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

Matt Stevenson, Allan Wailoo
School of Health and Related Research, University of Sheffield

[Where this report contains confidential information, the respective text is blacked out]

14 August 2009

Contents

ABBREVIATIONS AND DEFINITIONS

AC	Appraisal Committee
ACD	Appraisal consultation document
AR	Assessment report
BMD	Bone mineral density
BMI	Body mass index
BPs	Bisphosphonates
Consultee	Organisations that accept an invitation to participate in the appraisal
CRF	Clinical risk factor
DSU	Decision Support Unit
FAD	Final appraisal determination
FRAX	WHO fracture risk assessment tool
GP	General practitioner
HTA	Health technology assessment
ICER	Incremental cost effectiveness ratio
NICE	National Institute for Health and Clinical Excellence
Osteoporosis	Bone mineral density is 2.5 standard deviations or more below the young adult mean (T-score <-2.5)
PSA	Probabilistic sensitivity analysis
QALY	Quality adjusted life-year
ScHARR	School of Health and Related Research
SE	Standard error
SD	Standard deviation
SR	Strontium ranelate
T-Score	The number of standard deviations from the average bone mineral density of healthy young women
TA	Technology appraisal
WHO	World Health Organization

1. INTRODUCTION

1.1. BACKGROUND

Technology Appraisals (TA) 160 and 161 were published in October 2008. These appraisals were initiated in August 2002 with a single scope for both primary and secondary prevention of osteoporotic fragility fractures in postmenopausal women, and included bisphosphonates (licensed at that time), raloxifene and teriparatide. In February 2004, the appraisal was split into two separate appraisals for primary and secondary prevention (the latter was issued as TA87). For primary prevention, additional work was carried out to incorporate the identification costs of women at high risk of fracture; to include epidemiological data pertaining to the unpublished WHO algorithm; and to explore various scenarios/assumptions resulting from consultation on the first Appraisal Consultation Document and discussions with the Guideline Development Group.

These additional analyses extended the timelines of the appraisal for primary prevention to coincide with the new appraisal of strontium ranelate (for both primary and secondary prevention, referred to NICE in June 2004). By this time there was also the need to consider an update of TA87. In June 2005, the Institute decided to align its work on osteoporosis and develop one technology appraisal on primary prevention and one on secondary prevention (the update on TA87), including all respective drugs referred at that stage.

ACDs for both appraisals were first issued in September 2005. Following consultation, further sensitivity analyses were carried out; new analyses were required following price reductions for generic alendronate; and discussions were held with the NICE Osteoporosis Guideline Development Group to facilitate an alignment of approaches. Early in 2007, FADs were issued that contained recommendations on the initiation of treatment with alendronate only. Following an appeal, the appraisals were referred back to the Committee which was asked by the NICE Appeal Panel to also include guidance for women who cannot take alendronate. FADs were issued early in 2008. These FADs recommended alendronate as first line treatment for women at high risk of fracture as defined by age, bone mineral density and other risk factors. The other drugs, being more expensive, were recommended for women who could not

take alendronate at the ages and bone mineral density values at which those drugs become cost effective. An appeal was received from the manufacturer of strontium ranelate, Servier, and was heard in September 2008. The Appeal Panel dismissed the appeal on all points and the two pieces of guidance were published the following month.

One important element of the economic modelling for these appraisals was that the model included data which were designated academic-in-confidence. Because of this the economic model for these appraisals was never released to consultees and commentators during the course of this appraisal.

Servier applied for a Judicial Review and a court hearing was held in January 2009. One of the arguments made by Servier was that NICE's appraisal had been unfair because consultees and commentators (including Servier) could not access the economic model, because it contained confidential information. The High Court requested NICE to continue negotiating permission to release the confidential information. Following the court ruling, NICE was able to reach agreement on the release of the economic model for consultation with the owner of the confidential information, and the economic model for the appraisals was offered to all consultees and commentators for an 8 week period of consultation.

Eight consultees and commentators requested the model and provided the necessary confidentiality undertakings, that is, in addition to the standard agreement associated with the release of an executable model with NICE, and a confidentiality undertaking with the owner of the confidential information. These consultees and commentators received a CD ROM containing the confidential executable version of the economic model, a Word document with instructions for running the model and a pro-forma to document comments on the model. Four consultees and commentators provided responses on the pro-forma.

This review

The Decision Support Unit (DSU) was requested by NICE to review responses from the consultees on the model and report on its findings to the Appraisal Committee. Only comments on the economic model were considered. 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 were not reviewed.

The DSU was asked to provide an expert view on whether:

- The comments received on the executable model provide a justifiable case to question the reliability of the model used for the formulation of the guidance.
- Individual comments on the model have been made with enough justification, supporting information, and details of the implementation in the model for them to be replicated.
- The individual comments on the model on the cost effectiveness results have been established correctly.

The role of the DSU was not to provide an opinion on inputs or assumptions that had been previously agreed by the Committee and used in the model. The DSU's consideration of comments provided by consultees is limited to those which relate to the analyses agreed by the Committee and used for the formulation of the guidance. It does not include consideration of comments relating to analyses based on other inputs or assumptions which were discussed previously by the Committee during the appraisal process but not agreed as inputs for the model. Similarly, the DSU has not considered comments on analyses conducted with new data which has become available subsequent to the Committee's deliberations.

2. METHODS

All comments submitted by the consultee organisations have been considered as to whether they meet the criteria for consideration by the DSU as outlined in the project specification; that is, whether they relate to the model used in the appraisals published as TA160 and TA161 (the executable model being consistent with the description of it provided to the Appraisal Committee) AND relating to the analyses used for the formulation of the guidance.

Pro-forma comments considered to potentially meet these criteria have been assessed and a response detailed in the table in Annex 1 with further explorations in Annexes 2-5. Comments not meeting the criteria are considered outside of the scope for this report, as indicated in the table in Annex 1, and are not considered further in this document.

The DSU has also reviewed whether individual comments on the model have been made with sufficient supporting information, and whether the impact of the individual comments on the cost effectiveness results from the economic model have been established correctly.

Responses to the individual comments are provided in the table in Annex 1. In order to provide a full response to some consultee comments, the DSU carried out some exploratory sensitivity analysis. These analyses are described in Annexes 2-4. These sensitivity analyses were conducted using the executable model provided to consultees, except for the exploration of the estimation of fracture risks associated with different values for Body Mass Index (BMI), which was undertaken using the WHO algorithm provided to the Assessment Group by the owner of the WHO algorithm. An exploration of the correlation between T-scores and clinical risk factors was established with the raw data underpinning the WHO algorithm provided to the Assessment Group. The latter two explorations cannot be shown in detail because only the model was released from the confidentiality agreement with the owner of the data.

3. COMMENTS RECEIVED AND SUMMARY OF RESPONSES FROM THE DSU

Six responses were received from consultees and commentators following completion of the consultation period on the model. Two of these consultee organisations did not provide comments on the model. The pro forma comments from 4 consultees and commentators, Servier, the Bone Research Society (BRS), the National Osteoporosis Society (NOS) and the Society for Endocrinology (SocEnd) have been collated and are shown in Annex 1, along with specific and detailed responses from the DSU addressing key issues raised following consultation on the model.

A summary of the key themes raised by the consultees and commentators are listed below. The organisations raising the issue are included in parenthesis for ease of reference.

3.1. COMMENTS CONSIDERED AS RELEVANT TO THE MODEL

Inability to assess validity of model, leading to claims that the model is not fully executable (Servier Issue 1)

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.

Inadequate documentation of the model (Servier Issue 1, BRS Issue 6, NOS Issue 1, Soc End Issue 1)

In addition to the documentation in the Assessment Reports [1-5] and associated correspondence, the model has been detailed in an HTA monograph [6] and was peer reviewed both by the reviewers for the HTA and by reviewers for the Journal of Operational Research Society [7].

Further correspondence has occurred in *Bone* [8] indicating that the results of the model are relatively similar to those of a contemporary model [9] if the same assumptions or inputs are used. Further evidence submitted to *Osteoporosis International*, but not published, presenting the data to support this finding, is provided in Annex 5.

