions used in the NICE model. The difference between the NICE model and the 'replica' model was less using all NICE assumptions than when the model was populated with empirical data. The difference became even less when a ten year time horizon was used. Thus differences between models was partly explained by the assumptions that could be varied and partly by the 'bolt-ons' used by NICE.

28. At lower T-scores the difference between the models was larger which diminished with increasing T-score. The NICE model provided ICERs that were consistently higher at all T-scores for prior fracture as the CRF, whereas for parental fracture the NICE model initially gives higher ICERs at T-scores up to -1.5 SD and then became lower than the comparator model.

29. Thus the Gaussian functions seem to be more conservative in low risk patient groups than in the comparator model. However, the reason for this could not be investigated since the individual state transition model was not available for review.

30. The DSU report confirms discrepancies between the NICE model and an external model used by NICE

Step-wise analysis of NICE and the comparator model

31. To analyse the main drivers for the difference between the models a step by step approach was taken where the differences in cost-effectiveness between the models were examined by step changes applied to both models. The steps included side effects, the time horizon, preventable deaths, QALY gains beyond 10 years, efficacy in women with CRFs, compliance, utility values for vertebral fracture, mortality after fracture and discount rates.

32. The two components shown to have largest impact on the difference in ICERs between the models were the efficacy assumed in women with clinical risk factors and the long-term mortality associated with fractures.

33. When only assuming that 50% of the efficacy assumed for other CRFs (i.e. other than BMD, age and prior fracture) in the NICE model, the ICER decreased and the comparator model provided the lower ICER of the two models.

Cost-effectiveness of identification strategies

34. There are several errors identified in the computation of the costs of identification outlined in our earlier report. These are not adequately addressed by the DSU.

35. Contrary to the claim by NICE, the approach does not follow the guidance of the Royal College of Physicians, so that the acquisition costs are inflated with an adverse effect on cost-effectiveness

Conclusions

36. Insofar as the model can be examined we conclude that a major difference between the NICE model and the review model resides in the assumptions used to populate the model. Nevertheless, when these are excluded, there are systematic and non-systematic differences that are likely to impact on cost-effectiveness. In addition, the numerous errors found in the accessible parts of the model are likely to impair significantly the stratification of risk and thus the effective targeting of treatment.

37. The use of a fixed BMI, age intervals, intervals of BMD and median risks related to CRFs introduces errors of accuracy that impair markedly the ability to stratify risk of individuals.

38. Intervention thresholds are readily devised on the basis of fracture probability, but not considered feasible by the DSU or NICE

39. The reviewers do not support the view of the DSU that there are no issues that have been raised by consultees which cause it to doubt the validity of the model or that raise justifiable doubts about the appropriateness of the use of the model to inform the guidance.

Introduction

The clinical consequences of osteoporosis reside in the fractures that arise, particularly hip fracture which accounts for the major direct costs. In 1990, the number of osteoporotic fractures estimated in Europe was 2.7 million, with an estimated direct cost in 2004 of €36 billion (£24.5 billion), of which €24.3 (£16.6) billion were accounted for by hip fracture. Costs are expected to rise to €76.8 (£52.4) billion by the year 2050 [Kanis & Johnell, 2005] because of the increasing numbers of the elderly.

Against this background of the burden of osteoporosis, there has been an increase in the number of agents available that have been shown in well controlled studies to decrease the risk of fractures [Delmas, 2002]. Recommendations concerning the use of these agents in the UK and several other countries have been placed in a health economic setting in order to justify resource allocation and form the basis for the development of clinical guidelines. The agency responsible for this in the UK is the National Institute for Health and Clinical Excellence (NICE), which has published several appraisals on the treatment and prevention of osteoporosis [NICE, 2005, 2007a,b], most recently in June 2008 [NICE, 2008a,b]. There has been a great deal of concern about the model construct and the assumptions used to populate the model used in these appraisals [Delmas and Siris, 2008; Kanis et al, 2008f, g; Kanis and Compston, 2008]. Several analyses have revealed major differences in cost-effectiveness measures from those published by NICE for alendronate, strontium ranelate and risedronate [Kanis et al, 2008g; Borgström et al 2009a,b].

Following an unsuccessful appeal to NICE, Servier were given leave to seek a Judicial Review. A major argument was that NICE had not been transparent in providing access to the model used for the appraisal process. NICE argued that transparency was not possible because of the confidential nature of information provided to NICE for use in the model. The High Court found in favour of Servier since NICE had not acted reasonably in securing the release of the information under an appropriate confidentiality arrangement [Mills, 2009]. As a consequence NICE released a version of the model to interested consultees for comment and responses were submitted to NICE for evaluation. This included a report written for Servier [Kanis & Borgstrom, 2009], the content of which was partly edited by NICE. The Decision Support Unit (DSU) issued a response to the comments on the 14th August 2009 [Stevenson and Wailoo, 2009]. No details are provided on the composition of the DSU, but the authors of the report are the architects of the original economic appraisal.

The present report updates the authors' review of the model supplied by NICE to its consultees in the light of the DSU response with a particular focus on the cost-effectiveness of strontium ranelate.

A note on FRAX®

The data supplied in confidence to NICE comprised that used in the development of FRAX®. FRAX® is a computer based algorithm (http://www.shef.ac.uk/FRAX) that provides models for the assessment of fracture probability in men and women [Kanis et al, 2008a,b; WHO, 2008]. The approach uses easily obtained clinical risk factors (CRFs) to estimate 10 year fracture probability. The estimate can be used alone or with

femoral neck bone mineral density (BMD) to enhance fracture risk prediction. In addition, FRAX® uses Poisson regression to derive hazard functions of death as well as fracture. These hazard functions are continuous as a function of time which permit the calculation of the 10-year probability of a major osteoporotic fracture (hip, clinical spine, humerus or wrist fracture) and the 10-year probability of hip fracture. Some of the risk factors affect the risk of death as well as the fracture risk. Examples include increasing age, low body mass index (BMI), low BMD and smoking. Most other risk engines calculate the probability of a clinical event (e.g. a myocardial infarct) without taking into account the possibility of death from other causes. In addition, the FRAX® model can be calibrated for different countries [Kanis et al, 2008a, b; WHO, 2008].

Probability of fracture is calculated in men or women from age, body mass index (BMI) computed from height and weight, and dichotomised risk variables that comprise;

A prior fragility fracture,
Parental history of hip fracture,
Current tobacco smoking,
Ever long-term use of oral glucocorticoids,
Rheumatoid arthritis,
Other causes of secondary osteoporosis,
Daily alcohol consumption of 3 or more units daily.

These variables are entered onto the web site. Femoral neck BMD can additionally be entered as a machine-specific BMD or as a T-score derived from the NHANES III database for female Caucasians aged 20-29 years [Looker et al, 1998]. When entered, calculations give the 10-year probabilities as defined above with the inclusion of BMD.

The relationships between risk factors and fracture risk incorporated within FRAX® have been constructed using information derived from the primary data of nine population based cohorts from around the world, including centres from North America, Europe, Asia and Australia and has been validated in 11 independent cohorts (mainly women) with a similar geographic distribution with in excess of 1 million patient years [Kanis et al, 2007]. The use of primary data for the model construct permits the determination of the predictive importance in a multivariable context of each of the risk factors, as well as interactions between risk factors, and thereby optimises the accuracy by which fracture probability can be computed. The large sample permits the examination of the general relationship of each risk factor by age, sex, duration of follow up and, for continuous variables (BMD and BMI), and the relationship of risk with the variable itself in a manner hitherto not possible. The use of primary data also eliminates publication bias.

In addition to the clinical risk factors, fracture probability varies markedly in different regions of the world [Kanis et al, 2002]. Thus the FRAX® models need to be calibrated to those countries where the epidemiology of fracture and death is known. At present FRAX® models are available for Austria, China, Germany, France, Italy, Japan, Spain, Sweden, Switzerland, Turkey, and the UK and US.

The obvious application of FRAX® is in the assessment of individuals to identify those who would be candidates for pharmacological intervention, and it has been widely used since the launch of the web site, currently receiving on average 55,000 hits daily. Probability-based guidelines are available for many European Countries, including the UK, North America and Japan [Johansson et al, 2009; Kanis et al,2008c;Kurth et al, 2006; Fujiwara et al, 2008; Siris and Delmas, 2008; Siminoski et al, 2007; Dawson-Hughes et al, 2008; Kanis et al, 2008d]. The UK guidance for the identification of individuals at high fracture risk has been developed by the National Osteoporosis Guideline Group (NOGG) based on the translation of existing guidance provided by the Royal College of Physicians (RCP) [RCP, 1999, 2000, 2002] into probability based assessment [Kanis et al, 2008d; National Osteoporosis Guideline Group, 2009].

The model supplied

The economic model (defined as the NICE model in this report) that was supplied for review was based in Excel. The structure, data and assumptions used are described in HTA reports [Stevenson et al, 2005, 2007b]. The excel model estimates the cost-effectiveness based on Gaussian regression functions which are derived from an individual state transition model. The Gaussian functions have been estimated by simulating the cost-effectiveness over intervals for several of the input parameters in the individual state transition model. This approach has been described in an article by the developers of the model [Stevenson et al, 2004]. The Gaussian functions could not be evaluated since the individual state transition model was not included in the model that was provided by NICE. Thus, it is not possible to fully evaluate the model and it cannot be considered, therefore, to be fully executable - at least from the information that has been provided to date. Following a letter to NICE, the individual state transition model was forwarded to consultees, but no extra time was afforded to evaluate this. In addition, the opinion of NICE was that the individual state transition model was not the relevant model. This appears to be an extraordinary position given the detail afforded in the HTA report to the individual state transition model [Stevenson et al, 2007b].

The DSU re-emphasise this position of NICE by noting that '*The model that was provided to consultees was the model with all its functionality as used for the development of the recommendations for TA160/161. It appears that the consultee is requesting a model with additional functionality (where certain inputs could be changed), but this would be a different model, not the one used for the appraisals'*. Indeed it stressed, that it was only the current model that was the subject of this consultation exercise. The statement implies that the two models are to be viewed as separate entities and have no direct connection. In reality, they are very much linked since the current model relies on the Gaussian functions that are estimated by using the previous model. The current model would incapable of producing any ICERs without the existence of the previous model.

This view of the DSU implies that it is not possible to question the preceding individual simulation model or any of the data used in its construction, even though this would be a critical step in verifying the adequacy and accuracy of the current model. Indeed the "model" supplied does not fit with any description of a "fully executable" unredacted model even if these outputs were the only ones used to guide the appraisal.

The following variables cannot be changed to undertake sensitivity analyses

- Discount rate for QALYs
- Discount rate for costs
- Body mass index
- Mortality adjustments to the general population.
- Mortality adjustments in the presence of clinical risk factors
- Baseline population risk of fracture
- Time horizon
- Combinations of CRFs other than the 19 CRF combinations pre-specified

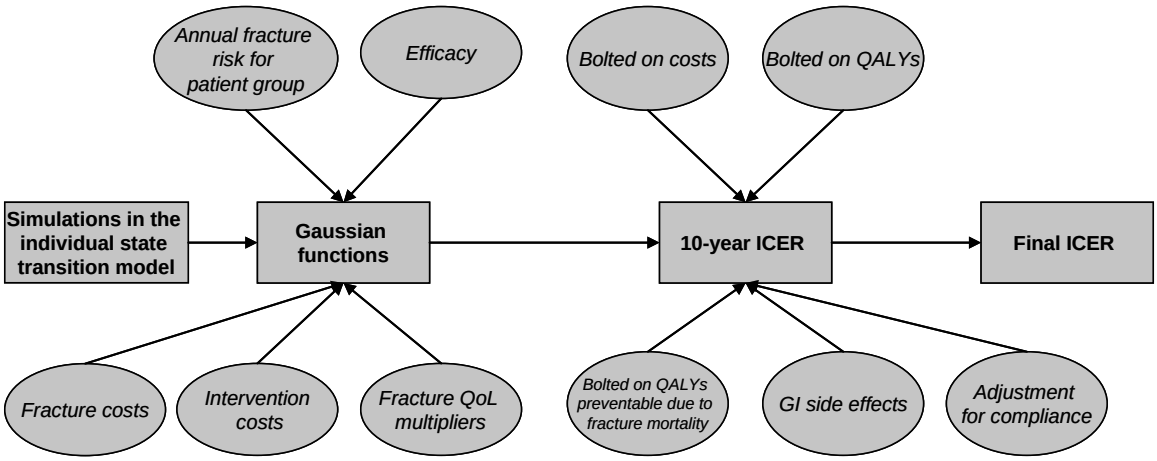
The DSU confirm that all but one of these variables are invariant and support the view that the model cannot be validated or be subjected to sensitivity analysis. The DSU claim that all possible combinations of CRFs are available, but this is not provided for in the model as supplied. The DSU refer to the appraisal of Strontium ranelate for further information [Stevenson et al, 2005c], but this sheds no further light on this claim

What can be reviewed are those components that have been added on top of the Gaussian functions and which are not included in the individual state transition model. The following variables can be changed

- Efficacy of treatment
- Efficacy related to additional CRFs
- Drug costs
- Fracture related costs
- Disutility associated with fractures
- Annual fracture risk at start of treatment
- Compliance (proportion of patients that stop within the first 6 months)

The process for obtaining the incremental cost-effectiveness ratio (ICER) from the individual simulations to the final ICER in the NICE model is depicted in Figure 1.

Figure 1 The process to obtain the ICER in the NICE model



The final ICER, as described in the figure above, for a defined patient group was not the end output for the interpretation of the cost-effectiveness of a treatment. After all the ICERs had been estimated for different numbers of CRFs they were grouped together and the costs related to strategies for identifying these patients (primarily in the assessment of prevention treatment) were added.

The components that can be directly addressed are:

- The incorporation and use of the FRAX® algorithms
- Bolt-ons. These are adjustments to compensate for limitations in the model and include adjustments for considering a lifetime perspective and adjustments for QALYs lost assumed to be preventable due to fracture mortality
- Estimation of the cost-effectiveness for identification strategies

These different components were separately reviewed and are described in separate sections below. In the absence of a fully executable model, the only approach available was to compare the output of the NICE model to a reference or comparator model. It is paradoxical that the DSU declined to review this strategy, cited as providing new and therefore inadmissible information when the DSU present new information (Annex 5) based on essentially the same model. The model used as a comparator is described briefly below.