Instructions were also supplied to consultees and commentators along with the executable version of the model.

At the request of one of the consultees, NICE gave consultees and commentators the opportunity to receive a copy of the executable version of a previous model. This previous model had been used in the osteoporosis appraisal up to April 2005 and

released to consultees. It had been used to generate data inputs to construct the current model used in the appraisal. The previous model was not necessary to run the executable version of the current model which was used to develop TA160/161. NICE was of the view that as a read-only version of the previous model had been made available to organisations that were consultees in April 2005, it was right to offer access to an executable version of this previous model to all consultees of TA160/161. However, NICE stressed that it was the current model that was the subject of this consultation exercise.

Generally, whilst NICE can release the model which it uses in an appraisal, it cannot engage with consultees who are for whatever reason unable to fully understand the functionality of the model over and above the extent it has done so in this case. The obligation is to provide access to material actually held, not to generate additional material.

Appropriateness of population data used (BRS Issue 2, NOS Issue 2)

The Holt *et al* data [10] were used in this appraisal, because consultees highlighted in the first consultation on this appraisal during December 2003 that these were the most appropriate data to use in a UK context. This is documented on p16-17 of the 2005 Strontium Ranelate Assessment Report. [1] The DSU review confirms that the data in the model is consistent with the data underpinning the Holt *et al* paper [10].

This point also focuses on the distribution of BMD T-scores in the model compared with data underpinning the publication by Holt *et al* [10]. The DSU have reviewed the T-score values used in the model with those in the database underpinning the Holt *et al* paper [10] provided to the Assessment Group by the data owners in 2004 (Annex 4).

The DSU undertook exploratory analyses using the data underpinning the Holt *et al* paper [10] to test the assumption of normality assumed for T-Score distribution. These analyses show that any assumptions about normality and a standard deviation of 1, as used in the model, are likely to be favourable to the treatments appraised (Annex 4).

Inability to change certain variables in the model (Servier Issue 1, NOS Issue 6)
There are several parameters that are fixed in the model. Several of these parameters are discussed in other sections of this report, for example, mortality associated with clinical risk factors.

These parameters are fixed because they are based on standard tables or other data that is not subject to change in the context of this appraisal; for example, population mortality data are based on standard life tables as documented.

The WHO algorithm used to generate estimates of fracture risk was not embedded in the model. Fracture risks were calculated using regression analysis based on the WHO coefficients and algorithm and these fracture risks were imported into the model. Parameters relating to the calculation of the fracture risks from the academic in confidence WHO algorithm are not part of the cost effectiveness model and therefore cannot be altered.

Disagreement with certain modelling approaches - fixed BMI (Servier Issue 3, BRS Issue 4, NOS Issue 6)
The DSU confirms that a BMI of 26 kg/m^2 was used for all women in the model, which was the mean value from the data underpinning the Holt et al paper [10]. Consultees claim that using a fixed BMI of 26kg/m^2 is unfavourable to interventions and also quote a paper indicating that a lower BMI should be associated with a greater risk of fracture [13]. The DSU has explored the impact of BMI on the fracture risk estimated by the WHO algorithm to test this consultees' claim.

This exploration was done with the WHO algorithm. Fracture risks were calculated using BMI values of 20 kg/m^2 and 32 kg/m^2. Because the WHO algorithm itself is confidential the detailed analyses and results are confidential (Annex 2). However, these exploratory analyses show that, once BMD is known, BMI values of 20 kg/m^2 and 32 kg/m^2 do not lead to higher fracture risks than a BMI of 26 kg/m^2 and that the relationship in the publication quoted [13] is not exhibited in the WHO algorithm provided to the Assessment Group.

Disagreement with certain modelling approaches - weight applied to risk factors (Servier Issue 5)

The model uses the risk factors with the 'weight' (or relative risk) as provided in the WHO algorithm, that is, the model itself can and has been used to assess cost effectiveness for specific combinations of risk factors according to those relative risks (see Strontium Ranelate Assessment Report; reference 1). Because of the complexities of giving recommendations for different combinations of risk factors, the Appraisal Committee, supported by the opinion of the Clinical Guideline Development Group, requested that in each age and T-Score group the results be categorised by the number of clinical risks factors that a woman has. To estimate a likely ICER for women with a specific number of clinical risk factors, all eligible permutations of risk factors were calculated with the model, and then the median ICER was taken to assume it represented the entire population for that T-score and age group.

Therefore, treatment of risk factors in the recommendations is a practical response to the complexity of the combinations of patient characteristics. This approach was agreed by the Appraisal Committee and is not an issue about the validity of the model.

Disagreement with certain modelling approaches – rationale for choice of clinical risk factors considered (Servier Issue 4, BRS Issues 7 and 8, NOS Issues 4 and 5)

The WHO algorithm was used as supplied to the Assessment Group to formulate the fracture risks used as inputs to the cost effectiveness model. The subsequent use of individual risk factors in the recommendations was a matter of Committee decision-making. Therefore, it is not related either to the use of the WHO algorithm itself or the cost effectiveness modelling. Following the decision of the Committee to exclude certain risk factors, these specific factors were not disaggregated from the modelling, as this would require major changes to the macros within the model used to aggregate the cost effectiveness data. The DSU has not explored this further.

The DSU expects, however, that based on the WHO coefficients, the inclusion of glucocorticoids, for example, as a risk factor in the model would favour the interventions. This is because the ICERs for women taking glucocorticoids were calculated in the model and were part of calculating the median ICER per age and T-

score group. Because the ICERs for women taking glucocorticoids were relatively lower than the average, this would favour the interventions appraised.

Omission of certain clinical risk factor interactions (Servier Issue 2, BRS Issues 5 and 9)

The model used the WHO algorithms supplied, at the time of the appraisal, as academic-in-confidence (Strontium Ranelate Assessment Report 2005 page 20-23; reference 1). The DSU can confirm that the interactions referred to -prior fracture:age and BMD:age have been appropriately incorporated as inputs into the model.

The fracture risks used within the model have been directly calculated from the WHO algorithm using all interactions. Thus, the interactions included in the algorithm were incorporated into the fracture risks calculated. These risks were calculated outside of the model using regression analyses with the coefficients from the WHO algorithm. The resulting fracture risk values were subsequently imported into the model.

As only the model was covered in the confidentiality agreement with the owner of the data, the regression formula and coefficients calculating the risks could not be released to consultees.

Disagreement with the annual risk associated with clinical risk factors (Servier Issues 2 and 7)

The DSU has reviewed the calculation of fracture risk based on the WHO algorithm supplied to the Assessment Group. The calculations are performed correctly to calculate fracture risk for each T-score.

The translation of these point estimates for use across T-Score bands (as used in the cost effectiveness model) was based on the mean value between the band values e.g. a T-score of -3.75SD would be assumed for those women within the T-Score band of -3.5SD to -4.0SD. The DSU note that in osteoporotic women such an assumption would be favourable to interventions as there would be more women with a T-Score of -3.5SD than women with a T-Score of -4.0SD.

The DSU also note that the values used for each band were 0.01SD lower than the halfway point between the T-Score band. For example, for the T-score band -3.5SD to

-4.0SD, -3.74SD was entered into the WHO algorithm instead of -3.75SD. However, given that the use of the intended midpoint value, of for example -3.25, as described above is favourable to the interventions, the DSU believes that the use of -3.24 would not change this conclusion.

Inability to directly assess integrity of the application of the WHO algorithm in the model (Servier Issue 2, NOS Issue 6)

The WHO algorithm used to generate estimates of fracture risk was not embedded in the model. Therefore, consultees are unable to directly assess the application of the algorithm by looking at the model. While the model contains inputs that concern fracture risk it calculates costs and effects for patient subgroups, but not fracture risk. The WHO algorithm itself is the work of a third party and so individual elements of it are not meaningfully open to comment. Whilst the DSU is satisfied that, apart from the small error described above related to entering the T-score into the WHO algorithm, the WHO algorithm has been correctly applied. The DSU notes that consultees have asked to verify this for themselves. The DSU understands that NICE intends to approach the owner of the WHO algorithm for permission to release it to consultees who have commented on the model.

Omission of mortality risks associated with clinical risk factors (Servier Issue 2)

The DSU has established that increases in mortality associated with clinical risk factors were not taken account of in the model. The DSU has not explored this quantitatively. However, 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.

Uncertainty around methodology used to extend 10-year time horizon (Servier Issue 6, NOS Issue 6)

The DSU confirms that an explanation of the derivation of the additional QALYs that may be gained by treatment after the 10-year time horizon is contained in the Strontium Ranelate Assessment Report [1] and the HTA report [6]. The explanation is relatively limited in its description.

The DSU has also established that following an update based on more recent evidence, this methodology had been amended for use in the modelling from 2006 onwards and incorporated into the appraisal from thereafter. However, this had not been captured clearly in the Assessment Reports from 2006 onwards. The DSU provide a clarification of the 'bolt-on' methodology in the first part of Annex 3. The DSU confirm that the updates to the mortality modelling described here were included in the model from 2006 onwards, that is, they are included in the model used for the development of TA160/161.

Using sensitivity analyses on the executable model provided to consultees, the DSU demonstrate that the ICERs are not sensitive to the benefits accrued beyond the 10 year time horizon and to the mortality assumptions associated with vertebral and proximal humerus fractures (Annex 3). A doubling of the base case estimation of the additional QALYs associated with extension in time horizon from 10 years to patient lifetime does not have any meaningful impact on the output of the modelling.

Amalgamation of clinical risk factors leads to inaccuracy in the estimates of cost effectiveness (Servier Issue 13, BRS Issue 3)

The DSU confirm that the guidance issued by NICE groups patients according to the number of clinical risk factors and by prior fracture. Consultees highlighted that the absolute fracture risk can vary substantially according to the specific risk factor, and claim that therefore cost effectiveness estimates for some patients are unfavourable to interventions when using the cost-effectiveness model. Consultees claim that FRAX provides the mechanism to compute the cost-effectiveness according to the specific risk factor. The DSU understands FRAX to be a tool to calculate fracture risk, not cost effectiveness. The DSU also caution against the use of absolute fracture risk as a measure to determine cost-effectiveness as each absolute risk could have different cost per QALY values dependent on the proportion of risk related to the individual risk factor that was associated with hip fractures. Table 33 of the HTA report [6] demonstrated that there is not a direct correlation between absolute risk and cost-effectiveness ratios. Thus, basing treatment decisions on absolute risk alone has limitations for the reasons outlined in the HTA report.

The DSU has established that the model has the capability to produce results for all combinations of specific clinical risk factors; indeed such results have been previously estimated [1].

The decision to present cost effectiveness results by groups defined by the number of clinical risk factors (amongst other factors) was made by the Appraisal Committee for practical reasons. The point raised is therefore not relevant to the operation of the model.

The consultees' claim that using the median ICERs underestimates the cost-effectiveness is incorrect because all subgroups will contain patients for whom an intervention is more cost effective than the average, as well as patients for whom the intervention is less cost effective.

3.2. COMMENTS RECEIVED BUT NOT CONSIDERED NOT RELEVANT TO THE EXECUTABLE MODEL CONCERNED

Discount rate (Servier Issue 8)

The model applies a discount rate of 6% to costs and 1.5% to benefits, as was appropriate to the NICE process governing this appraisal.

The DSU confirm that the discount rates used were described in the Strontium Ranelate Assessment Report and no comments were received from consultees on the discount rate during the appraisal.

The DSU comments that it is expected that were the current rates (3.5% for both benefits and costs) used then the cost effectiveness of all interventions would be reduced (potentially markedly) as a successful intervention would be associated with acquisition costs in the initial treatment period (which would increase using alternative discount rates) and a long-term utility benefit (which would decrease using alternative discount rates).

Vertebral fracture utility value (Servier Issue 12)

As this is a comment on an input parameter, previously known to consultees and commentators, and agreed upon by the Committee, and therefore one on which consultees have previously been able to make representations, no DSU response is provided.

Side effects disutility sensitivity analysis (Servier Issue 10, BRS Issue 1, NOS Issue 3)

As this is a comment on an input parameter, previously known to consultees and commentators, and agreed upon by the Committee, and therefore one on which consultees have previously been able to make representations, no DSU response is provided.

Side effects – the same disutility for Strontium Ranelate and bisphosphonates (Servier Issue 10)

As this is a comment on an input parameter, no detailed DSU response is provided. However, the DSU confirms that the ScHARR review of adverse events [12] was based on data for bisphosphonates. No formal review for strontium ranelate was available to the Committee. Strontium ranelate is not associated with the same adverse gastrointestinal effects as bisphosphonates, and the clinical data for strontium ranelate indicate other adverse effects related to thromboembolism. The latter was not formally included in the model, and the Committee considered it appropriate to set the base case assumptions for strontium ranelate to the same as for bisphosphonates. The DSU additionally note that the sensitivity analysis using a higher disutility of side effects was only applied to alendronate, not to strontium ranelate.

Costs associated with fracture (Servier Issue 11)

As this is a comment on an input parameter, previously known to consultees and commentators, and agreed upon by the Committee, and therefore one on which consultees have previously been able to make representations, no DSU response is provided.

Identification of women at high risk (Servier Issue 13)

The DSU confirm that the identification of women at high risk is fully documented in the Strontium Ranelate assessment report (p100) which was available to consultees.

[1] The way in which the identification of women at high risk was carried out in the model was understood and agreed by the Appraisal Committee. As such no further comment is made by the DSU.

Compliance (Servier Issue 9)

The assumptions regarding compliance are fully documented in the Strontium Ranelate Assessment Report (p58) which was available to consultees.[1] The assumptions about compliance were understood and agreed by the appraisal committee and thus no comment from the DSU is provided.

4. CONCLUSIONS

The DSU has considered in detail the comments received on the executable model from Servier Laboratories, the Bone Research Society, the National Osteoporosis Society and the Society for Endocrinology.

We consider that a large number of comments did not relate to the executable model but to the appropriateness of specific parameter values that were previously set out in the relevant documentation and the Appraisal Committee were aware of for these appraisals. Consultees have previously been aware of these parameter values and have been able to comment on them in previous consultations. The Appraisal Committee has previously consulted on, considered and decided on the appropriate parameter values to adopt. Therefore, these do not relate to the validity of executable model, these comments have not been investigated in detail by the DSU.

The DSU agrees that some parameters in the executable model are fixed, including those with small uncertainty and that are usually considered fixed such as standard mortality rates. In particular, the WHO algorithm used to generate estimates of fracture risk was not embedded in the model. These fracture risks are inputs to the cost effectiveness model and do not form part of the cost effectiveness model itself. Comparisons with FRAX were made by several consultees, on the basis that the WHO algorithm supplied to the Assessment Group and FRAX are identical. It is not possible to verify any such analyses without access to the FRAX algorithm, which the DSU does not have. The DSU can confirm that the estimation of fracture risk used as

an input to the cost effectiveness model as consistent with the WHO algorithm supplied to the Assessment Group.

Our review and sensitivity analyses suggest that none of the consultee comments relating to the modelling approach would lead to significant improvements in the cost effectiveness of the interventions, either cumulatively or in isolation. Indeed, in several instances the modelling approach that has been adopted in the appraisal appears favourable to the technologies i.e. suggesting that the ICERs generated by the model provide underestimates of the ICERs.

The DSU concludes 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.

5. REFERENCES

Please note, references 7 onwards support information contained in the Annexes.