Comparator model

Even though the individual state transition model is not available, the results based on the Gaussian functions can be compared to another model. If the same data and assumptions are used, the estimated ICERs from both models should be similar, if not identical. Therefore a comparator model which was populated with data and assumptions was used to reproduce the calculations of NICE with the assumptions described in the HTA report [Stevenson et al, 2005, 2007]. The cost-effectiveness was estimated for a number of different assumptions and patient groups in order to analyse differences and key drivers for differences between the models. The cost-effectiveness was primarily analysed using the costs and efficacy related to strontium ranelate.

The simulation model was based on Markov cohort methodology. The model has been extensively used to evaluate the cost-effectiveness of treatments for osteoporosis and hormone replacement therapy in several countries, including the UK [Borgström et al 2006a,b, 2009a, b; Kanis et al 2004, 2005e; Jonsson et al 1995, 1999; Johnell et al 2003, Zethraeus et al 1999, Kanis et al, 2008f]. The model has also been used to compute intervention thresholds, predict fracture rates and mortality making it well validated and calibrated [Kanis 2005b,c; Borgström et al 2006c,d] and provides a reference model for the International Osteoporosis Foundation [Zethraeus et al 2007]. In the model version used for the comparison with the NICE model, the cycle length was set to 6 months and all patients were followed until they died, reached the age of 100 years or the sum of the event probabilities was above 1.

All patients began in the healthy state where each 6 months they had a probability of a fracture of the hip, forearm, spine, or other site or dying. When a fracture occurred, the patient moved to the corresponding fracture health state (i.e. *hip*, *vertebral*, *wrist* or *other fracture)*. The long-term consequences of hip and vertebral fractures were considered in separate health states. Wrist fracture and other osteoporotic fracture were assumed to have an impact on costs and morbidity only in the first year after fracture, and the patient was thus considered to have regained full health one year after the fracture. After a hip fracture, the patient was only at risk for another hip fracture or dying. After a vertebral fracture, the patient was at risk of sustaining a hip or a vertebral fracture or dying. This conservative simplification was adopted because there are few available data on the costs and effects of multiple fractures and, given the low probability of having a vertebral or a wrist fracture after a hip fracture, this discrepancy will have a minor impact on the ICER.

The data used to populate the model were based whenever possible on information from the UK and were the same as those used by NICE in their assessments, unless indicated otherwise.

RESULTS

Validation.

When developing cost-effectiveness models it is important to validate the model to ensure that it is both internally and externally rigid. Internal validation is to ensure that the model calculates correctly according to its specification and the data used. External validation is that the model accurately reflects epidemiology (e.g. fracture risk and mortality), treatment effect and characteristics of the target patient groups both in terms of the data, assumptions and model structure.

The internal validation can be conducted in several ways. One method is to rebuild the model in another software (sometimes also done by a different programmer) in order to replicate results to ensure that there is no programming errors. Another approach is to compare the outputs (such as fracture risks) of the model to other estimations of the output using the same or similar data.

The external validation is also sometimes referred to as methodological uncertainty which arises when comparing study results based on different methods. This most often originates in a disagreement between researchers about the most appropriate method, data and assumptions to be used. This type of uncertainty is often best handled by sensitivity analysis and agreement upon a reference case model.

Unfortunately, the NICE model provided did not have the simulated risk as an output of the model which made it not possible to validate the model through estimated fracture risks or mortality. Nor was it possible to determine the accuracy with which the model reproduced the epidemiology of osteoporosis in the UK. The DSU confirm this position and ‘trust’ that the model is valid rather than to provide a model where the relevant outputs are provided, accessible or previously published. The DSU notes that an HTA

report and a further paper [Stevenson et al, 2005b, 2007] were peer reviewed, but neither article assesses the validity of the model.

In the absence of external validation, internal validation had to be conducted by comparing the specification of the NICE model in the HTA report compared to actual programming and by comparing the ICER outputs with a reference cost-effectiveness model. The DSU states that it believes that the validity of the model structure can be inferred by comparison with another published osteoporosis model that has been used as a reference model for the International Osteoporosis Foundation [Zethraus et al, 2007]. This view can be challenged and is addressed in a subsequent section (see Comparison of models).

The use of risk factors

Annual risk of fracture

The annual risk of fracture was computed by NICE from the data supplied in confidence by the WHO Collaborating Centre for Metabolic Bone Disease at the University of Sheffield, and is now made available for inspection. The FRAX® algorithm uses fracture hazards and death hazards to compute 10-year fracture probabilities for any combination of clinical risk factors (CRFs). The NICE model does not permit the calculation of 10-year fracture probabilities, despite advice from the GDG to the contrary, so that the integrity of the NICE application of FRAX® cannot be directly addressed. Notwithstanding, the DSU states that it does not consider it necessary for 10-year fracture rates to be provided.

By contrast, the NICE model uses a one year time frame. In the NICE model, the annual risks are entered directly as values in the excel sheets and it is not possible, therefore, to evaluate how the actual calculation of the risks were derived. In our original report we noted discrepant values for fracture probability as calculated by NICE and by ourselves, illustrated for a woman aged 70 years and a T-score at the femoral neck of -2.5 SD.

The DSU report offers possible reasons for the discrepancies including the use of the mean age (age 70 years is meant to be age 72.5 years) and a small but uncorrected error in the T-score used by NICE. The DSU state that, apart from the error with entering the T-score into the algorithm, no evidence of other errors was discovered. The DSU states that the WHO algorithm may have changed, but this is not the case.

We have re-estimated the risks using the FRAX with these settings. The differences are now smaller, but they are still not exactly similar. The difference in hip fracture risk differs only by a very small amount. The difference in the risk a major fracture is a little greater which is now slightly higher using FRAX® (lower before). There remains a larger difference when it comes to prior fracture. For example, the one year risk of a major fracture in a woman aged 72.5-year with a BMD T-score of -2.74SD is calculated as 2.78% by us and as 2.38% by NICE.

The errors may arise because NICE do not consider mortality hazards (the DSU has admitted that mortality associated with clinical risk factors was not considered in the model – see below) and that NICE used BMD values from Holt et al [2002] to estimate Z-scores rather than using the NHANES reference standards.

Mortality

The FRAX® algorithms can also be used to assess the probability of death related to any combination CRFs i.e. FRAX® can be used to adjust the mortality for a specific patient group. This part of the FRAX® has not been implemented in the NICE model [Stevenson et al, 2007b]. An exception is the increase in mortality assumed for low BMD values, but it is unclear whether appropriate adjustments are made for the mortality consequences of fracture

The DSU has confirmed that increases in mortality associated with clinical risk factors were not taken account of in the model because they 'could not be easily incorporated'. It notes that 'the incorporation of mortality associated with clinical risk factors is likely to make the ICER estimates higher since the benefits of fracture prevention would result in fewer QALYs being generated. The model therefore is expected to favour the interventions appraised'. This view is not invariably correct since patients with CRFs that are not associated with excess mortality would accrue more QALYs. Indeed, survival is significantly higher the fewer the CRFs.

Interactions.

There are a number of interactions that are incorporated into FRAX®. For the fracture hazards they include 'BMD·BMD', 'prior fracture·age', 'BMD·age' and 'BMI·BMI'. There are additional interactions for the death hazards. The DSU states that the interactions prior fracture·age and BMD·age were incorporated into the model, but make no reference to other interactions.

It may be relevant that the beta coefficients for BMD, prior fracture and BMI would differ in the absence of the interaction terms. Interactions with BMI have not been used, and the consequence of this is evaluated below.

BMI

NICE uses a fixed BMI in the computations of fracture risk [Stevenson et al, 2007b]. It is fixed at 26kg/m^2 for all simulations and cannot be changed in the model as supplied. It is also noted that BMI is used as a dichotomous risk variable by NICE in their case finding strategy. The threshold used is a BMI of 22kg/m^2 [NICE, 2008a, b]. The effect of omitting BMI as a continuous variable is shown in Table 1 for women aged 70 years with a prior fracture. In the absence of BMD, the ten-year probability of a major fracture varied more than two-fold, ranging from 26% with a BMI of 15kg/m^2 to 13% at a BMI of 40kg/m^2. The range of hip fracture probabilities was even greater (from 2.3% to 14%). Variations were less marked, but still evident with the inclusion of BMD at a T-score of -2.5 SD.

Table 1 The effect of BMI on fracture probability for women aged 70 years with a prior fracture. Ten-year fracture probabilities are shown without including BMD and with a T-score for femoral neck BMD is set at -2.5 SD.

	T-score -2.5		No BMD	
BMI	Major	Hip	Major	Hip
15	16	4.7	26	14
20	19	5.3	22	8.4
25	22	5.8	20	4.8
30	21	5.4	17	3.8
35	20	5.0	15	2.9
40	19	4.6	13	2.3

It is evident that the use of BMI as a fixed variable is not consistent with the construct of FRAX®. The deficit decreases the accuracy of all risk estimates except at the value used by NICE. The effect is very marked when BMD is not used to estimate risk. This will have implications where management decisions are given for women without BMD (e.g. with a prior fracture aged 75 years or more). Though the impact is less, there are errors of accuracy incurred when BMD is added to the model.

The DSU has explored the impact of BMI on the fracture risk estimated by the WHO algorithm to test the claim already tested above. Unfortunately, the DSU neglected to test the impact of a low BMI, affecting a substantial minority of women (approximately 15%). Thus the analysis in Annex 2 of the DSU report is inappropriate and the conclusion that 'the use of a BMI of 26kg/m^2 appears to be favourable to intervention, is not sustainable.

The use of a fixed BMI introduces other errors of accuracy in the computation of fracture probability. There is a significant interaction of BMI with BMI and for some outcomes with age [De Laet et al, 2005]. Thus the significance of a step change in BMI differs at different values of BMI and age. There is also a significant effect of BMI on mortality. The phenomenon is illustrated in Table 2 which gives the ratio of fracture probabilities at low values for BMI compared to average values (25kg/m^2) at the ages of 50 and 70 years. At the age of 50 years and a BMI of 15kg/m^2 the 10 year probability of a major fracture is increased by 40%. At the age of 70 years the probability of a major fracture is decreased by 22%. These important interactions are not accommodated in the NICE model.

Table 2 The effect of low BMI on fracture probability ratios for women aged 50 or 70 years with a prior fracture and with a T-score for femoral neck BMD set at -2.5 SD. The ratio of ten-year fracture probabilities are shown at each BMI compared to a BMI of 25kg/m^2 in an individual of the same age.

	Age 50 years		Age 70 years	
BMI	Major	Hip	Major	Hip
15	1.4	1.2	0.78	0.88
20	1.2	1.1	0.92	0.94
25	-	-	-	-

The potential impact on cost effectiveness of strontium ranelate for a woman aged 70 years and a family history of hip fracture is shown in table 3, where, in the absence of BMD, cost effectiveness ranged from £24,300 to £36,100/QALY gained over a modest range of BMI

Table 3 The effect of low BMI on fracture risk and cost-effectiveness of strontium ranelate for women aged 70 years with a family history of hip fracture.

	BMI=20		***BMI=26***		***BMI=32***	
CRFs	major	hip	major	hip	major	hip
Absolute risk (%)	2.46	0.67	2.76	0.70	1.61	0.29
Relative risk	1.29	2.29	1.26	1.16	1.08	0.86
ICER (£000/QALY gained)	24.3		30.6		36.1	

The DSU argue that there is a significant but poor correlation between BMI and BMD and, for this reason, it was decided only to use BMD rather than BMI and BMD. This would only be a logical argument if there were a strong correlation i.e. if BMI could be predicted from BMD which is clearly not the case.

Additional problems arise with the use of BMI in case finding which are reviewed later (see *Cost-effectiveness of identification strategies*)

Intake of alcohol

The FRAX® model accommodates alcohol intake as a dichotomous risk variable. The threshold is set at an average intake of 3 or more units daily and is associated with an increased risk of hip fracture and a major fracture [Kanis et al, 2005f]. The HTA report indicates incorrectly that a threshold value of >2 units daily was used. Notwithstanding, the NICE appraisal chose a threshold of >4 units daily. This is associated with a higher relative risk for fracture than either of the thresholds given above (Table 4). For example, the relative risk of hip fracture (without BMD) is 1.92 for an intake of 3 or

more units daily, but 2.26 at an average intake of 4 or more units daily. Thus the use of the original FRAX® coefficient by NICE underestimates the fracture risk when the threshold is altered.

Table 4 Risk ratio for fracture and 95% confidence intervals according to the intake of alcohol with and without adjustment for femoral neck BMD [Kanis et al, 2005f].

Consumption	Without BMD		Adjusted for BMD	
(units/day)	RR	95% CI	RR	95% CI
Osteoporotic fracture				
>2	1.38	1.16-1.65	1.36	1.13-1.63
>3	1.55	1.26-1.92	1.53	1.23-1.91
>4	1.70	1.30-2.22	1.64	1.24-2.27
Hip fracture				
>2	1.68	1.19-2.36	1.70	1.20-2.42
>3	1.92	1.28-2.88	2.05	1.35-3.11
>4	2.26	1.35-3.79	2.39	1.39-4.09

The DSU confirmed that the coefficient for alcohol used was that given in the FRAX model.

The DSU has not quantitatively explored the effect of the Appraisal Committee's decision to use a threshold of more than 4 units of alcohol on the estimated fracture risks and subsequent ICERs. The error underestimates cost-effectiveness.

Use of risk factors to compute ICERs

The final ICER, as described in Figure 1 above, for a defined patient group was not the final output for the interpretation of the cost-effectiveness of treatment. After all ICERs had been estimated for different numbers of CRFs, they were grouped together. Thus, whereas FRAX® provides the mechanism to compute the cost-effectiveness according to the specific risk factor, NICE weights all risk factors equally.

The impact of this on fracture probability is shown in Table 5. For example the average ten year probability for women aged 65 years with two risk factors and a T-score of -2.0 SD is 20%, but varies more than two-fold (13 to 29%) depending on the risk factor. Other examples are available on the FRAX® web site. The impact of this on resource use is discussed towards the end of the report.

Table 5 Ten-year probability of a major osteoporotic fracture (%) according to BMD T-score at the femoral neck in women aged 65 years from the UK. [Data from FRAX® web site]

Number of CRFs	BMD T-score (femoral neck)					
	-4.0	-3.0	-2.0	-1.0	0	1.0
0	27	15	9.7	7.1	5.9	5.0
1	37 (33-41)	22 (18-26)	14 (10-18)	10 (7.1-14)	8.5 (5.7-12)	7.3 (4.8-10)
2	49 (42-58)	30 (23-40)	20 (13-29)	15 (8.6-23)	12 (6.8-19)	10 (5.6-17)
3	62 (53-72)	41 (30-55)	27 (17-42)	20 (11-34)	17 (8.7-29)	15 (7.2-26)
4	73 (63-81)	52 (42-65)	36 (26-51)	27 (18-41)	23 (14-36)	20 (11-32)

A similar situation pertains when CRFs are accorded equal weights in the absence of BMD. For example, the average ten year probability for women aged 65 years with two risk factors and a BMI of 25 kg/m^2 is 19%, but varies more than two-fold (11 to 29%) depending on the risk factor. Other examples are given in Table 6 and on the FRAX® web site.

Table 6 Ten-year probability of osteoporotic fractures (%) according to body mass index (BMI) in women aged 65 years from the UK. [Data from FRAX web site]

Number of CRFs	BMI (kg/m^2)						
	15	20	25	30	35	40	45
0	11	9.3	8.6	7.4	6.5	5.6	4.9
1	16 (12-21)	14 (10-18)	13 (9.2-16)	11 (7.9-14)	9.8 (6.9-12)	8.5 (5.9-11)	7.4 (5.1-9.5)
2	24 (16-34)	21 (13-31)	19 (11-29)	17 (9.8-26)	14 (8.4-23)	13 (7.3-20)	11 (6.3-18)
3	35 (24-49)	30 (19-45)	27 (16-43)	24 (14-38)	21 (12-34)	18 (10-30)	16 (8.7-27)
4	48 (35-62)	42 (30-57)	38 (26-54)	34 (22-49)	30 (19-44)	26 (16-39)	23 (14-35)

A similar inaccuracy results from the presentation of age and BMD in categories. Thus NICE present ICERs in age bands (e.g. 55-59 years) and T-score bands (e.g. T= -3.0 to -3.5 SD). This makes direct comparisons with the results of NICE problematic because a mean value will differ from a point estimate at a specific age and a specific BMD. For example, cost effectiveness for strontium ranelate is given at £57,500/QALY for women with a prior fracture aged 55-59 years, with a T-score that ranges between -3.0 and -3.5 SD and no clinical risk factors [Stevenson, 2008b]. In the presence of one additional clinical risk factor (assumed to be a prior fracture in the context of the NICE

appraisal), the cost-effectiveness decreased to £46,800 and in the presence of two clinical risk factors was £34,000. The analysis gives an inaccurate estimate of cost effectiveness, since it does not provide information at a specific T-score (e.g. at -3.0 or at -3.5 SD) and a specific age (e.g. at 55 years or at 60 years). Moreover, the cost-effectiveness varies according to the specific risk factor, whereas NICE weights all risk factors equally.

The error is illustrated from the example below in Table 7. The cost-effectiveness for women with a prior fracture aged 55-59 years, with a T-score that lies between -3.0 and -3.5 SD and two clinical risk factors is given by a single estimate in the NICE appraisal of £46,800/QALY gained [Stevenson, 2008b]. In Table 7, cost-effectiveness ranged from £19,200 to £30,100 depending on the T-score, age and the nature of the clinical risk factor. In other words there was a greater than 1.5-fold variation in cost-effectiveness, covered by NICE as a single estimate.

Table 7 Estimates of cost effectiveness for strontium ranelate in women aged 55-59 years with a T-score of -3.0 to -3.5 SD according to the presence or absence of clinical risk factors (CRF)

	NICE	Present study			
	Age 55-59	Age 55		Age 60	
CRF	T=-3.0 to -3.5	T=-3.0	T=-3.5	T=-3.0	T=-3.5
Base case					
Prior fracture	57.5	36.3	28.8	36.0	28.9
Additional CRF	46.8				
Prior fracture + alcohol	na	30.1	23.7	30.0	24.0
Prior fracture + parental history	na	22.3	19.2	22.2	19.2

na = not available

Similar conclusions are reached using point estimates provided as an addendum to the FADs by NICE [NICE 2008c]. For example, the range of cost effectiveness in a woman with a prior fracture and a T-score of -3.0 SD at the age of 65 years is given as £38,499. With a 0.5 decrement in T-score and 5 year increment in age, the ICER decreased to £14,986 - a greater than two-fold variation in cost-effectiveness.

The DSU report that the median ICER was used for ‘simplicity’ and acknowledged that this would favour those women who have a CRF which conferred lower than median increased risk, but would disfavour women who have a CRF which conferred a higher than median increased risk. Thus NICE could not possibly deny that the “median” solution would differ to an ICER which is estimated by properly weighing the different CRFs. The fact that the errors are introduced after the model output and not by the model itself does not lessen the errors. Thus the point is conceded by the DSU that the consideration of age groups, median coefficients and T-score groups decreases the accuracy of the information by which patients’ risk and cost-effectiveness can be stratified.

NICE, however, argue that the consideration of individual CRFs is too complex to provide practical advice, an opinion that they state is supported by the GDG. Against this view, the GDG have consistently recommended that NICE report on fracture probabilities and base intervention thresholds on probabilities using individual CRFs with their appropriate weightings. The argument that this is too complex is flawed; negated by the development of the National Osteoporosis Guidelines by NOGG [NOGG, 2008] supported by many learned societies and patient support organisations such as the National Osteoporosis Society. These guidelines provide practical advice based on the accurate assessment of fracture probability and are increasingly used throughout the UK [Praities, 2009]. Indeed the NOGG website (www.shef.ac.uk/NOGG) receives more than 4000 hits daily. A supplementary argument by NICE that it is not possible to develop intervention thresholds based on probabilities is unfounded. The DSU, however, did not fully consider the manner by which the problem can be remedied (see Intervention thresholds, below).

Conclusion

There are discrepancies between estimates of fracture risk using FRAX® and the estimates derived by NICE. The NICE appraisal neglects the impact of CRFs on the death hazards. The NICE appraisal does not take account of all variable interactions intrinsic to FRAX®. The NICE model makes inappropriate use of BMI, alcohol intake, age and T-score that introduces errors of accuracy which impact significantly on the ICER.

Time horizon

Introduction

Several health economic assessments have drawn attention to discrepancies in estimates of cost effectiveness produced by NICE and other models [Borgström et al 2009a, b; Kanis et al, 2007b, 2008f, g; Kanis and Compston, 2008]. It is difficult to determine why the results differ, but ultimately reasons reside in either the construct of the model or the assumptions used to populate the model. With regard to construct, the NICE model uses predominantly a ten-year time horizon which, as shown in previous sensitivity analyses and elsewhere [Kanis et al, 2007b; Kanis et al, 2008f], has a large effect on apparent cost-effectiveness.

In order to overcome this deficit, the NICE model preserved the time frame but 'bolted on' adjustments to overcome this flaw in the model construct. There are two types of bolt-on factors used in the model to adjust the incremental values estimated from the Gaussian functions. The first adjustment is related to additional QALYs gained beyond 10-years to account for preventable deaths due to avoided fractures during the treatment period of 5 years. These adjustments are described in appendix 10 in the HTA report [Stevenson et al, 2007b]. In short, the expected remaining QALYs for a patient alive at the end of 10-years are multiplied with the number of potentially prevented fracture deaths during the 5-year treatment period. This calculation in the model is consistent with the description in the report. However, in the model there are two values called *wristbonusat2.5* and *phbonusat2.5* that are also added on to the QALYs which are not

described in the report. If these bonuses are also related to preventable deaths it seems to have been assumed that wrist, rib, scapular, clavicular and sternal fractures increase mortality, whereas the report [Stevenson et al, 2007b] indicates otherwise. The DSU could not establish the rationale for including the variables 'phbonusat2.5' and 'wristbonusat2.5', even though the authors of the DSU report are the model architects. Rather than amending the model, the DSU only comment that excluding these variables from the model would be unfavourable to the interventions, but provide no details to document the extent or basis for this claim.

Another issue is that these adjustments only are related to preventable deaths during the 5 years of treatment. However, during the offset period after the intervention where there still is assumed to be a residual effect of treatment there should be an impact on the number of preventable deaths which seems not to be accounted for in the NICE model. Also, the number of fractures and deaths will differ between the comparator interventions even after the 10-years which have an impact on both QALYs and costs which seem not to be accounted for in the model.

The other type of ´bolt-on´ in the model called 'global cost and QALY increases' which provides values that are multiplied with incremental disease related costs and QALYs gained. The estimations of the bolt-on cost consequences which are included in the NICE model are not transparent since they are not mentioned in the HTA report and merely entered as values in the model with no information on how they are derived. The publication of the 'bolt-on' adjustments states that this took account of QALYs related to deaths occurring after 10 years [Stevenson et al, 2005], but none of the other consequences of fracture. The spread sheets provided by NICE suggest that this may be misleading in that it may also account for the cost consequences beyond 10 years, though not the long term effects of fracture on quality of life.

The DSU report that 'following an update based on more recent evidence', this methodology was amended for use in the modelling from 2006 onwards and incorporated into the appraisal thereafter. However, this had not been adequately documented in the Assessment Reports.

Annex 3 of the DSU report redresses this. It also provides some 'sensitivity analysis' of the adequacy of the methodology, but this is done indirectly in such a manner that the DSU has to couch its conclusions with uncertainties. Examples include 'the likely magnitude of the underestimation of the QALYs gained'; 'The likely gain in QALYs'; 'This methodology is likely to be favourable to the interventions';'The DSU believe that on balance the methodology is likely to slightly favour the interventions'; 'In order to calculate the likely degree of underestimation of QALYs gained' ;'to provide an indication of the likely error';' to produce an estimate of the likely underestimation of QALYs' etc.

Unfortunately there are no data that test the sensitivity of the NICE model to changes in the time horizon and no way to test the adequacy of the bolt-on to overcome the intrinsic deficit in the model. The information provided below and later in this report indicate that the belief of the DSU is unfounded

Estimation of 'bolt ons'

In the absence of the fully transparent original model, the only approach available to us was to derive 'bolt on' coefficients from our own FRAX®-incorporated health economic model. As noted previously, the model is well validated and is the model that the DSU used to present supplementary information in their report. Notwithstanding, the DSU declined to comment on our analysis shown below.

In order to examine the adequacy of the 'bolt-ons' we compared the effects of a 10-year time horizon and a lifetime horizon on costs and QALYs in a 70 year old woman with a family history of hip fracture at different T-scores for BMD. For example, the cost at a T-score of -1 SD was £118 using a 10-year horizon, but in reality was greater (£137) as seen when using the lifetime horizon. Thus the cost coefficient for this scenario was 1.16 (137/118). Cost coefficients varied from 1.11 to 1.18 depending on the T-score (Table 8). The QALY coefficients ranged from 1.26 to 1.3.

Table 8 Comparison of costs and QALYs using a lifetime horizon or a 10-year time horizon in women aged 70 years and a family history of hip fracture at the T-scores shown

	T-score					
	-1	-1.5	-2	-2.5	-3.0	≤-3.5
(a) Costs						
Lifetime	-137	-166	-214	-285	-383	-517
10-year	-118	-142	-182	-245	-335	-466
multiplier	***1.16***	***1.17***	***1.18***	***1.16***	***1.14***	***1.11***
(b) QALYs						
Lifetime	0.02	0.019	0.0235	0.0299	0.04	0.050
10-year	0.01	0.015	0.02	0.0234	0.03	0.038
multiplier	***1.26***	***1.27***	***1.27***	***1.28***	***1.28***	***1.30***

These coefficients were compared with the NICE assumptions in Table 9. It is evident that there are discrepancies in the coefficients which are consistently higher in the NICE model than that calculated by ourselves. The actual impact on the results of these differences is that the NICE model will provide more optimistic ICER compared to our comparison model.

Table 9 Multipliers used by NICE and derived from our model to adjust from a lifetime horizon to a 10-year time horizon in women aged 70 years and a family history of hip fracture.

	T-score (SD)					
	≥-1	-1.5	-2	-2.5	-3	≤-3.5
QALYs gained						
NICE	1.40	1.40	1.39	1.38	1.38	1.37
Our model	1.26	1.27	1.27	1.28	1.28	1.30
Incremental disease related costs						
	-1	-1.5	-2	-2.5	-3	≤-3.5
NICE	1.33	1.33	1.32	1.32	1.31	1.30
Our model	1.16	1.17	1.18	1.16	1.14	1.11

Conclusion

There are consistent discrepancies in the coefficients used by NICE and our estimates to calculate both the long term costs and QALYs which adjust a 10-year time horizon to a lifetime horizon.

Risk multipliers for fracture risk

The model cannot accommodate all types of osteoporotic fracture but has states for hip, wrist and humerus fractures. In order to overcome this, fractures at other sites are allocated to an existing state. For example fractures of the femoral shaft are considered to be like hip fractures. Thus the risk of 'femoral' fractures needs to be uplifted to take account of fractures that resemble hip fractures etc.

The risk multipliers found in the report differ from those used in the model. The DSU acknowledge the error made within the Strontium Ranelate Assessment Report [Stevenson et al, 2005], but state that the correct values were used in the model.

Discount rates

The costs and the effects in the NICE model are based on discount rates (costs: 6%, effects: 1.5%) that are not in line with the current NICE recommendations (3.5% for both costs and effects). The discount rates are fixed in the model and cannot be changed. The DSU acknowledge this, but elected not to remedy this, citing historical precedent, even though the new NICE guidelines (2004) came into force before the appraisal process for strontium ranelate and long before the generation of the 2006 ScHARR model used for its appraisal. Using the older discount rates is likely to provide

underestimations of the ICER since the higher discount rates on effects (3.5%) will decrease the QALYs gained.

Adjustments for compliance

In the HTA reports it is assumed that 50% of the patients stop treatment within the first month. The patients that drop out of treatment are not simulated in the model. The patients that are simulated in the model are only those that persist on treatment for the whole intervention period. This is probably because compliance functionality was not implemented at the time it was decided to produce the Gaussian functions. Instead, an adjustment is made on the cost side to account for non-compliers by adding on one additional month of intervention costs. Any adjustment on the effect side is not necessary since non-compliers are not assumed to have any effect of treatment. This approach to account for compliance will overestimate both the incremental costs and QALYs gained [Ström et al, 2009] so that there may not be a major impact on the ICER compared to an approach where all patients are simulated in the model. This has, however not been tested. The DSU elected to provide no comment since this was. considered not relevant to the executable model.

Comparison of cost-effectiveness

The model as supplied did not provide a basis for validation or for estimating the important drivers of cost-effectiveness. With this lack of detail our approach was to reproduce the NICE model using assumptions identical to those given by NICE.

We then compared the cost effectiveness of strontium ranelate with an identical model populated with the assumptions used by NICE with the following exceptions as detailed below with an explanation for the changes.