1. 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). 2005

2. Decision Support Unit economic evaluation and systematic review. Available at http://www.nice.org.uk/guidance/index.jsp?action=folder&o=38448

3. Additional analyses requested by the Committee. Available at http://www.nice.org.uk/guidance/index.jsp?action=folder&o=38451

4. Evaluation Report, February 2007. Available at http://www.nice.org.uk/guidance/index.jsp?action=folder&o=37690

5. Analyses of cost-effective BMD scanning and treatment strategies for generic alendronate, and the cost-effectiveness of risedronate and strontium ranelate in those people who would be treated with generic alendronate, April 2008. Available at http://www.nice.org.uk/guidance/index.jsp?action=download&o=40243

6. Stevenson M, Davis S, Lloyd Jones M et al. The clinical and cost-effectiveness of strontium ranelate for the prevention of osteoporotic fragility fractures in post-menopausal women. Health Technol Assess 2007; 11 (4) pp1-134

7. Stevenson MD, Brazier JE, Calvert NW, Lloyd-Jones M, Oakley J, Kanis JA. Description of an individual patient methodology for calculating the cost-effectiveness of treatments for osteoporosis in women. Journal of Operational Research Society. 2005; 56 (2): 214-221

8. Stevenson M. The population of health economic models is critical. Bone 2008; 43 (1): 214

9. Kanis JA, Adams J, Borgstrom F et al. The cost-effectiveness of alendronate in the management of osteoporosis. Bone 2007 42 (1): 4-15

10. Holt G, Khaw KT, Reid DM 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 2002; 75:736–42.

11. Zethraeus N, Borgstrom F, Strom O et al. Cost effectiveness of the treatment and prevention of osteoporosis—a review of the literature and a reference model. Osteoporos Int 2007;18:9–23

12. Lloyd Jones M, Wilkinson A. Adverse events and persistence with therapy in patients taking oral alendronate, etidronate or risedronate: systematic reviews. 2006. Available at: http://www.nice.org.uk/guidance/index.jsp?action=download&o=38450

13. De Laet C, Kanis JA, Oden A et al. Body Mass as a predictor of fracture risk: a meta-analysis. Osteoporos Int 2005; 16:1330-8

14. Government Actuary Department. Expectation of life: United Kingdom females. London: GAD; 1999.

15. Browner WS, Seeley DG, Vogt TM & Cummings SR. Study of Osteoporotic Fractures Research Group. Non-trauma mortality in elderly women with low bone mineral density. Lancet 1991; 338:355–8.

16. Kind P, Dolan P, Gudex C et al. Variations in population health status: results from a United Kingdom national questionnaire survey. BMJ 1998; 316: 736-741.

17. Todd CJ, Freeman C, Camilleri-Ferrante C et al.. Anglian audit of hip fracture 2. Cambridge: Cambridge Health Services Research Group, University of Cambridge; 1999

18. Parker MJ & Anand JK. What is the true mortality of hip fractures? Public Health 1991; 105: 443-446.

19. Jalava T, Sarna S, Pylkkanen L et al. Association between vertebral fracture and increased mortality in osteoporotic patients. J Bone Miner Res 2003; 18:1254–60

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

21. Kanis JA, Compston JE. NICE continues to muddy the waters of osteoporosis. Osteoporos Int 2008; 19(8):1105-7.

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

Annex 1: DSU responses to comments received from consultees and commentators on the osteoporosis model

Comments received from Servier

Pro-forma field	Comment	DSU response
Issue	Issue 1 Transparency and validation	
Description of problem	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 The validity of the model cannot be assessed from the data supplied, nor is there any previous publication available to demonstrate its validity. It is not possible to test the manner by which mortality, fracture risks are accommodated in the model supplied. 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 model provided to consultees and commentators was the model used for the appraisals. No other model functionality was used for the appraisals. Instructions were included along with the executable version of the model. The DSU believe 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 [11].The DSU note that the results produced by the 'NICE' model and the reference model are similar when populated with similar input parameters (Annex 5) Alterations of the discount rates were not necessary as these values were set by NICE process. These appraisals were started before

Pro-forma field	Comment	DSU response
		2004 and therefore the 1.5/6% discount rates applied. Response to the set BMI value is provided in Annex 2. It is seen that setting this value at 26 kg/m2 for all women appears favourable to treatment. Population mortality data were taken from standard life tables (see Strontium Ranelate Assessment Report 2005, page 28/29). The DSU does not see a reason why these need to be changeable variables in a model. The DSU has established that mortality associated with clinical risk factors was not considered in the model. As the model was calculated based on results from the individual patient model calculated prior to the WHO algorithm being available, such mortality hazards could not be easily incorporated within the analyses and thus were not included. There has been no formal quantitative investigation of the impact of the impact of not including mortality hazards associated with clinical risk factors. However, it is hypothesised that the cost-effectiveness of treatments in women with clinical risk factors are likely to be favourable to treatment in the NICE model. This is because

Pro-forma field	Comment	DSU response
		preventing fractures and/or mortality in women with clinical risk factors would result in fewer QALYs being accrued as the woman would be expected to die earlier. The methodology concerning the incorporation of mortality after the 10 year time horizon is described in Appendix 9 of the Strontium Ranelate Assessment Report 2005, pages 137-139 [1]. The DSU has established that this methodology has been amended after 2005 for use in the modeling from 2006 onwards. This amended methodology used for estimating the change in results if the model was to be extended to a lifetime horizon rather than a 10-year period are contained in Annex 3. The DSU carried out sensitivity analyses in order to explore the effect of alternative mortality assumptions and the time horizon using the executable models provided to the consultees (Annex 3). These sensitivity analyses indicate that when the potential underestimations in QALYs were assumed to be double that used in the base case then this has only a minor effect on the results. The executable model contained the fracture risks associated with all combinations of

Pro-forma field	Comment	DSU response
		clinical risk factors, T-Score and age bands. In order to reduce the computational time of the models, the WHO algorithm was not embedded in the model. Instead, the fracture risks were computed using the regression formula from the WHO algorithm and the resulting risk values imported into the executable model. As only the executable models were allowed to be released to consultees following the agreement with Professor John Kanis the regression formula from the WHO algorithm for calculating fracture risks could not be released to C&Cs.
	The following variable cannot be changed for the sensitivity analysis: Baseline population risk of fracture.	The baseline risk of fracture was taken from the WHO algorithm and was assumed to be correct. No correlations were provided between the variables within the WHO algorithm, and thus only the midpoint estimates could be used without the risk of sampling algorithms that would not fit the underlying data.
	Nor was it possible to determine the accuracy with which the model reproduced the epidemiology of osteoporosis in the UK.	The epidemiology of osteoporosis within the model was driven by the WHO algorithm and the dataset from the Holt *et al* publication [10] provided to the assessment group; both of these have been taken on trust and assumed to be correct.

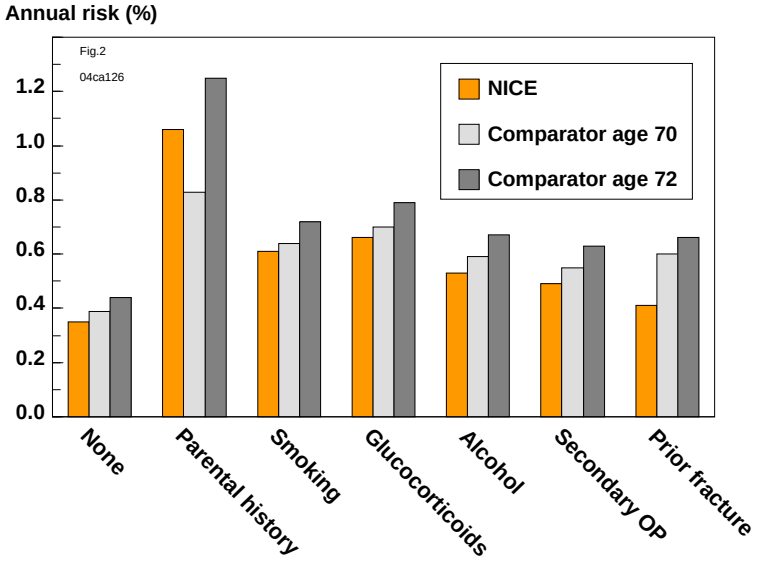
Pro-forma field	Comment	DSU response
Description of proposed amendment	Amend process to allow full re-assessment and comment on all model used as a part of the current appraisal	
Result of amended model or expected impact on the result	Provide an opportunity for an open and educated debate on the validity of the cost effectiveness model used as a basis for Appraisal Committee decisions.	