1. Costs and effects were discounted at 3.5% as recommended by NICE, whereas, the NICE appraisals of osteoporosis used discount rates of 6% for costs and 1.5% for benefits. As noted above, the DSU acknowledge this but elected not to remedy this, citing historical precedent, even though the use of the 2006 ScHARR model in the appraisal of strontium ranelate violates the same precedent.
2. NICE assumed that intervention in women without clinical risk factors had greater efficacy than that shown in clinical trials, and conversely assumed that intervention in women with clinical risk factors had lesser efficacy [NICE 2008a, b] Take, for example, a woman aged 65 years with a T-score of -2.5 SD and no clinical risk factors for fracture. The probability of a major osteoporotic fracture is 12% with a body mass index of 23.8 kg/m^2 (www.shef.ac.uk/FRAX). For an intervention with an efficacy of say 50% (RRR = 0.5), NICE would assume greater efficacy – say 55%. In the same woman who additionally had a family history, the fracture probability rises to 21%. The effects of treatment on the incremental risk (the difference between 12% and 21%) were assumed to be half that of the trial results (an efficacy of 25% in this example). The adjustment was set so that if intervention was used in the phase III setting, the overall efficacy would remain unchanged. The manipulation needs to assume (unlikely, but untested) that the prevalence of clinical risk factors is the same in Phase III studies as in the general population. The justification is based on the view that treatment of women with CRFs is less effective at any given BMD. This has been shown to be untrue in the

many phase III studies addressing this question (reviewed recently in Kanis et al, 2008a, i).

To test the hypothesis directly that a candidate risk factor identified a risk amenable to treatment, it would be necessary to recruit individuals selected on the basis of the risk factor(s) to a randomised controlled trial (RCT). The risk factor that is best evaluated in this way is BMD, and indeed the majority of therapeutic studies have recruited on the basis of low BMD as recommended by regulatory agencies in Europe [CHMP, 2006,]. In recent years, other trials have recruited on the basis of age, gender, a prior vertebral fracture and current exposure to glucocorticoids irrespective of BMD, and have shown therapeutic effects similar to those noted in RCTs based on BMD selection [Saag et al, 1998; Adachi et al, 2001; Reginster at al, 2000].

For other individual risk factors, comparable data are lacking, but several considerations suggest that this concern is misplaced in the contest of the FRAX® risk factors. First, several studies have shown that intervention in the general population induces therapeutic results similar to those expected in individuals selected to be at high risk [Roussow et al, 2002; McCloskey at al 2007b; Chapuy et al, 1994]. Second, studies have shown no significant interaction between response to treatment and the presence or absence of the risk factors used in the present study including age, height, family history of fracture, low body weight or BMI, smoking, alcohol intake or prior non-vertebral fracture [Kanis et al, 2005d; McCloskey el al, 2004; Roux et al. 2006; Marcus et al, 2003; Johnell et al, 2004] including studies with strontium ranelate. Third, the clinical risk factors are not totally independent of BMD and when clinical risk factors alone are used in women aged 70 years or more, BMD is approximately 1 SD lower in the high risk group compared with a low risk group [Johansson et al, 2004]. Perhaps the best evidence is that response to intervention in elderly women recruited from the general population is greater, the higher the probability of fracture estimated without the inclusion of BMD from FRAX® [McCloskey el al, 2009]. Similar finding are reported for the SERM bazedoxifene. In this phase III intervention study, relative risk reduction compared to placebo was greater in women with the higher baseline fracture probabilities [Kanis et al, 2009a]. These considerations suggest that the risk factors chosen are appropriate in that they identify a risk that is amenable to pharmacological intervention. This leads to the conclusion that the NICE assumptions bias cost-effectiveness and unfairly discriminate against women with CRFs. NICE edited this argument from our report and the point was therefore not addressed by the DSU

3. Side-effects were not included in the base case since randomised studies of efficacy have shown few persistent differences between placebo and actively treated patients. Indeed, a disease specific instrument (QUALIOST) showed improvements in quality of life in patients treated with strontium ranelate and a trend in the same direction for a generic instrument (SF-36) [Marquis et al, 2008]. By contrast, NICE assumed that the prevalence and disutility of side-effects for strontium ranelate was the same as that assumed by ScHARR for the bisphosphonates [Lloyd Jones et al 2006]. In the NICE appraisal, the prevalence and consequences of treatment, taken from non-randomised studies, assumed that there would be 23.5 additional GP consultations per 1000 patient months in the

initial treatment period and 3.5 GP consultations subsequently, and the use of a proton pump inhibitor. Symptoms were assumed to persist for 1 month with a utility loss equivalent to a multiplier of 0.91 [Lloyd Jones et al 2006].

The DSU confirms that the ScHARR review of adverse events was based on data for bisphosphonates and that strontium ranelate is not associated with the same adverse gastrointestinal effects as bisphosphonates. The DSU elected to provide no further comment since this was considered not to be relevant to the executable model and provided no explanation on what basis the Appraisal Committee made its decision.

4. Costs of fracture were taken from Stevenson et al [2006] as used previously to determine cost-effectiveness of intervention in glucocorticoid-induced osteoporosis [Kanis et al, 2007b]. These differ greatly from those used by NICE, which were based on now out-dated Health Resource Group codes and are unrealistically low as judged by empirical data in the case of hip fracture, unavailable for vertebral fractures and inappropriate for forearm fractures in the elderly, since a substantial proportion of forearm fractures occur in young individuals [Stevenson et al, 2006]. In addition the incorrect HRG coding was chosen for hip fracture.

   For the present analysis, average in-patient and out-patient costs used were those estimated by ScHARR, and were £10,760 for hip fracture, £9,236 for pelvic fracture, £13,771 for other femoral fractures, £1,706 for vertebral fracture, £527 for forearm fracture, £147 for ribs and sternal fractures, £141 for scapular fractures, £1112 for humeral fractures and £3,864 for fractures of the leg. These did not include any cost for home help. Costs were age-weighted [Borgström et al, 2006c] and included nursing home admissions after hip fracture that increased from 6.7% between the age of 50-59 years to 22.6% at the age of 90 years or more [Zethraeus et al, 2006: McLellan et al, 2007]. These costs, substantially higher than those used by NICE, may be underestimated and empirical data from the UK suggest even higher costs [Lawrence et al, 2005; Johal et al, 2007; Inglesias et al, 2009]. The more recent study estimated costs at £15,133, £2753, £1863, £1331 and £3498 for hip, wrist, arm, vertebral and other fractures, respectively. No response was provided by the DSU, even though one of the authors of the report concedes the inaccuracy.

5. The NICE appraisals took account of the mortality associated with hip and vertebral fractures by assuming that approximately 42% and 28% of deaths from hip and vertebral fractures, respectively are causally attributed to the fracture event [Parker et al, 1991; Kanis et al, 2004a, 2003]. As noted above, the appraisals, however, did not take account of any mortality consequences associated with the presence or absence of other clinical risk factors. These consequences were incorporated in the present analysis.

6. The impact on quality of life the first year after a fracture (hip, vertebral and forearm) was based on empirical estimates [Borgström et al, 2006d]. The quality of life estimates for other fractures were based on expert opinion [Kanis et al, 2004b]. The quality of life in subsequent years after a hip fracture was assumed to be 91% of that of a healthy individual. Forearm fractures were estimated to have no quality of life reduction in the second and subsequent years. The quality of life

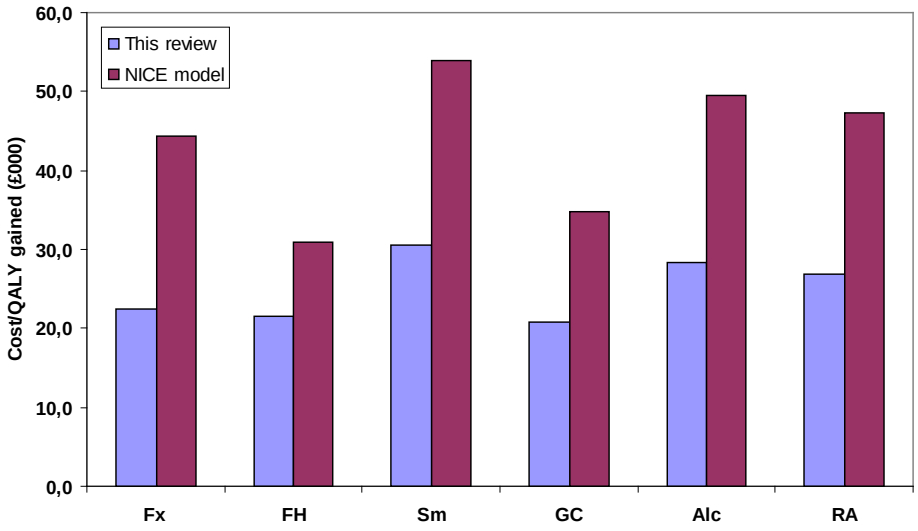
in subsequent years after a vertebral fracture was reduced by 7.1% derived from empirical observations. In an international study when the clinical vertebral fracture may have occurred at a previously unknown time [Oleksik et al, 2000], the utility loss was 9%. These multipliers were used together with the population tariff values for the UK [Kind et al, 1998]. These values are similar to those used by NICE except for vertebral fracture where the utility multiplier in the first year was arbitrarily reduced by the appraisal committee by 27% from 0.626 to 0.792, despite empirical evidence to the contrary at the time of the assessment and now supported by a systematic review by ScHARR [Peasgood et al, 2009]. The item was not considered to be relevant to the executable model and not commented on by the DSU.

The results comparing cost effectiveness between the present study and NICE are summarised in Table 10 and figure 2 for women at the age of 70 years with a T-score of -2.5 SD. The cost-effectiveness ratio was consistently lower in the present study than in the appraisal given by NICE. Indeed recommendations concerning treatment would differ taking a willingness to pay threshold of £20,000 or £30,000/QALY gained. The rank order of cost-effectiveness differed in the present study compared to that of NICE.

Table 10 Incremental cost effectiveness ratio (ICER) of strontium ranelate (£000/QALY gained) compared with no treatment in women aged 70 years with a T-score of -2.5 SD in the presence of a single clinical risk factor (CRF).

	This study		NICE	
CRF	ICER	Ranking	ICER	Ranking
Prior fracture	22.5	3	44.4	3
Family history	21.5	2	31.0	1
Current smoking	30.5	6	53.9	6
Glucocorticoids	20.8	1	34.7	2
Alcohol >3 units daily	28.3	5	49.5	5
Rheumatoid arthritis	26.9	4	47.3	4

Figure 2 Cost-effectiveness of treatment with strontium ranelate (£000/QALY gained) in women aged 70 years with a femoral neck T-score of -2.5 SD and a clinical risk factor as judged by the NICE appraisal and this review



Fx, prior fracture; FH, family history of hip fracture; Sm, smoking; Alc, alcohol intake 3 or more units daily; RA, rheumatoid arthritis.

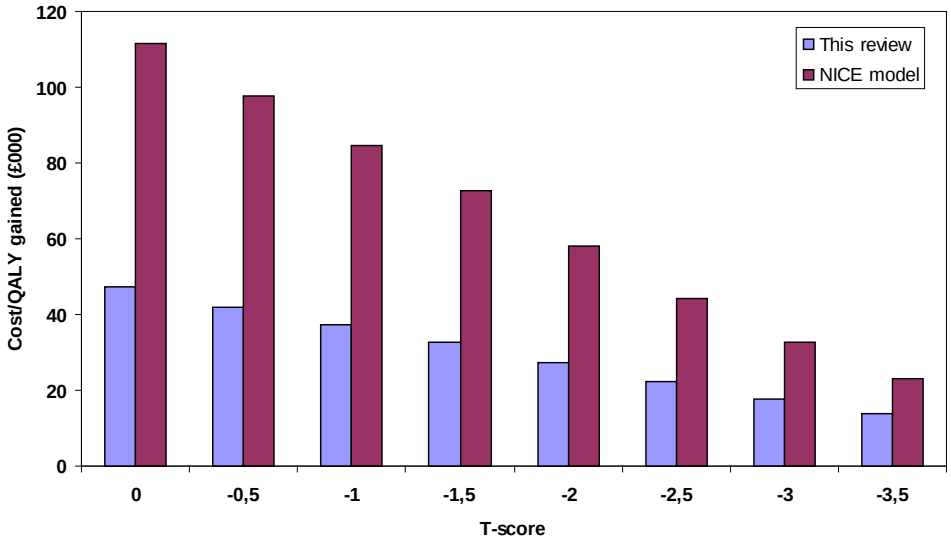
The results comparing cost effectiveness between the present study and NICE are summarised in Table 11 for women at the age of 70 years over a range of T-scores. As noted for a T-score of -2.5 SD, the cost-effectiveness ratio was consistently higher in the present study than in the appraisal given by NICE. As noted above, recommendations concerning treatment would differ taking a willingness to pay threshold of £20,000 or £30,000/QALY gained and the rank order of cost-effectiveness differed in the present study compared to that of NICE.

Table 11 Incremental cost effectiveness ratio (ICER) of strontium ranelate (£000/QALY gained) compared with no treatment in women aged 70 years according to the T-score for BMD in the presence of a single clinical risk factor (CRF).

	T=-1.0		T=-2.0		T=-3.0	
CRF	This study	NICE	This study	NICE	This study	NICE
Prior fracture	37.3	84.5	27.3	58.2	17.6	32.7
Family history	38.2	80.5	27.2	45.5	16.9	19.5
Current smoking	56.9	126.6	39.5	75.4	23.2	36.8
Glucocorticoids	35.7	75.2	26	47.4	16.5	24.3
Alcohol ≥3 units daily	50.6	107.7	36	67.3	21.1	35.0
Rheumatoid arthritis	46.9	99.6	33.8	63.5	22	33.9

The difference in cost-effectiveness ratio was more marked at higher T-scores than at lower T-scores (Figure 3)

Figure 3 Cost-effectiveness of treatment with strontium ranelate (£000/QALY gained) in women aged 70 years with a prior fracture according to femoral neck T-score as judged by the NICE appraisal and this review



In the presence of more than one clinical risk factor the ICER depended on the weight of the clinical risk factor. In the absence of information on BMD, the combination of the weakest two risk factors gave an ICER of less than £30,000 (£27,300) at the age of 70 years. In the presence of the strongest two clinical risk factors (family history and prior fracture) and in the absence of information on BMD test, the ICER lay below £20,000/QALY at the age of 70 years (data not shown). In women aged 70 years with a BMD test and two weak CRFs, the ICER was below £30,000/QALY gained with a T-score of -2.5 SD or less and below £20,000/QALY gained with a T-score of -3.5 SD or less. With two strong CRFs treatment was cost-effective irrespective of BMD.

NICE edited these results from our report and they were therefore not addressed by the DSU

Intervention thresholds

A strength of FRAX® is the ability to express risk as fracture probabilities which are more readily understood than T-scores by physicians and patients. The inappropriateness of the use of a single T-score to direct intervention is now widely acknowledged. Thus probability-based assessment is now becoming the norm for treatment guidelines [Compston et al, 2009; Fujiwara et al, 2008; Tosteson et al, 2008; Dawson-Hughes et al, 2008; Siminoski el, 2007; Kurth et al, 2006;Kanis et al, 2008h,i; NOGG, 2008; Brown et al, 2002; Kanis et al, 2002b; Kanis, 2002, 2008b; WHO, 2007; CHMP, 2006]. The FRAX® models were supplied to avail NICE of the opportunity to use probabilities as intervention thresholds as recommended by the Guideline Development Group of NICE. For reasons discussed below, the NICE appraisal did not consider these feasible or appropriate, a position supported by the DSU.

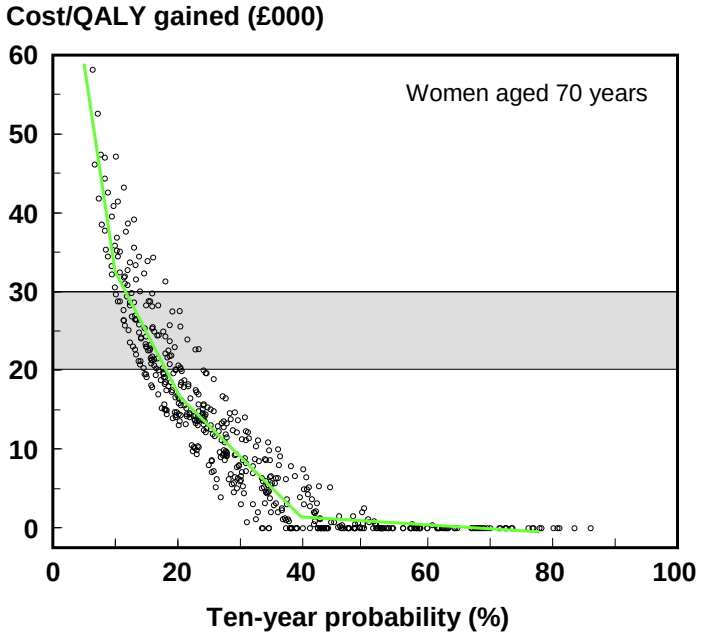
We investigated the validity of this claim. Fracture probabilities were computed using the FRAX® tool calibrated to the epidemiology of fracture and death in the UK [Kanis et al, 2008a]. The algorithms (http://www.shef.ac.uk/FRAX/index.htm) provide two outputs; namely the ten year probability of a major osteoporotic fracture (clinical spine, hip, forearm or humerus) and the ten year probability of a hip fracture alone [Kanis et al 2008a, b; WHO. 2007].

For the purpose of determining intervention thresholds we computed probabilities of a major osteoporotic fracture (rather than hip fracture), for reasons previously argued [Kanis et al 2008i]. Intervention thresholds at each age were determined from the relationship between fracture probabilities and the cost-effectiveness of all possible combinations of CRFs at BMD T-scores between 0 and -3.5 SD in 0.5 SD steps (512 combinations) with a BMI set to 26 kg/m^2. Note that this was not a population simulation, but an array of all possible combinations.

Linear regression was used to estimate the mean value for probability for any willingness to pay. The regression was step-wise linear with four break-points at a probability of 5%, 10%, 20% and 30% for the ages between 50 to 60 years. For older ages break points were set at 10%, 20%, 30% and 40%. The data were not censored at high probabilities since the use of step-wise regression diminishes markedly the effect of high probabilities on low probabilities.

At each age, there was a close correlation between the probability of a major osteoporotic fracture as determined by FRAX® and cost-effectiveness. The relationship is illustrated in Figure 4 for women at the age of 50 years.

Figure 4 Correlation between the ten year probability of a major osteoporotic fracture and cost-effectiveness of strontium ranelate at the age of 70 years in women (BMI set to 26 kg/m^2). Each point represents a particular combination of clinical risk factors



The point estimates for the correlations permit the calculation of the mean fracture probability for any willingness to pay as shown Table 12 for a WTP of £20,000 and £30,000. There was rather little difference in the threshold probability at which

treatment became cost-effective at different ages with a mean value of 37.8% at a WTP of £20,000 and 21.6% and at a WTP of £30,000. Thus, with a WTP of £30,000, any recommendations for intervention should ensure that individuals have a fracture probability that exceeds 21.6%. Probability based thresholds have also been determined for alendronate [Tosteson et al, 2008; NOGG, 2008; Kanis et al, 2008i] and risedronate [Borgström et al, 2006a]

Table 12 Ten year probabilities (mean and 95% confidence intervals; CI) of a major osteoporotic fracture (%) by age at or above which treatment with strontium ranelate becomes cost-effective

	10 year probability of osteoporotic fracture (%) with BMD at a WTP of			
Age	£20 000/QALY		£30 000/QALY	
(years)	Probability	95% CI	Probability	95% CI
50	46.1	16.6-138	17.0	10.0-51.6
55	47.5	19.3-112	19.7	14.4-52.5
60	46.5	22.0-91.6	23.8	16.4-52.8
65	27.8	18.1-52.9	17.1	12.4-25.2
70	25.0	17.0-40.0	15.4	10.3-21.0
75	35.3	21.4-137	26.2	16.0-39.7
80	36.4	20.9- 94.4	32.1	17.3-69.5
mean	37.8		21.6	

To date, treatment of osteoporosis has largely been directed by the level of BMD. The appreciation that age and a variety of clinical risk factors modulate risk and therefore cost-effectiveness, reinforce the view that treatment should be directed on the basis of fracture probability, rather than on a BMD threshold [Brown et al, 2002; Kanis et al, 2002b; Kanis, 2002, 2008b; WHO, 2007; CHMP, 2006]. The preferred metric is the probability of fracture, e.g. the ten-year fracture probability that integrates not only fracture hazards, but also competing death hazards. Thus, from a health economic perspective, an intervention threshold represents the fracture probability at which treatment becomes cost-effective. Intervention thresholds have previously been estimated for the UK [Kanis et al, 2005b; Borgström et al, 2006c], but were based on hip fracture probability alone and not on specific interventions. The analysis above uses the FRAX® tool to determine the average fracture probability above which treatment became cost-effective with the use of strontium ranelate. At a WTP of £20,000, intervention with strontium ranelate became cost-effective at or above a 10-year fracture probability of 37.8% and at or above 21.6% with a WTP of £30,000. Such data could be used to inform clinical practice guidelines, as has been done in the case of alendronate [NOGG, 2008; Kanis et al, 2008i] and risedronate [Borgström et al, 2006a]

NICE have been reluctant to adopt this approach for three reasons, which are variously ill-founded or spurious [NICE 2008a, b] but supported by the DSU. It is claimed that the Appraisal Committee did not have access to the FRAX® algorithms. This, however, is not the case since this was given (in confidence) several years ago. Unfortunately the data were incompletely used and in some instances inappropriately used. A second reason given by NICE for avoiding probability based treatment thresholds in favour of those based on T-scores for BMD is the argument that absolute fracture risk does not provide a single measure of cost effectiveness. This is correct as shown in Figure 4, Table 12 and elsewhere [Stevenson et al, 2007]. For example, at a WTP of £30,000, it was on average cost-effective to intervene in a woman aged 70 years with a 10 year probability of a major fracture of 15.4%. From the different permutations of risk factors this might range from 10.3 to 21.0% i.e. a 2-fold range in the probability estimate. The argument is spurious given that the use by NICE of T-score ranges, median weights for CRFs and age ranges for their economic assessment give an equivalent or greater variance (reviewed earlier). A third reason given by NICE is that the Appraisal Committee 'was not persuaded that the drugs under consideration had been unequivocally shown to reduce fracture risk that was attributable to risk factors not mediated through low BMD and age.' This is ironic given that NICE vary intervention thresholds according to the number of risk factors and even accord less efficacy to strontium ranelate in the presence of risk factors. More disturbing is that the view does not accord with the available evidence [Kanis et al, 2008b, f, i; McCloskey et al, 2009; Kanis et al, 2009a, b: and reviewed in this report].

Comparison of models

Comparison of the reviewers

In order to further explore the validity of the NICE model, we undertook a 'sensitivity analysis' of the NICE model by varying some of the assumptions and parameters to identify the drivers for differences in the results between the NICE model and the comparator model. The comparator model was populated with the data and assumptions used in the NICE model as closely as possible. For all scenarios, we used the fracture cost stipulated by NICE rather than those recommended to NICE [Stevenson et al, 2006]. The efficacy of strontium ranelate was that used by NICE as was the annual cost of strontium ranelate (£334). We omitted the costs related to identification strategies (further elaborated below) as undertaken by NICE.

A comparison of the models populated with identical assumptions is shown in table 13 for women aged 70 years with a T-score of -2.5 SD in the presence of a single clinical risk factor (CRF). Both models were set to account for full efficacy related to the entire range of fracture risk, compliance was set to 50% and disutility after vertebral fracture was set to be equal to hip fracture disutility. It is notable that, in the base case, the comparator model provided higher ICERs than the NICE model except for prior fracture. An explanation for this could be the more conservative 'bolt-on' multipliers, differences in the incorporation of the GI side effects, the inclusion or exclusion of interaction terms for FRAX®, different approaches to account for compliance and the higher annual fracture risks derived from the FRAX® used in the NICE model.

Table 13 Incremental cost effectiveness ratio (ICER) of strontium ranelate (£000/QALY gained) compared with no treatment in women aged 70 years with a T-score of -2.5 SD in the presence of a single clinical risk factor (CRF).

	Comparator model		NICE model	
CRF	ICER	Ranking	ICER	Ranking
Prior fracture	50.5	3	53.3	3
Family history	47.4	1	43.4	1
Current smoking	72.3	6	59.6	6
Glucocorticoids	47.9	2	46.2	2
Alcohol $\geq$3 units daily	66.2	5	56.8	5
Rheumatoid arthritis	63.2	4	55.4	4

Table 14 compares the ICERs when excluding the ´bolt-ons´(i.e. a pure ten-year horizon), GI effects and assuming full compliance. It is now notable that differences between the two models are less. NICE results for starting treatment at 70-years showed slightly higher values for prior fracture, but for other risk factors estimates from the reviewers' model gave higher values for ICER than the NICE model.

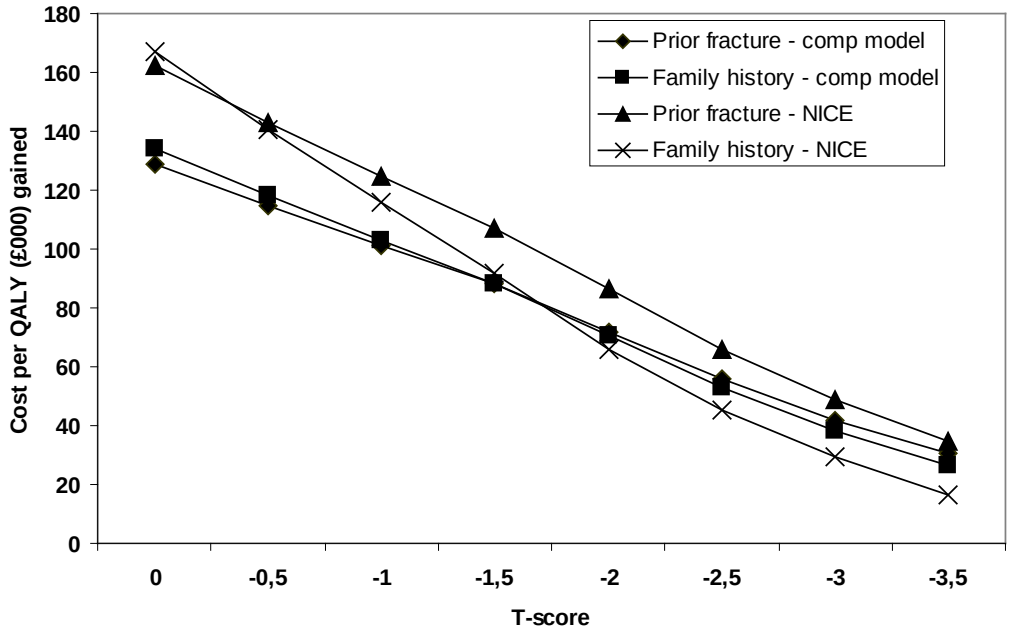
Table 14 Incremental cost effectiveness ratio (ICER) of strontium ranelate (£000/QALY gained) compared with no treatment in women aged 50 and 70 years with a T-score of -2.5 SD in the presence of a single clinical risk factor (CRF).

	50-years		70-years	
CRF	Comparator model	NICE	Comparator model	NICE
Prior fracture	165. 1	192.3	55.7	66.0
Family history	182.2	192.7	53.0	45.4
Current smoking	266.3	232.1	80.1	76.6
Glucocorticoids	188.3	174.9	53.3	51.8
Alcohol $\geq$3 units daily	251.4	231	73.0	71.8
Rheumatoid arthritis	243.3	229.9	69.5	69.2

In Figure 5 the ICER based on the same setup over varying T-score for 70-year old women are shown based on prior and parental fracture as risk factors. At lower T-scores the difference between the models is larger which diminishes with increasing T-score. A

larger difference in low risk patients can to some extent be expected since the QALY gain is very small in these patient groups and small differences in the QALY gain may have a marked impact on the ICER. However, it is notable that the NICE model provides ICERs that are consistently higher at all T-scores for prior fracture as the CRF, whereas for parental fracture the NICE model initially gives higher ICERs at T-score up to -1.5 SD and then becomes lower than the comparator model.

Figure 5 Incremental cost effectiveness ratio (ICER) of strontium ranelate (£000/QALY gained) compared with no treatment in women aged 70 years with a T-score between 0 to -3.5 SD in the presence of a prior fracture or parental fracture



To further explore the reasons for differences between the models we entered our own derived annual FRAX® risks in the NICE model (still using the setup as in the previous analysis shown in table 13 previously. As can be seen (Table 15), the ICERs between the models were very close for women aged 70 years, especially at a T-score of -2.5 SD. This could be seen as an indication that the estimation of the ICER using the Gaussian functions works fairly well. However, the Gaussian functions seems to be better at predicting the ICER in high risk patients since the NICE model consistently gives a higher ICER at T-scores below -2.5 SD and at younger ages. Thus the Gaussian functions seem to be more conservative in low risk patient groups than in the comparator model. However, the reason for this could not be investigated since the individual state transition model was not available for review.