Pro-forma field	Comment	DSU response
Issue	Issue 2 Hip fracture estimates	
Description of problem	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. For the calculation of annual fracture risk it is not given whether this is applied to specific ages or to an age range. Irrespectively, there are discrepancies between the reviewers and NICE in the calculation of annual risks associated with clinical risk factors (CRFs). There are also discrepancies in the rank order of importance of the CRFs. Possible reasons for the discrepancies may relate to an erroneous assumption that none of the risk factors were associated with excess mortality. An alternative or additional explanation is that NICE derived the risks of clinical spine, forearm and humeral fractures incorrectly by subtracting the risk of hip fracture from the risk of a major fracture. 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. This part of the FRAX® has not been implemented in the NICE model. There are a number of significant interactions that are incorporated into FRAX® that appear to have been omitted from the NICE model. These include prior fracture·age and BMD·age, the omission of both will adversely affect cost-effectiveness at younger ages 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. In the NICE model the annual risks are entered directly as values in the	The DSU established that the model uses the annual probability of fractures provided by the WHO algorithm and an underlying risk of mortality. The DSU do not consider it necessary for 10-year fracture rates to be provided. The model is based on the evidence and data available at the time of guidance production, which was the WHO algorithm. The model used the WHO algorithms supplied as academic-in-confidence (Strontium Ranelate Assessment Report 2005 page 20-22, and Appendix 12) [1] These presented the risk of fracture at the hip and non-hip sites in the forthcoming year, i.e. on a yearly basis. The translation of these point estimates for use across T-Score bands (as used in the cost effectiveness model) was based on the mean value between the band values e.g. a T-score of -3.75SD would be assumed for those women within the T-Score band of -3.5SD to -4.0SD. The DSU note that in osteoporotic women such an assumption would be favourable to interventions as there would be more women with a T-Score of -3.5SD than women with a T-Score of -4.0SD.

Pro-forma field	Comment	DSU response
	excel sheets and it is not possible, therefore, to evaluate how the actual calculation of the risks were derived. Risks with different risk factors alone or in combination are given in Table 1 and Figure 1. All computations using FRAX® gave different values for annual risks compared to the estimates used in the NICE model. Moreover we could not reproduce the values derived by NICE from the methods described in the HTA report [p6, Stevenson et al, 2007b]. In the case of a major osteoporotic fracture (hip, clinical spine, forearm and humerus fracture), the NICE estimates were higher than those derived from FRAX®. An important exception was the risk estimate associated with a prior fracture where the risk estimate was lower with the NICE assumptions. The same findings were observed when comparing the annual risks in younger ages (Table 2). **Table 1** Annual risk of fracture (%) as given in the NICE model and computed from FRAX®. Risks are given for hip fracture and a major fracture (hip, clinical spine, forearm and humerus) NICE FRAX: major 1.66, hip 0.35 Review FRAX 70-year: Major 1.52, hip 0.39 Review FRAX 72-year: major 1.58, hip 0.44 CRFs: None NICE FRAX: major 2.82, hip 1.06 Review FRAX 70-year: Major 2.58, hip 0.83 Review FRAX 72-year: major 2.78, hip 1.25 CRFs: Parental history	The DSU also note that the values used for each band were 0.01SD lower than the halfway point between the T-Score band. For example, for the T-score band -3.5SD to -4.0SD, -3.74SD was used instead of -3.75SD. However, given that the use of the intended midpoint value, of for example -3.25, as described above is favourable to the interventions, the DSU believes that the use of -3.24 would not change this conclusion. The DSU has explored the methodology used for converting the results from the WHO algorithm into risk of fracture. Apart from the above error with entering the T-score into the algorithm, no evidence of errors was discovered. The DSU has established that mortality associated with clinical risk factors was not considered in the model. As the model was calculated based on results from the individual patient model calculated prior to the WHO algorithm being available, such mortality hazards could not be easily incorporated within the analyses and thus were not included. There has been no formal quantitative investigation of the impact of the impact of not including mortality hazards associated with clinical risk factors. However, it is hypothesised that the cost-effectiveness of treatments in women with clinical risk factors are likely to be

Pro-forma field	Comment							DSU response
	Smoking	1.86	0.61	1.68	0.64	1.75	0.72	favourable to treatment in the model. This is because preventing fractures and/or mortality in women with clinical risk factors would result in fewer QALYs being accrued as the patient would be expected to die earlier. The DSU confirm that the interactions referred to -prior fracture:age and BMD:age- have been incorporated within the NICE model. The fracture risks used within the model are contained in the age 50 through to age 75 worksheets of the CPQ Calc Est and CPQ Calc Prev sheets and have been directly calculated from the WHO algorithm using all interactions. These interactions were incorporated into the fracture risk calculated using the regression formula from the WHO algorithm and the resulting risk values imported into the executable models. As only the executable models were exempt of the AIC, the regression analysis calculating the risks could not be released to C&Cs. The DSU cannot respond to a general comment regarding errors, because of the lack of sufficient detail provided but have tried to address the specific issues raised. The DSU note that all parts of the executable model are accessible.
	Glucocorticoids	2.80	0.66	2.47	0.70	2.53	0.79	
	Alcohol	2.07	0.53	1.92	0.59	2.00	0.67	
	Secondary OP	2.19	0.49	2.01	0.55	2.08	0.63	
	Prior fracture	2.38	0.41	2.47	0.60	2.50	0.66	
	Parental history + smoking	3.54	1.86	2.95	1.37	3.38	2.03	
	Parental history + Glucocorticoids	4.89	2.02	4.17	1.49	4.49	2.21	
	Parental history + alcohol	3.69	1.63	3.29	1.26	3.66	1.90	

Pro-forma field	Comment	DSU response
	Parental history + secondary OP 3.78 1.52 3.41 1.17 3.71 1.77 **Table 2** Annual risk of fracture (%) as given in the NICE model and computed from FRAX® in women at the age of 50 years. Risks are given for hip fracture and a major fracture (hip, clinical spine, forearm and humerus) NICE FRAX: major 0.64, hip 0.18 Review FRAX 50-year: Major 0.61, hip 0.13 Parental history: 1.17, 0.19, 1.16, 0.14 Smoking: 0.76, 0.32, 0.68, 0.23 Glucocorticoids: 1.09, 0.35, 1.03, 0.24 Alcohol intake: 0.82, 0.28, 0.76, 0.20	The risk of fracture were calculated directly from the WHO algorithm and then imported into the appropriate executable model spreadsheets. The midpoint age was used, thus for the age band 50-54 years the age of the woman was set to 52.5 years. The use of the mid point age may partly explain the apparent discrepancies presented in Table 1. The model used an age of 72.5 years; with the exception of a prior fracture the value presented for a major osteoporotic fracture at 72 years could be consistent with those used in the model. However this would neither explain the difference in risks with a prior fracture nor the risks of hip fracture. The DSU have checked that the coefficients used within the NICE model are consistent with the data in the WHO algorithm provided to the assessment group. It is possible that since 2004 the WHO algorithm was adjusted following peer review or emergent data and that FRAX uses marginally different coefficients than the WHO algorithm provided to the assessment group. This cannot be confirmed without having access to the

Pro-forma field	Comment	DSU response
	Secondary osteoporosis 0.85 0.26 0.80 0.18 Prior fracture 1.06 0.27 1.17 0.28 **Figure 1** Annual risk of hip fracture (%) computed by NICE and our comparison at the age of 70 and 72 years  	algorithm used by FRAX to calculate the annual risk of fracture. It is expected that the results presented in Table 2 would be similar were an age of 52.5 years assumed. The fracture rates presented in Table 1 for Review FRAX are likely to be more favourable to interventions than the NICE model due to the assumed increase in hip fracture.