Table 15 Incremental cost effectiveness ratio (ICER) of strontium ranelate (£000/QALY gained) compared with no treatment in women aged 50 and 70 years with a T-score of -1.5 and -2.5 SD when the same annual risks of fracture for each CRF is used in the models

	Comparator	***NICE***	***Comparator***	***NICE***
CRF	*T-score=-1.5*		*T-score=-2.5*	
Age 50years				
Prior fracture	274	313	165	177
Family history	263	307	182	205
Current smoking	492	555	266	281
Glucocorticoids	313	358	188	203
Alcohol >3 units daily	429	489	251	270
Rheumatoid arthritis	401	460	243	264
Age 70years				
Prior fracture	88	95	56	57
Family history	88	93	53	52
Current smoking	138	145	80	79
Glucocorticoids	87	92	53	53
Alcohol >3 units daily	120	127	73	73
Rheumatoid arthritis	112	119	70	70

NICE did not permit the DSU to consider this analysis. Nevertheless, some additional information is available from the DSU report.

Comparison by the DSU

The DSU believes that the validity of the model structure can be inferred by comparison with another published osteoporosis model that has been used as a reference model for the International Osteoporosis Foundation [Zethraeus et al, 2007]. The DSU note that the results produced by the 'NICE' model and the reference model are similar when populated with similar input parameters with regard to the cost-effectiveness of alendronate (Annex 5). The author of the DSU report also notes that the adaptations made to the model to allow for effects beyond the initial 10-year time horizon appear to be appropriate. The data, in the form of a letter to Osteoporosis International was not accepted for publication.

Unfortunately the argument fails on several counts. The first is that this is not a test of external validity. The second is that the results presented in Annex 5 do not show concordance (Table 16). It is notable that the conclusion is based on only 11 numerical examples. Also, the sampling frame is biased by only considering very cost-effective scenarios. It cannot draw any inference that cost ineffective scenarios using the NICE model would also be cost-ineffective using the reference model. In any event, there appears to be large numerical discrepancies when comparing models. For, these reasons we do not support the conclusions of the DSU, and may be the reason why the letter was not accepted for publication.

Table 16 The cost per QALY gained (£000) for women treated with alendronate at a T-Score of –2.5 SD with and without a previous fracture from a previously published analysis [Kanis et al, 2008f] and the NICE model populated with the same assumptions Annex 5 of DSU report]

Age (years)	Kanis et al. [2008f]	NICE model	Difference (%)
T-score= −2.5 SD no previous fracture			
50	14.7	26.0	+77
55	16.2	21.0	+30
60	14.3	17.7	+24
65	7.0	14.0	+100
70	3.7	6.1	+65
75	3.0	1.7	-43
T-score= −2.5 and previous fracture			
50	6.7	8.5	+27
55	7.3	7.4	+1
60	7.3	6.6	-10
65	2.9	5.0	+72
70	0.8	1.4	+75
75	c.s	c.s	-
75	3.0	1.7	-43

c.s. Cost saving

Step-wise analysis of NICE and the comparator model

There are several assumptions and data (e.g. mortality after fracture, CRF related efficacy) that differ between the HTA reports and other cost-effectiveness studies. The impact of these differences was analysed by comparing the ICERs when changing various assumptions and data in the models. To analyse the main drivers for the difference between the models a step by step approach was taken where the differences

in cost-effectiveness between the models were examined by step changes applied to both models. The sequence of changes is outlined in Table 17. As a starting point the models were as similar as possible in terms of data and assumptions. The analysis was conducted for 70-year old women with a T-score of -2.5 SD and parental history of fracture as a CRF.

The results from the step-wise analysis are shown in Table 18. In step 1, the models were kept as similar as possible, but it should be noted that the models were not adjusted to account for the difference in annual fracture risks. The comparator model gives a slightly higher ICER than the NICE model.

The GI effects (step 2) had a higher impact on the costs in the NICE model whilst in the comparator model, the QALY was most affected. However, the ICER in both models increased by a similar magnitude.

In step 3, when adding on a lifetime horizon (by simulation in the comparator model and by adding on the 'bolt-ons' in the NICE model), the ICER decreased more in the NICE model which is related to the higher cost and QALY multipliers used (as shown in a previous section of this report).

In step 4, fracture related QoL gain beyond the ten years was incorporated in the comparator model. The addition of this part in the comparator model decreases the difference the ICER difference between the models.

When only assuming that 50% of the efficacy assumed for other CRFs (i.e. other BMD, age and prior fracture) in the NICE model (step 5) the ICER decreases and the comparator model now provides the lower ICER of the two models.

In step 6 when assuming 50% compliance, the ICER in both models decreased by about the same magnitude in absolute numbers which is expected if it has been implemented correctly. However, the different changes in the incremental values show the different approaches to adjust for adherence. In the NICE model the compliance is adjusted by adding extra intervention costs while in the comparator model 50% of the patients are taken off treatment after 6 months but remain in the model simulation giving an almost halved QALY gained.

Increasing the first year vertebral QoL multiplier provides a lower ICER in both models (step 7). It is notable that changing this QoL multiplier had a larger effect on the cost-effectiveness in the NICE model than in the comparator model.

Including a long-term impact on mortality after vertebral and hip fractures in the comparator model (step 8 and 9) led to a higher QALY gain by avoiding fractures and thus a lower ICER. When using the currently recommended discount rates in the comparator model the ICER decreases due to the higher discount rate for the effects.

Table 17 Step-wise variations in the NICE and comparator models

	GI effects	*Time horizon*	*Additional QALYs related to preventable deaths in NICE model*	*Fracture related QoL gain beyond 10-years in review model*	*CRF efficacy in NICE model*	*Compliance*	*Vertebral disutility*	*Mortality after fracture*	*Discount rates in comparator model**
Step 1	No	10 years	No	No	100%	100%	As hip fracture	First year after hip and vertebral only	Costs: 6%; Effects: 1.5%
Step 2	**Yes**	10 years	No	No	100%	100%	As hip fracture	First year after hip and vertebral only	Costs: 6%; Effects: 1.5%
Step 3	Yes	**Life-time**	No	No	100%	100%	As hip fracture	First year after hip and vertebral only	Costs: 6%; Effects: 1.5%
Step 4	Yes	Life-time	**Yes**	No	100%	100%	As hip fracture	First year after hip and vertebral only	Costs: 6%; Effects: 1.5%
Step 5	Yes	Life-time	Yes	**Yes**	100%	100%	As hip fracture	First year after hip and vertebral only	Costs: 6%; Effects: 1.5%
Step 6	Yes	Life-time	Yes	Yes	**50%**	100%	As hip fracture	First year after hip and vertebral only	Costs: 6%; Effects: 1.5%
Step 7	Yes	Life-time	Yes	Yes	50%	**50%**	As hip fracture	First year after hip and vertebral only	Costs: 6%; Effects: 1.5%
Step 8	Yes	Life-time	Yes	Yes	50%	50%	**QoL multiplier=0.696**	First year after hip and vertebral only	Costs: 6%; Effects: 1.5%
Step 9	Yes	Life-time	Yes	Yes	50%	50%	As hip fracture	**First and following years after hip in review model**	Costs: 6%; Effects: 1.5%
Step 10	Yes	Life-time	Yes	Yes	50%	50%	As hip fracture	**First and following years after hip and vertebral in review model**	Costs: 6%; Effects: 1.5%
Step 11	Yes	Life-time	Yes	Yes	50%	50%	As hip fracture	First and following years after hip and vertebral in review model	**Costs: 3.5%; Effects: 3.5%**

* The discount rates in the NICE model could not be changed

Table 18 Results for the step-wise variation analysis in the NICE and comparator models

	NICE model			Comparator model			
	QALYs gained	*Incremental cost*	*ICER (£000)*	*QALYs gained*	*Incremental cost*	*ICER (£000)*	*ICER difference*
Step 1	0.0262	1188	45.4	0.0234	1241	53.0	-7.64
Step 2	0.0248	1207	48.6	0.0216	1242	57.5	-8.90
Step 3	0.0391	1126	28.8	0.0280	1202	42.9	-14.09
Step 4	0.0391	1126	28.8	0.0351	1202	32.5	-3.69
Step 5	0.0301	1221	40.5	0.0370	1202	32.5	8.00
Step 6	0.0301	1306	43.4	0.0190	685	36.1	7.30
Step 7	0.0346	1306	37.7	0.0200	685	34.3	3.40
Step 8	0.0346	1306	37.7	0.0238	689	29	8.70
Step 9	0.0346	1306	37.7	0.0307	694	22.6	15.10
Step 10	0.0346	1306	37.7	0.0249	716	28.8	8.90

The two different components that were shown to have largest impact on the difference of the ICERs between the models were assumption of only accounting for 50% of the efficacy related to other CRFs and long-term mortality with fractures. Both these components are primarily related to different beliefs in the most appropriate data and assumptions to use to reflect the effect of treatment and epidemiology rather than differences in the structure of the models. Nevertheless, differences remain. It is possible that the differences could be resolved by re-estimating the ´bolt-ons´ but this would further increase the uncertainty of the validity for these adjustments. Another option would be re-estimation of the Gaussian function using a life-time perspective in the individual state transition model. The most correct approach is to rebuild the model.

Cost-effectiveness of identification strategies

The estimated ICERs for specific combinations of CRFs were not directly used in the interpretation of the results. Rather, the cost-effectiveness of treatment included the costs of an identification strategy based on age, T-score and number of CRFs. The first step in the evaluation of the cost-effectiveness of the identification strategy for a given age range is that the average incremental cost and QALYs gained for each T-score range and combination of CRF (grouped in 0, 1, 2 and 3 CRFs) is estimated. If the ICER for a specific scenario is either above the WTP threshold value or no treatment is dominating the incremental values are set to zero. The derived increments are then used to estimate the average cost-effectiveness for a specific number of CRFs for different ages and T-scores.

To assess whether an overall identification strategy is cost-effective at a given age, the incremental values are multiplied with the number in the entire population for England

and Wales that is estimated to fall within each combination of T-score level and number of CRFs at different ages. An example derived from the model for 70-74 year olds is shown in Table 19. This distribution over T-score, ages and CRFs is stated to be based on the data used to develop the FRAX® algorithm. The total population incremental values are then summarised over all T-scores for each number of CRFs. The total costs for the identification strategy, which are derived by multiplying the costs for BMD and physician time with population numbers, are then added to the total population incremental cost. The total identification costs for each number of CRFs are then summarised and divided by the total QALY gained to obtain the cost per QALY gained for the entire identification strategy at each age. If the cost per QALY gained is below the threshold value (£20,000) then the identification strategy is considered cost-effective. If the overall identification strategy is cost-effective then they go back to average ICERs for each T-score level and number of CRFs to determine which patients (based on age, T-score and number of CRFs) should be treated.

Table 19 Number of women in population that are estimated to fall within each combination of T-score level and number of risk factors at the age of 70-74 years

	CRFs			
T-score intervals	*0*	*1*	*2*	*3*
<-5 to -5.5	252	86	22	1
<-4.5 to -5.6	1 086	371	95	5
<-4 to -4.5	4 304	1 468	376	20
<-3.5 to -4	13 345	4 553	1 166	61
<-3 to -3.5	32 385	11 050	2 829	149
<-2.5 to -3	61 514	20 988	5 374	283
<-2 to -2.5	91 462	31 206	7 990	421
<-1.5 to -2	106 455	36 322	9 300	490
<-1 to -1.5	96 998	33 095	8 474	447
<-0.5 to -1	69 186	23 606	6 044	319
<0 to -0.5	38 630	13 180	3 375	178
<0.5 to 0	16 882	5 760	1 475	78
<1 to 0.5	7 704	2 629	673	35
Total	*540 204*	*184 313*	*47 194*	*2 487*

There are several limitations in this approach. Firstly, an average ICER is used to determine the population that would be identified as suitable for treatment. The use of

the average ICER assumes that the prevalence of each CRF is equal. This is clearly not the case [Kanis et al, 2008b, d], and weighted averages should have been used.

A further error is that in the derivations of the identification strategy, cost-effectiveness the NICE model also included the ICERs based on alcohol intake (where the incorrect coefficient was used), and smoking and exposure to glucocorticoids which were CRFs not considered to be relevant risk factors in the NICE appraisal. It further did not include a low BMI as a risk variable – a weakness acknowledged in the HTA report to disadvantage younger women with CRFs, and a low BMI.

A third error is that the distribution of clinical risk factors over T-score and age (said erroneously to be based on the data used to develop the FRAX® algorithm) assumes an identical prevalence of CRFs over the entire range of T-score (see Table 18 above) which is clearly inappropriate. Indeed women above a threshold of probability on the basis of CRFs have a T-score that is approximately 1 SD lower than women below the threshold [Johansson et al, 2004]. Thus, the distribution of risk factors by age does not conform to their known distribution [Kanis et al, 2008i, 2004c].

A further error is in the distribution of the T-score in the population which does not conform to the population from which it was derived [Holt et al, 2002]. The assumed distribution adversely affects cost-effectiveness of the identification strategy, particularly in younger women.

In the case of alendronate, the cost of drug is modelled at twice its actual cost which will adversely affect cost-effectiveness.

A further flaw is that the acquisition algorithm claims to follow the guidance of the Royal College of Physicians. This guidance indicates that women with CRFs would be eligible for a BMD test, and treatment offered to those with a T-score of -2.5 SD. But an important exception is given for women with a prior fragility fracture where intervention may be considered without recourse to BMD testing [RCP, 1999, 2000]. The guidance of the RCP mirrors that of many other clinical guidelines in Europe and North America [Kurth et al, 2006; Kanis et al, 2008h; NOGG, 2008; Lippuner et al, 2009; Siminoski et al, 2007; Dawson-Hughes et al, 2009; EC, 1998; NOF, 2003]. The omission of this aspect of the guidance increases the requirement for BMD tests in the identification strategy and thus inflates the cost. For example, the number of BMD tests to identify a patient for treatment between the ages of 70-74 years is given as 4.6 with a WTP of £20,000 and 5.8 with a WTP of £30,000 [Stevenson et al, 2007b, Table 59]. By contrast, when the approach used by NOGG that follows the RCP guidance is used for the same age range, the average requirement is 0.4 BMD scans per patient identified for treatment [Kanis et al, 2008i]

The DSU note that the identification of women at high risk is fully documented and agreed by the Appraisal Committee. For this reason little further comment was made by the DSU, despite the mathematical errors. The DSU acknowledge some of these errors such as the need to reconsider the way in which risk factors are handled, but declined to explore this since this would require major changes to the macros within the model used to aggregate the cost effectiveness data.