Pro-forma field	Comment	DSU response
Description of proposed amendment	Correctly utilise FRAX® by using the co-efficients to adjust the mortality for a specific patient group and include the interactions that have been omitted.	
Result of amended model or expected impact on the result	Will reduce ICER of SR	The DSU do not concur that the use of mortality associated with clinical risk factors would reduce the ICER of SR. The DSU do not believe that any interactions were omitted.

Pro-forma field	Comment	DSU response
Issue	Issue 3 Body Mass Index	
Description of problem	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 It is not known how the BMI value was set by NICE, nor could this be tested since BMI cannot be changed in the NICE model. 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 70 years or more). Though the impact is less, there are errors of accuracy incurred when BMD is added to the model. 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 an important effect of BMI on mortality. The phenomenon is illustrated in Table 3 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	In TA 160/161 BMD is measured in women who receive treatment except in exceptional circumstances ('those women aged 75 years or over if the responsible clinician considers a DXA scan to be clinically inappropriate or unfeasible'). The issue presented is that the model is not sensitive to BMI which is unfavourable to interventions in some women. The DSU has established that it is correct that BMI is set at 26kg/m^2. Furthermore, the DSU has explored the impact of BMI on the fracture risk estimated by the WHO algorithm to see if it is correct that using a fixed BMI of 26kg/m^2 is unfavourable to interventions. The DSU could not establish why the coefficient of fracture prediction associated with BMI within the WHO algorithm changed markedly when BMD was known as the correlation between the variables in the Holt et al dataset was small. (R^2 = 0.079) (Annex 2). Nor was there a strong correlation between BMI and age (R^2 = 0.004).

<table>
<tr><th>Pro-forma field</th><th>Comment</th><th>DSU response</th></tr>
<tr><td></td><td>22%. These important interactions do not appear to be accommodated in the NICE model.

Table 3 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/m2 in an individual of the same age.

<table><tr><th></th><th colspan="2">Age 50 years</th><th colspan="2">Age 70 years</th></tr><tr><th>BMI</th><th>Major</th><th>Hip</th><th>Major</th><th>Hip</th></tr><tr><td>15</td><td>1.4</td><td>1.2</td><td>0.78</td><td>0.88</td></tr><tr><td>20</td><td>1.2</td><td>1.1</td><td>0.92</td><td>0.94</td></tr><tr><td>25</td><td>-</td><td>-</td><td>-</td><td>-</td></tr></table></td><td>Due to the large number of potential permutations of T-Score bands and BMI bands the NICE model did not subdivide women into both T-Score and BMI categories; only T-Score was used to categorise women in addition to age and prior fracture status (Annex 2).

A BMI of 26 kg/m^2, was used for all women in the model, which was the mean value from the Holt et al database. Exploratory analyses of the risk of fracture using BMI values of 20 kg/m^2 and 32 kg/m^2 (which encompass over 85% of women in the Holt et al database) within the WHO algorithm were used to assess the change in fracture risk.

The exploratory analyses show that, once BMD is known, BMI values of 20 kg/m^2 and 32 kg/m^2 do not lead to higher fracture risks than a BMI of 26 kg/m^2.</td></tr>
</table>

Pro-forma field	Comment	DSU response
Description of proposed amendment	Utilise FRAX ® appropriately to estimate the risk associated with BMI ranges instead of a fixed value.	
Result of amended model or expected impact on the result	Underestimates cost effectiveness of SR in patients with lower BMI.	As shown by the data in Annex 2 this statement does not appear to be compatible with fracture risks taken from the WHO algorithm. Indeed the use of a BMI of 26kg/m^2 appears to be favourable to intervention.

<table>
<tr><th>Pro-forma field</th><th>Comment</th><th>DSU response</th></tr>
<tr><td>Issue</td><td>Issue 4 Intake of alcohol</td><td></td></tr>
<tr><td>Description of problem</td><td>
The risk associated with alcohol intake is incorrect for the exposure recommended by NICE and will adversely affect cost-effectiveness.

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, 2005]. 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, 2005].
<table>
<tr><td>Consumption</td><td colspan="2">Without BMD</td><td colspan="2">Adjusted for BMD</td></tr>
<tr><td>(units/day)</td><td>RR</td><td>95% CI</td><td>RR</td><td>95% CI</td></tr>
<tr><td>Osteoporotic fracture</td><td></td><td></td><td></td><td></td></tr>
<tr><td>>2</td><td>1.38</td><td>1.16-1.65</td><td>1.36</td><td>1.13-1.63</td></tr>
</table>
</td><td>
The DSU has established that the coefficients for alcohol used in the model was consistent with that in the WHO algorithm, i.e. for >2 units of alcohol intake per day.

The DSU has not quantitatively explored the effect of the Committee decision to used a threshold of more than 4 units of alcohol on the estimated fracture risks and subsequent ICERs.

The DSU has estimated that were midpoint values used then it is expected that the risks would increase in women who consumed >4 units of alcohol compared with >2 units of alcohol. The DSU note however that the confidence intervals around the risk ratios are wide and that no data on statistical significance has been provided by the consultee. As such a positive correlation between risk ratio and alcohol consumption may not exist. If this were true then the weighted midpoint may rise only marginally compared with the value of >2 units.
</td></tr>
</table>

Pro-forma field	Comment					DSU response
	>3	1.55	1.26-1.92	1.53	1.23-1.91	
	>4	1.70	1.30-2.22	1.64	1.24-1.27	
	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	
Description of proposed amendment	Correct the accounting for alcohol intake					
Result of amended model or expected impact on the result	Improve the cost effectiveness of treatments					

<table>
<tr><th>Pro-forma field</th><th>Comment</th><th>DSU response</th></tr>
<tr><td>Issue</td><td>Issue 5 Weighting of Risk Factors</td><td></td></tr>
<tr><td>Description of problem</td><td>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. 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).

Table 5 Ten-year probability of osteoporotic fractures (%) according to BMD T-score at the femoral neck in women aged 65 years from the UK. [Data from FRAX® web site]
<table>
<tr><th rowspan="2">Number of CRFs</th><th colspan="6">BMD T-score (femoral neck)</th></tr>
<tr><th>-4.0</th><th>-3.0</th><th>-2.0</th><th>-1.0</th><th>0</th><th>1.0</th></tr>
<tr><td>0</td><td>27</td><td>15</td><td>9.7</td><td>7.1</td><td>5.9</td><td>5.0</td></tr>
<tr><td>1</td><td>37 (33-41)</td><td>22 (18-26)</td><td>14 (10-18)</td><td>10 (7.1-14)</td><td>8.5 (5.7-12)</td><td>7.3 (4.8-10)</td></tr>
<tr><td>2</td><td>49 (42-58)</td><td>30 (23-40)</td><td>20 (13-29)</td><td>15 (8.6-23)</td><td>12 (6.8-19)</td><td>10 (5.6-17)</td></tr>
<tr><td>3</td><td>62 (53-72)</td><td>41 (30-55)</td><td>27 (17-42)</td><td>20 (11-34)</td><td>17 (8.7-29)</td><td>15 (7.2-26)</td></tr>
<tr><td>4</td><td>73 (63-81)</td><td>52 (42-65)</td><td>36 (26-51)</td><td>27 (18-41)</td><td>23 (14-36)</td><td>20 (11-32)</td></tr>
</table></td><td>The DSU understands FRAX to be a tool to calculate fracture risk, not cost effectiveness. The DSU has established that the model has the capability to produce results for all combinations of clinical risk factors; indeed such results have been previously estimated [6]. For example in Table 33 of this report, it is seen that the T-Score threshold for a 70 year old woman with a prior fracture and parental history of fracture to receive strontium ranelate was –2.4 SD, which fell to –3.1 SD for a 70 year old woman with a prior fracture who smoked. The absolute risk of fracture difference was less marked (4.66 and 4.49 respectively) indicating that there is not a direct correlation between absolute risk and cost-effectiveness ratios. Thus basing treatment decisions on absolute risk alone has limitations for the reasons outlined in the report.

In the Strontium Ranelate AR [1], the results from the model were presented in a less aggregated way, with different thresholds produced at each age for all combinations of risk factors and with the T-Score bands divided into steps of 0.1SD rather than 0.5SD; the change in T-Score banding was at the request of the</td></tr>
</table>

Pro-forma field	Comment	DSU response
	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/m2 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]	Appraisal Committee. Results presented in terms of specific CRF combinations were also presented to the NICE Guideline Development Group. Given that there was not a risk calculator available at this time there was a strong recommendation that a comprehensive list of CRF combinations would present GPs with a logistical problem and that the recommendations should be grouped by the number of risk factors possessed by a woman. For simplicity the median ICER was used; it is 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.

Number of CRFs	BMI (kg/m²)						
	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)

Pro-forma field	Comment	DSU response
Description of proposed amendment	Implement the FRAX algorithm accurately to allow a more accurate assessment of fracture risk and cost effectiveness that aids implementation and deals more fairly with inter-patient variation..	
Result of amended model or expected impact on the result	Underestimates cost effectiveness of SR for some patients.	The DSU agree with the fact that the ICERs calculated by the model will be unfavourable to intervention in women with risk factors conferring a higher than median risk. However, the DSU note that conversely the cost effectiveness will be overestimated in women with risk factors conferring a lower than median risk.

Pro-forma field	Comment	DSU response
Issue	Issue 6 Time horizon	
Description of problem	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. The estimation 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 there is no information on how they are derived. 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 publication of the 'bolt-on' states that this took account of deaths occurring after 10 years [Stevenson et al, 2005]. The 'bolt-on' does not appear to accommodate preventable deaths during the offset period or after 10 years. The publication describing the 'bolt-on' states that this took account of deaths, but none of the other consequences of fracture. The spread sheets provided by NICE suggest that this may be untrue 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. Some adjustment is made for forearm fractures, the nature of which is not explained. If these adjustments are related to preventable deaths this would assume that wrist, rib, scapular, clavicular and sternal fractures increase mortality, whereas the report indicates otherwise. 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. However, in the model there are two values called *wristbonusat2.5* and	The DSU have explored • if the 'bolt on ' methods were described • if other consequences of fracture beyond 10 years were taken into account in the model, • if any adjustments were done for any particular types of fracture, The DSU confirms that an explanation of the derivation of the additional QALYs that may be gained by treatment after the 10-year time horizon is contained in the Strontium Ranelate Assessment Report [1] and the HTA report [6]. The explanation is relatively limited in its description. The DSU has also established that following an update based on more recent evidence, this methodology had been amended for use in the modelling from 2006 onwards and incorporated into the appraisal from thereafter. However, this had not been captured clearly in the Assessment Reports from 2006 onwards. The DSU provide a clarification of the 'bolt-on'

Pro-forma field	Comment	DSU response
	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. Another issue is that these adjustments only are related to preventable deaths during the 5 years of treatment.	methodology in Annex 3. The DSU confirm that the updates to the mortality modelling described here were included in the model from 2006 onwards, that is, they are included in the model used for the development of TA160/161. The DSU could not establish the rationale for including the variables 'phbonusat2.5' and 'wristbonusat2.5' Excluding these variables from the model would be unfavourable to the interventions. The methodology for the adjustments being undertaken only to those fractures within the first 5 years of treatment are contained in Annex 3. The DSU believe that the methodology employed may be slightly favourable to interventions. Sensitivity analyses performed using the executable models provided to consultees were conducted to assess the robustness of the results to changes in the assumptions regarding the time horizon and mortality estimates. (Annex 3) It is noted that when the base case adjustment for QALYs accrued beyond the 10-year horizon were doubled then there are only slight changes

Pro-forma field	Comment	DSU response
		in the results.
Description of proposed amendment	Amend or completely re-write the model to account for the ability to include the quality of life and mortality effects as mentioned.	
Result of amended model or expected impact on the result	Improve the accuracy of the estimate of costs and benefits and improve the cost effectiveness of treatment.	

<table>
<tr><th>Pro-forma field</th><th>Comment</th><th>DSU response</th></tr>
<tr><td>Issue</td><td>Issue 7 Risk multipliers for fracture risk</td><td></td></tr>
<tr><td>Description of problem</td><td>The risk multipliers found in the NICE report differ from those used in the NICE model
Table 7 Fracture risk multipliers cited in the report and those used in the NICE model
<table>
<tr><th></th><th colspan="3">Site of fracture</th></tr>
<tr><th>Age (years)</th><th>hip</th><th>wrist</th><th>humerus</th></tr>
<tr><td colspan="4">Values in the model</td></tr>
<tr><td>50</td><td>1.27</td><td>1.79</td><td>2.12</td></tr>
<tr><td>55</td><td>1.25</td><td>1.40</td><td>1.69</td></tr>
<tr><td>60</td><td>1.23</td><td>1.24</td><td>1.37</td></tr>
<tr><td>65</td><td>1.21</td><td>1.35</td><td>1.44</td></tr>
<tr><td>70</td><td>1.20</td><td>1.52</td><td>1.41</td></tr>
<tr><td>75</td><td>1.19</td><td>1.77</td><td>1.35</td></tr>
<tr><td colspan="4">Values in the report</td></tr>
</table></td><td>The DSU has established that there were typographical error made within the Strontium Ranelate Assessment Report (Table 4 p25) [1].The values contained within the model are correct and therefore this typographical error has no impact on the results produced.</td></tr>
</table>

Pro-forma field	Comment				DSU response
	50	1.27	1.79	2.12	
	55	1.25	***1.38***	1.69	
	60	1.23	***1.21***	1.37	
	65	1.21	***1.34***	1.44	
	70	1.20	***1.47***	1.41	
	75	1.19	***1.76***	1.35	
Description of proposed amendment	Amend the model or the report to gain consistency.				
Result of amended model or expected impact on the result	Not known				

Pro-forma field	Comment	DSU response
Issue	Issue 8 Discount rates	
Description of problem	Discount rates used are not those recommended by NICE. The model does not allow changes in the discount rates for costs or QALYs	The discount rates which applied at the start of the appraisal (6% for costs and 1.5% for benefits) were used throughout the work in accordance with NICE process. The DSU confirm that the discount rates used were described in the Strontium ranelate Assessment Report [1] and no comments were received from consultees on the discount rate during the appraisal. It is expected that were the current rates (3.5% for both benefits and costs) used then the cost effectiveness of all interventions would be reduced (potentially markedly) as a successful intervention would be associated with acquisition costs in the initial treatment period (which would increase using alternative discount rates) and a long-term utility benefit (which would decrease using alternative discount rates).
Description of proposed amendment	Amend model or consider new model capable of changing discount rates	
Result of amended model or expected impact on the result	Probably reduce cost effectiveness of treatment	The DSU would concur, although it is very likely that the ICER would rise. It is expected that only those scenarios where an intervention is cost-saving (due to the

Pro-forma field	Comment	DSU response
		reduction in fracture costs) would not experience an increase in the cost per QALY.

Pro-forma field	Comment	DSU response
Issue	Issue 9 Compliance	
Description of problem	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. 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 assumptions regarding compliance are fully documented. Strontium Ranelate Assessment Report (p58) [1]. These were understood and agreed by the appraisal committee and thus no comment from the DSU is provided.
Description of proposed amendment	Model compliance appropriately to remove the over estimate of costs and QALYs gained.	
Result of amended model or expected impact on the result	Not known	

Pro-forma field	Comment	DSU response
Issue	Issue 10 Side effects	
Description of problem	NICE have used same disutility for side effects for all treatments even though SR does not have the same as profile as BPs	The DSU confirms that the ScHARR review of adverse events was based on data for bisphosphonates. No formal review for SR was available to the Committee. Strontium ranelate is not associated with the same adverse gastrointestinal effects as bisphosphonates, and the clinical data for strontium ranelate indicate other adverse effects related to thromboembolism. The latter was not formally included in the model, and the Committee considered it appropriate to set the base case assumptions for strontium ranelate to the same as for bisphosphonates. As such no further comment is made by the DSU.
Description of proposed amendment	Use evidence from SR studies see p 27	
Result of amended model or expected impact on the result	Underestimates cost effectiveness of SR	