The DSU has established that the price of alendronate used in the model was that at the time of the appraisal. The DSU have no further comment, even though the change invalidates the applicability of the whole appraisal.

Annex 5 of the DSU report that examines the distribution of BMD completely misses the point concerning the errors in the distribution of clinical risk factors over T-score and age and the distribution of T-scores used in the identification strategy.

Discussion

This review arises from concerns surrounding the technology appraisals for osteoporosis. A major difficulty has been the lack of transparency in the model construct and the manner in which the model has been populated. A stumbling block was the reluctance on the part of NICE to negotiate the restricted release of confidential information supplied to NICE by the WHO Collaborating Centre for Bone Diseases at Sheffield. This impediment is now resolved, but the problems remain.

The model supplied for the consultation period remains opaque. It is based on an individual state transition model. The authors claim that individual patient simulations are superior to cohort models in the accuracy with which they are populated and their flexibility. The authors concede, however, that results are unlikely to differ from those using cohort methodology populated with the same assumptions. It is ironic that the model construct appears to be inflexible in that, rather than rebuild the model to fit the requirements of the NICE appraisal, the model has been successively adapted with transformations and add-ons which makes it susceptible to accuracy errors. The adaptations are so extensive that it is quite uncertain whether to define it as a state transition model any longer. Indeed the model supplied is a 'model of a model'.

Unfortunately, the manner whereby these transformations and add-ons were computed and the assumptions used are for the most part opaque and supplied in neither the HTA reports nor the Appraisal documentation. This is only partly redressed in the DSU report. Thus, the reviewers consider that a fully transparent model has not been supplied for evaluation. This is a matter for concern. The concern is accentuated by the observation that, where the reviewers were able to deconstruct the model, in very few instance could we replicate the findings of the authors of the model.

The lack of transparency has a number of consequences. Firstly, the model cannot be externally validated. Secondly, internal validation is problematic. Indeed a component of this review has been to address its internal validation. Thirdly, the transformations have meant that many variables necessary for sensitivity testing cannot be accessed or varied. As a trivial example, it is not even possible to model changes in the discount rates for QALYs and costs, with the result that the rates used do not conform to those recommended by NICE.

A potential strength of the model is the incorporation of the FRAX® algorithms. This permits the integration of multiple risk factors for fracture (and death) to be integrated for the assessment of fracture probability. The obvious application of FRAX® is in the assessment of individuals to identify those who would be candidates for pharmacological intervention, and it has been widely used since the launch of the web site, currently receiving on average 55,000 hits daily. Probability-based guidelines are replacing guidelines based on BMD thresholds and are available for many European Countries. Unfortunately, NICE has not used this methodology for the computation of 10-year probability or for the computation of intervention thresholds. Indeed, it is not

even possible to compute 10-year fracture probability from the model. The arguments of NICE and the DSU for not basing intervention thresholds on fracture probabilities are not robust, and indeed intervention thresholds for strontium ranelate are given within this report.

These considerations apart, a second reason for NICE to use the information provided by FRAX® was to improve the accuracy in stratifying risk and therefore cost-effectiveness. Indeed it is integral to the present model. Since the FRAX® variables were supplied by one of the reviewers, the manner in which the information has been used has in some measure been more readily deconstructed than most of the other add-ons or transformations. The use made of FRAX® is problematic. The death hazards are ignored or inappropriately used, continuous variables ignored, the categorisation of risk factors changed, and risk factors inappropriately used for costing. In addition numerous errors have been identified, each of which may be minor, but compounded have an uncertain effect on accuracy. It is unfortunate that accuracy could not be directly tested since probability estimates are not available from the model supplied.

Several health economic assessments have drawn attention to discrepancies in estimates of cost effectiveness produced by NICE and other models. It is difficult to determine why the results differ, but ultimately reasons reside in the construct of the model, the assumptions used to populate the model or both. With regard to construct, the NICE model uses predominantly a ten-year time horizon, which, as shown in previous sensitivity analyses and in this report has a large effect on apparent cost-effectiveness. The adequacy of the add-ons was impossible to address due to the lack of transparency of the model.

In order to address this issue we have had to build a 'replica' model. One conclusion from the replica model is that a large component of the difference in cost-effectiveness resides in the assumptions used to populate the model. The assumptions that we used were based on empirical observation rather than expert opinion as detailed in the present report. In addition, there were systematic and non systematic differences when the replica model was populated with the same assumptions as the NICE model. The largest discrepancies between the NICE model and the replica model were in the impact of the efficacy assumed in women with clinical risk factors and in the long-term mortality associated with fractures. In addition, at lower T-scores the difference between the models was larger which diminished with increasing T-score. The NICE model provided ICERs that were consistently higher at all T-scores for prior fracture as the CRF, whereas for parental fracture the NICE model initially gives higher ICERs at T-scores up to -1.5 SD and then became lower than the comparator model. The DSU report confirms discrepancies between the model of NICE and an external model. The reasons for this could not be addressed.

Insofar as the model can be examined we conclude that a major difference between the NICE model and the review model resides in the assumptions used to populate the model. Nevertheless, when these are excluded, there are systematic and non-systematic differences that are likely to impact on cost-effectiveness. In addition, the numerous errors found in the accessible parts of the model impair significantly the stratification of risk and thus the effective targeting of treatment.

It is unfortunate that the DSU could or would not consider comments on the effect of different values of input parameters on model outputs (as these had been apparent in the

initial appraisal) or comments made on aspects of the model that had previously been described in Assessment Reports or other consultation documents. This has led to rejecting or ignoring empirical data, preferring to defer to expert opinion. This does not lend itself to a claim that cost-effective analysis is evidence based. This apart, the conclusion of the DSU that there are no issues that have been raised by consultees which cause it to doubt the validity of the model or that raise justifiable doubts about the appropriateness of the use of the model to inform the guidance seem at best to be very overoptimistic.

Acknowledgements and competing interests

Servier Laboratories provided financial support for this study. Servier had no role in study design, in the collection, analysis, and interpretation of data and in the writing of the report.

JAK and FB act as advisors to and have received funding from many pharmaceutical companies involved in marketing products for treatment of osteoporosis. JAK was a member of the NICE Osteoporosis Guideline Development Group (GDG), and was asked by NICE to leave the GDG in 2006. JAK is President of the International Osteoporosis Foundation and FB serves on its Committee of Scientific Advisors.

References cited

Adachi JD, Saag KG, Delmas PD, Liberman UA, Emkey RD, Seeman E et al. Two year effects of alendronate on bone mineral density and vertebral fracture in patients receiving glucocorticoids: a randomised, double-blind, placebo-controlled extension trial. Arthritis Rheum. 2001; 44: 202-11

Borgström F, Carlsson A, Sintonen H, Boonen S, Haentjens P, Burge R et al. The cost-effectiveness of risedronate in the treatment of osteoporosis: an international perspective. Osteoporos Int. 2006a;17: 996-1007

Borgström F, Jonsson B, Ström O, Kanis JA. An economic evaluation of strontium ranelate in the treatment of osteoporosis in a Swedish setting: based on the results of the SOTI and TROPOS trials. Osteoporos Int. 2006b; 17: 1781-3.

Borgström F, Johnell O, Kanis JA, Jonsson B, Rehnberg C. At what hip fracture risk is it cost-effective to treat? International intervention thresholds for the treatment of osteoporosis. Osteoporos Int. 2006c; 17: 1459-71.

Borgström F, Zethraeus N, Johnell O, Lidgren L, Ponzer S, Svensson O et al. Costs and quality of life associated with osteoporosis-related fractures in Sweden. Osteoporos Int. 2006d;17:637-50.

Borgström F, Ström O, Coelho J, Johansson H, Odén A, McCloskey E, Kanis JA (2009a) The cost-effectiveness of strontium ranelate in the UK for the management of osteoporosis. Osteoporos Int, in press Jun 10. [Epub ahead of print] PubMed PMID: 19513577

Borgström F, Strom O, Coelho J, Johansson H, Odén A, McCloskey E, Kanis JA (2009b) The cost-effectiveness of risedronate in the UK for the management of osteoporosis. Osteoporos Int. In press

Borgström F, Zethraeus N, Johnell O, Lidgren L, Ponzer S, Svensson O et al. Costs and quality of life associated with osteoporosis-related fractures in Sweden. Osteoporos Int. 2006d;17:637-50

Peasgood T, Herrmann K, Kanis JA, Brazier JE. An updated systematic review of health state utility values for osteoporosis related conditions. Osteoporos Int. 2009; 20: 853-68.

Brown J, Josse RG, for the Scientific Advisory Council of the Osteoporosis Society of Canada. Clinical practice guidelines for the diagnosis and management of osteoporosis in Canada. Canad Med Assoc J. 2002; 167 (suppl 10); S1-S34.

Chapuy MC, Arlot ME, Delmas PD, Meunier PJ (1994) Effect of calcium and cholecalciferol treatment for three years on hip fractures in elderly women. BMJ 308: 1081-2

Committee for Medicinal Products for Human Use (CHMP). Guideline on the evaluation of medicinal products in the treatment of primary osteoporosis. Ref CPMP/EWP/552/95Rev.2. London, CHMP. Nov 2006.

Compston J, Cooper A, Cooper C, Francis R, Kanis JA, Marsh D, McCloskey EV, Reid DM, Selby P, Wilkins M; National Osteoporosis Guideline Group (NOGG). Guidelines for the diagnosis and management of osteoporosis in postmenopausal women and men from the age of 50 years in the UK. Maturitas. 2009 Feb 20;62(2):105-8.

Dawson-Hughes B, Tosteson AN, Melton LJ 3rd, Baim S, Favus MJ, Khosla S, et al. Implications of absolute fracture risk assessment for osteoporosis practice guidelines in the USA. Osteoporos Int. 2008; 19: 449-58.

De Laet C, Kanis JA, Oden A, Johanson H, Johnell O, Delmas P et al. Body mass index as a predictor of fracture risk: a meta-analysis. Osteoporos Int. 2005;16:1330-8.

Delmas PD, Siris ES. NICE recommendations for the prevention of osteoporotic fractures in postmenopausal women. Bone. 2008; 42:16-18.

Delmas PD. Treatment of postmenopausal osteoporosis. Lancet 2002; 359: 2018-26.

European Community. Report on osteoporosis in the European Community. 1998. EC, Strasbourg

Fujiwara S, Nakamura T, Orimo H, Hosoi T, Gorai I, Oden A et al. Development and application of a Japanese model of the WHO fracture risk assessment tool (FRAX™).Osteoporos Int 2008; 19: 429-448.

Holt IG, Khaw K, Reid DM, Compston JE Bhalla A, Wolff AD et al. Prevalence of osteoporotic bone mineral density at the hip in Britain differs substantially from the US over 50 years of age: implications for clinical densitometry . Br. J. . Radiol., 75 (2002), 736–742

Iglesias CP, Manca A, Torgerson DJ The health-related quality of life and cost implications of falls in elderly women Osteoporos Int 2009;20:869-78

Johal K, C. Boulton C, O. Sahota O, C. Moran C. Cost versus funding for hip fracture treatment: Financial suicide for NHS Trusts? Osteoporos Int (2007) 18 (Suppl 3):S245–S328

Johansson H, Oden A, Johnell O, Jonsson B, De Laet C, Oglesby A et al. Optimisation of BMD measurements to identify high risk groups for treatment – A test analysis. J Bone Miner Res. 2004; 19; 906-13.

Johansson H, Kanis JA, Borgström F, Ström O, Svensson O, Mellström D (2009) FRAX® ett stöd för frakturprevention. Läkartidningen, in press

Johnell O, Kanis JA, Black DM et al (2004) Association between baseline risk factors and vertebral fracture risk in the Multiple Outcomes of Raloxifene Evaluation (MORE) study. J Bone Miner Res 19: 764-772

Johnell O, Jonsson B, Jonsson L, Black D. Cost effectiveness of alendronate (Fosamax) for the treatment of osteoporosis and prevention of fractures. Pharmacoeconomics. 2003;21:305-14

Johnell O, Kanis JA, Oden A, Johansson H, De Laet C, Delmas P, Eisman JA, Fujiwara S, Kroger H, Mellstrom D, Meunier PJ, Melton LJ III, O'Neill T, Pols H, Reeve J, Silman A, Tenenhouse A (2005) Predictive value of BMD for hip and other fractures. Journal of Bone and Mineral Research; 20: 1185-1194.

Jonsson B, Christiansen C, Johnell O, Hedbrandt J. Cost-effectiveness of fracture prevention in established osteoporosis. Osteoporos Int. 1995;5:136-42.

Jonsson B, Kanis JA, Dawson A, Oden A, Johnell B . Effect and offset of effect of treatments for hip fracture on health outcomes. Osteoporos Int. 1999; 10: 193-9.

Kanis JA, Johnell O, Oden A, Sernbo I, Redlund-Johnell I, Dawson A, de Laet C, Jonsson B (2000a) Long-term risk of osteoporotic fractures in Malmo. Osteoporisis International 11; 669-674.

Kanis JA. Diagnosis of osteoporosis and assessment of fracture risk. Lancet. 2002; 359: 1929-36.

Kanis JA, Johnell O, De Laet C, Jonsson B, Oden A, Oglesby A. International variations in hip fracture probabilities: implications for riskassessment J Bone Miner Res 2002; 17: 1237-44.

Kanis JA, Black D, Cooper C, Dargent P, Dawson-Hughes B, De Laet C et al on behalf of the International Osteoporosis Foundation and National Osteoporosis Foundation. A new approach to the development of assessment guidelines for osteoporosis. Osteoporos Int. 2002b; 13: 527-36.

Kanis, JA, Oden A, Johnell O, De Laet C, Jonsson B, Oglesby AK. The components of excess mortality after hip fracture. Bone, 2003;. 32: 468-73.

Kanis JA, Borgstrom F, Johnell O, Jonsson B, Cost-effectiveness of risedronate for the treatment of osteoporosis and prevention of fractures in postmenopausal women. Osteoporosis International; 2004; 15: 862-71

Kanis, JA, Oden A, Johnell O, De Laet C, Jonsson B. Excess mortality after hospitalisation for vertebral fracture. Osteoporos Int. 2004a: 15; 108-12.

Kanis JA, Johnell O, Oden A, Borgstrom F, Zethraeus N, De Laet C et al. The risk and burden of vertebral fractures in Sweden. Osteoporos Int 2004b; 15: 20-6.

Kanis JA, Johnell O, De Laet C, Johansson H, Oden A, Delmas P, Eisman J, Fujiwara S, Garnero P, Kroger H, McCloskey EV, Mellstrom D, Melton III LJ, Pols H, Reeve J, Silman A, Tenenhouse A (2004c) A meta-analysis of previous fracture and subsequent fracture risk. Bone; 35: 375-382.

Kanis JA, Johnell O. Requirements for DXA for the management of osteoporosis in Europe. Osteoporos Int. 2005; 16: 229-238.

Kanis JA, Borgstrom F, Zethraeus N, Johnell O, Oden A, Jonsson B. Intervention thresholds for osteoporosis in the UK. Bone 2005b; 36; 22-32.

Kanis JA, Johnell O, Oden A, Borgstrom F, Johansson H, De Laet C et al. Intervention thresholds for osteoporosis in men and women: a study based on data from Sweden. Osteoporos Int. 2005c; 16: 6-14.

Kanis JA, Barton I, Johnell O (2005d) Risedronate decreases fracture risk in patients selected solely on the basis of prior vertebral fracture. Osteoporos Int 16: 475-482

Kanis JA, Borgstrom F, Johnell O, Oden A, Sykes D, Jonsson B. Cost-effectiveness of raloxifene in the UK. An economic evaluation based on the MORE study. Osteoporosis Int. 2005e; 16: 15-25

Kanis JA, Johansson H, Johnell O, Oden A, De Laet C, Eisman JA, Pols H, Tenenhouse A. Alcohol intake as a risk factor for fracture. Osteoporos Int. 2005f;16: 737-42

Kanis JA, Oden A, Johnell O, Johansson H, De Laet C, Brown J, et al . The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. Osteoporos Int 2007; 18: 1033-46

Kanis JA, Stevenson M, McCloskey EV, Davis S, Lloyd-Jones M. Glucocorticoid-induced osteoporosis: a systematic review and cost-utility. analysis. Health Technol Assess. 2007b; 11:1-256.

Kanis JA, Compston JE; National Osteoporosis Guideline Group of the UK. NICE continues to muddy the waters of osteoporosis. Osteoporos Int. 2008;19:1105-7.

Kanis JA, Johnell O, Oden A, Johansson H, McCloskey EV. FRAX™ and the assessment of fracture probability in men and women from the UK. Osteoporos Int 2008a; 19: 385-397.

Kanis JA on behalf of the World Health Organization Scientific Group. Assessment of osteoporosis at the primary health-care level. Technical Report. 2008b WHO Collaborating Centre, University of Sheffield, UK.

Kanis JA, Burlet N, Cooper C, Delmas PD, Reginster J-Y, Borgstrom F, Rizzoli R, on behalf of the European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO). European guidance for the diagnosis and management of osteoporosis in postmenopausal women. Osteoporos Int 2008ch; 19: 399-428.

Kanis JA, McCloskey EV, Johansson H, Strom O, Borgstrom F, Oden A and the National Osteoporosis Guideline Group. Case finding for the management of osteoporosis with FRAX® - Assessment and intervention thresholds for the UK. Osteoporos Int 2008di; 19: 1395-1408 Erratum 2009 Osteoporos Int 20, 499-502.

Kanis JA, Adams J, Borgström F, Cooper C, Jönsson B, Preedy D, Selby P Compston J. Cost-effectiveness of alendronate. Bone 2008f; 42: 4-15

Kanis JA, Adams J, Borgström F Cooper C , Jönsson B, Preedy D, Selby P, Compston J (2008g) Modelling cost-effectiveness for osteoporosis. Bone, 43:215-6.

Kanis JA and Borgstrom F 2009. The cost-effectivenss of strontium ranelate in the UK for the management of osteoporosis. An evaluation of the NICE model. A report prepared for Servier Laboratories, UK, June 26th.

Kanis JA, Johansson H, Oden A, McCloskey EV (2009a) Bazedoxifene reduces vertebral and clinical fractures in postmenopausal women at high risk assessed with FRAX®. Bone 44: 49-54.

Kanis JA, Oden A, Johansson H, Borgström F, Ström O, McCloskey E (2009b) FRAX® and its applications to clinical practice. Bone. 2009; 44: 734-743.

Kind P, Dolan P, Gudex C, Williams A. Variations in population health status: results from a United Kingdom national questionnaire survey. BMJ. 1998; 316: 736-41.

Kurth AA, Pfeilschifter J. [Diagnosis and treatment of postmenopausal osteoporosis and osteoporosis in men. German Guidelines Update 2006] Orthopade. 2007; 36: 683-90. German

Lawrence TM, White CT, Wenn R, Moran CG. The current hospital costs of treating hip fractures. Injury. 2005; 36: 88-91; discussion 92.

Lloyd Jones M, Wilkinson A. Adverse effects and persistence with therapy in patients taking oral alendronate, etidronate or risedronate: a systematic review. 2006 NHS R & D HTA. ScHARR

Lippuner K, Johansson H, Kanis JA, Rizzoli R. FRAX® assessment of osteoporotic fracture probability in Switzerland. Osteoporos Int. 2009 Jun 11. [Epub ahead of print] PubMed PMID: 19517155.

Looker AC, Wahner HW, Dunn WL, Calvo MS, Harris TB, Heyse SP. Updated data on proximal femur bone mineral levels of US adults. Osteoporos Int 1998; 8: 468-486.

McCloskey EV, Johansson H, Oden A, Vasireddy S, Kayan K, Pande K, Jalava T, Kanis JA (2009)Ten-year fracture probability identifies women who will benefit from clodronate therapy – additional results from a double blind, placebo controlled randomised study. Osteoporos Int. 20; 811-818.

McCloskey EV, Selby P, Davies M et al (2004) Clodronate reduces vertebral fracture risk in women with post-menopausal or secondary osteoporosis: results of a double blind placebo-controlled 3 year study. J Bone Miner Res 19: 728-736

McCloskey EV, Beneton M, Charlesworth D, Kayan K, deTakats D, Dey A et al (2007b) Clodronate reduces the incidence of fractures in community dwelling elderly women unselected for osteoporosis: results of a double-blind, placebo-controlled randomized study. J Bone Miner Res 22:135-141

McLellan AR. Reid DM, Forbes K, Reid R, Campbell C, Gregori A et al. Effectiveness of strategies for the secondary prevention of osteoporotic fractures in Scotland. CEPS99/03. www.nhshealthquality.org/nhsqis/controller?p_service=Content.show&p_applic=CCC&pContentID=2755 1999 (accessed 6th May 2007).

Marcus R, Wang O, Satterwhite J, et al (2003) The skeletal response to teriparatide is largely independent of age, initial bone mineral density and prevalent vertebral fractures in postmenopausal women with osteoporosis. J Bone Miner Res 18: 18-23

Marquis P, Roux C, de la Loge C, Diaz-Curiel M, Cormier C, Isaia G, Badurski J, Wark J, Meunier PJ (2008) Strontium ranelate prevents quality of life impairment in post-menopausal women with established vertebral osteoporosis. Osteoporos Int. 19:503-10.

National Institute for Clinical Excellence. Bisphosphonates (alendronate, etidronate, risedronate), selective oestrogen receptor modulators (raloxifene) and parathyroid hormone (teriparatide) for the

secondary prevention of osteoporotic fragility fractures in postmenopausal women. London, NICE, 2005.

National Institute for Health and Clinical Excellence. Addendum to Final appraisal determination. Alendronate, etidronate, risedronate, raloxifene, strontium ranelate and teriparatide for the primary/secondary prevention of osteoporotic fragility fractures in postmenopausal women. London, NICE. June, 2008c.

National Institute for Health and Clinical Excellence. Final appraisal determination. Alendronate, etidronate, risedronate, raloxifene, strontium ranelate and teriparatide for the secondary prevention of osteoporotic fragility fractures in postmenopausal women. London, NICE. June 2007a.

National Institute for Health and Clinical Excellence. Final appraisal determination. Alendronate, etidronate, risedronate, raloxifene and strontium ranelate for the primary prevention of osteoporotic fragility fractures in postmenopausal women. London, NICE. June 2007b.

National Institute for Health and Clinical Excellence. Final appraisal determination. Alendronate, etidronate, risedronate, raloxifene, strontium ranelate and teriparatide for the secondary prevention of osteoporotic fragility fractures in postmenopausal women. London, NICE. June 2008a.

National Institute for Health and Clinical Excellence. Final appraisal determination. Alendronate, etidronate, risedronate, raloxifene and strontium ranelate for the primary prevention of osteoporotic fragility fractures in postmenopausal women. London, NICE. June 2008b.

National Osteoporosis Foundation (NOF). Physician's Guide to prevention and treatment of osteoporosis. National Osteoporosis Foundation, 2003 Washington, DC

National Osteoporosis Guideline Group on behalf of the Bone Research Society, British Geriatrics Society, British Society of Rheumatology, Society of Endocrinology, British Orthopaedic Association, Primary Care Rheumatology Society, Osteoporosis 2000 and Osteoporosis Dorset. Osteoporosis: Clinical guideline for prevention and treatment. University of Sheffield Press, Sheffield, UK 2008

Oleksik, A., Lips P, Dawson A, Marshall ME, Shaw W, Cooper C et al. Health-related quality of life in postmenopausal women with low BMD with or without prevalent vertebral fractures. J Bone Miner Res. 2000;15: 1384-92.

Parker MJ and Anand JK. What is the true mortality of hip fractures? Public Health, 1991; 105: 443-6.

Praities N. Government backs rival to NICE guidance on fracture risk. Pulse. Saturday, 15 August 2009

Reginster JY, Minne HW, Sorensen OH et al (2000) Randomized trial of the effects of risedronate on vertebral fractures in women with established postmenopausal osteoporosis. Vertebral efficacy with risedronate therapy (VERT) study group. Osteoporos Int 11: 83-91

Roussow JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML et al (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. JAMA 288:321-333

Roux C, Reginster J-Y, Fechtenbaum J, Kolta S, Sawicki A, Tulassay Z et al. Vertebral fracture with reduction with strontium ranelate in women with postmenopausal osteoporosis is independent of baseline risk factors. J Bone Miner Res. 2006; 21: 536-42.

Royal College of Physicians and Bone and Tooth Society of Great Britain. Update on pharmacological interventions and an algorithm for management 2000. Royal College of Physicians, London UK.

Royal College of Physicians. Glucocorticoid-induced osteoporosis. Guidelines on prevention and treatment. Bone and Tooth Society of Great Britain, National Osteoporosis Society and Royal College of Physicians. 2002 Royal College of Physicians, London UK

Royal College of Physicians. Osteoporosis: clinical guidelines for the prevention and treatment. 1999. Royal College of Physicians, London

Saag KG, Emkey R, Schnitzer TJ, Brown JP, Hawkins F, Goemaere S et al. Alendronate for the prevention and treatment of glucocorticoid-induced osteoporosis. Glucocorticoid-induced Osteoporosis Intervention Study Group. New Engl J Med. 1998; 339: 292-99.

Siminoski K, Leslie WD, Frame H, Hodsman A, Josse RG, Khan A, Lentle BC et al. Recommendations for bone mineral density reporting in Canada: a shift to absolute fracture risk assessment. J Clin Densitom. 2007; 10:120-3.

Siris E and Delmas PD. Assessment of 10-year absolute risk: a new paradigm with worldwide application. Osteoporos Int 2008; 19: 383-384.

Stevenson M, Davis S, Lloyd-Jones M, Beverley C. The clinical effectiveness and cost-effectiveness of strontium ranelate for the prevention of osteoporotic fragility fractures in postmenopausal women. Health Technol Assess. 2007; 11: 1-134.

Stevenson M, Lloyd Jones M, De Nigris E, Brewer N, Davis S, Oakley J. A systematic review and economic evaluation of alendronate, etidronate, risedronate, raloxifene and teriparatide for the prevention and treatment of postmenopausal osteoporosis. Health Technol Assess. 2005a; 9: 1-160.

Stevenson MD, Oakley J, Chilcott JB (2004).Gaussian process modeling in conjunction with individual patient simulation modeling: a case study describing the calculation of cost-effectiveness ratios for the treatment of established osteoporosis. Med Decis Making; 24: 89-100.

Stevenson, M, Davis, SE, and Kanis, J., The hospitalization costs and outpatient costs of fragility fractures. Women's Health Medicine, 2006: 4: 149-151.

Stevenson MD, Brazier JE, Calvert NW, Lloyd-Jones M, Oakley J, Kanis JA (2005b) Description of an individual patient methodology for calculating the cost-effectiveness of treatments for osteoporosis. Journal of the Operational Research Society; 56: 214-221.

Stevenson M and Wailoo A (2009) Alendronate, etidronate, risedronate, raloxifene and strontium ranelate for the primary prevention of osteoporotic fragility fractures in postmenopausal women. Alendronate, etidronate, risedronate, raloxifene, strontium ranelate and teriparatide for the secondary prevention of osteoporotic fragility fractures in postmenopausal women. A review of comments submitted by consultees on the economic model. Report by the Decision Support Unit. School of Health and Related Research, University of Sheffield

Stevenson M, Davis S, Lloyd Jones M & Beverley C. The clinical effectiveness and cost effectiveness of strontium ranelate for the prevention of osteoporotic fragility fractures in postmenopausal women (Strontium ranelate assessment report). 2005c

Ström O, Borgström F, Kanis JA, Jönsson B (2009) Incorporating adherence in health economic modelling of osteoporosis. Osteoporosis International. 20: 23-34 with erratum page 35

Tosteson AN, Melton LJ 3rd, Dawson-Hughes B, Baim S, Favus MJ, Khosla S, Lindsay RL; National Osteoporosis Foundation Guide Committee (2008) Cost-effective osteoporosis treatment thresholds: the United States perspective. Osteoporos Int. 19: 437-47.

World Health Organization. Assessment of osteoporosis at the primary health care level. WHO, Geneva, 2007. (www.who.int/chp/topics/rheumatic/en/index.html)

Zethraeus N, Borgstrom F, Strom O, Kanis JA, Jonsson B. Cost-effectiveness of the treatment and prevention of osteoporosis - a review of the literature and a reference model. Osteoporos Int. 2007; 18: 9-23.

Zethraeus, N., Johannesson, M., and Jonsson, B., A computer model to analyze the cost-effectiveness of hormone replacement therapy. Int J Technol Assess Health Care, 1999; 15: 352-65.

Zethraeus N, Ström O, Borgström F. What is the risk of institutionalization after hip fracture? Osteoporos Int . 2006; 17 (Suppl 2): 60.