Pro-forma field	Comment	DSU response
Issue	Issue 11 Costs	
Description of problem	Hip fracture costs are out of date	As this is a comment on an input parameter,

	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 somewhat 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.	previously known to consultees and commentators, and agreed upon by the Committee, and therefore one on which consultees have previously been able to make representations, no DSU response is provided.
Description of proposed amendment	Use new data see p 27	
Result of amended model or expected impact on the result	Underestimates cost effectiveness of SR, Will reduce ICER of SR	

Pro-forma field	Comment	DSU response
Issue	Issue 12 QOL for vertebral fractures	
Description of problem	QOL data for vertebral fractures appears incorrect 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].	As this is a comment on an input parameter, previously known to consultees and commentators, and agreed upon by the Committee, and therefore one on which consultees have previously been able to make representations, no DSU response is provided.
Description of proposed amendment	Use best available evidence see p 27	
Result of amended model or expected impact on the result	Will reduce ICER of SR	

Pro-forma field	Comment	DSU response
Issue	Issue 13 Cost-effectiveness of identification strategies	
Description of problem	Identification strategies appear incorrectly costed and inappropriate. 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 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 to be based on the data used to develop the FRAX® algorithm). This assumes an identical prevalence of CRFs over the entire range of T-score which is clearly inappropriate. Indeed women with 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]. 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	The appraisal committee stated that they did not wish to treat women without a BMD scan except in exceptional circumstances ('those women aged 75 years or over if the responsible clinician considers a DXA scan to be clinically inappropriate or unfeasible'). The DSU confirm that the identification of women at high risk is fully documented in the Strontium Ranelate assessment report (p100) which was available to consultees [1]. The way in which the identification of women at high risk was carried out in the model was understood and agreed by the Appraisal Committee. As such no further comment is made by the DSU. The DSU has established that the model has the capability to produce results for all combinations of clinical risk factors; indeed such results have been previously estimated [6]. For example in Table 33 of this report, it is seen that the T-Score threshold for a 70 year old woman with a prior fracture and parental history of fracture to receive strontium ranelate was –2.4 SD, which fell to –3.1 SD for a 70 year old woman with a

Pro-forma field	Comment	DSU response
	does not conform to the population from which it was derived [Holt et al, 2002]. The assumed distribution adversely affects cost- effectiveness, 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 WHO approach is used for the same age range, the average requirement is 0.4 BMD scans per patient identified for treatment [Kanis et al, 2008i]	prior fracture who smoked. The absolute risk of fracture difference was less marked (4.66 and 4.49 respectively) indicating that there is not a direct correlation between absolute risk and cost-effectiveness ratios. Thus basing treatment decisions on absolute risk alone has limitations for the reasons outlined in the report [6]. In the Strontium Ranelate AR, the results from the NICE model were presented in a less aggregated way, with different thresholds produced at each age for all combinations of risk factors and with the T-Score bands divided into steps of 0.1SD rather than 0.5SD; the change in T-Score banding was at the request of the Appraisal Committee. Results presented in terms of specific CRF combinations were also presented to the NICE Guideline Development Group. Given that there was not a risk calculator available at this time there was a strong recommendation that a comprehensive list of CRF combinations would present GPs with a logistical problem and that the recommendations should be grouped by the number of risk factors possessed by a woman. For simplicity the median cost-effectiveness ratio was taken; it is acknowledged that this would favour those

Pro-forma field	Comment	DSU response
		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 average increased risk. The DSU has established that the coefficients for smoking and glucocorticoids used in the model were consistent with that in the WHO algorithm. The Appraisal Committee decided to not use smoking as a risk factor (section 4.3.8 in TA160/161), and the Institute decided to include glucocorticoid-induced osteoporosis as part of the clinical guideline (Pre-amble in TA160/161). The DSU has not explored this further, as this would require major changes to the model structure. The DSU expects, however, that based on the WHO coefficients, the inclusion of glucocorticoids, for example, as a risk factor in the model would favour the interventions. This is because the ICERs for women taking glucocorticoids were calculated in the model and were part of

Pro-forma field	Comment	DSU response
		calculating the median ICER per age and T-score group. Because the ICERs for women taking glucocorticoids were relatively lower than the average, this would favour the interventions appraised The DSU have checked the T-Scores values used in the model with those in the database underpinning the Holt et al paper provided to the assessment group by the data owners in 2004 (Annex 4). The DSU undertook additional analyses using the Holt et al database to test the assumption of normality assumed for T-Score distribution. (Annex 4). For some age bands this assumption did not hold. However, in these cases the most appropriate log-normal distribution would result in ICERs less favourable to the interventions. 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. The distributions used were based on the

Pro-forma field	Comment	DSU response
		Holt *et al* study, with raw data provided by the data owners to the Assessment Group. The DSU have checked these BMD distributions used within the NICE model to establish whether the assumption of normality and standard deviation of 1 T score are appropriate (Annex 4).
Description of proposed amendment	see p 40,41	
Result of amended model or expected impact on the result	Underestimates cost effectiveness of SR	The DSU agree with the fact that the ICERs calculated by the model will be unfavourable to intervention in women with risk factors conferring a higher than median risk. However, the DSU note that conversely the cost effectiveness will be overestimated in women with risk factors conferring a lower than median risk.

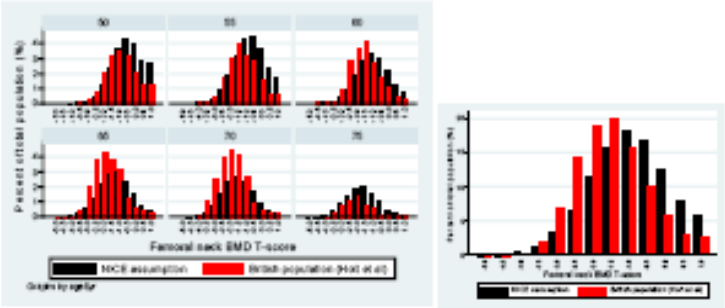
Comments received from Bone Research Society

Pro-forma field	Comment	DSU response
Issue	Issue 1 Alendronic acid assumed to have 10-fold the actual risk of side-effects that reduce quality of life	
Description of problem	The disutility associated with bisphosphonate use (eg alendronic acid) was over-estimated by a factor of 10 compared to the published literature. For those not familiar with the terminology of health economic modelling,	As this is a comment on an input parameter, previously known to consultees and commentators, and agreed upon by the

<table>
<tr><th>Pro-forma field</th><th>Comment</th><th>DSU response</th></tr>
<tr><td></td><td>disutility refers to the extent to which taking the drug is useless or counterproductive. It is quantitated according to the associated add-on costs of dealing with the disutility plus the reduction in quality-adjusted life years (QALYs) resulting from treatment that is attributable to the disutility.

Thus, when the disutility factor is increased for alendronic acid by a factor of 10, the benefits of treating those who receive treatment and still suffer no ill effects remain the same, while the numbers suffering disutility (or alternatively the impact of the disutility on the individual) are/is amplified ten-fold. The effect is to remove and sometimes reverse the benefit of treatment in those who stand to gain moderately from treatment in terms of fractures avoided.</td><td>Committee, and therefore one on which consultees have previously been able to make representations, no DSU response is provided.</td></tr>
<tr><td>Description of proposed amendment</td><td>Restore the side effect disutility to unity from its current value of 10-fold</td><td></td></tr>
<tr><td>Result of amended model or expected impact on the result</td><td>
<table>
<tr><th></th><th colspan="2">SE disutility = 10</th><th colspan="2">SE disutility = 2</th><th colspan="2">SE disutility = 1</th></tr>
<tr><th></th><th>CPQ</th><th>is it CE</th><th>CPQ</th><th>is it CE</th><th>CPQ</th><th>is it CE</th></tr>
<tr><td>Age 50</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td></tr>
<tr><td>Age 55</td><td>#DIV/0!</td><td>#DIV/0!</td><td>£113,616</td><td>-</td><td>£105,301</td><td>-</td></tr>
<tr><td>Age 60</td><td>£267,461</td><td>-</td><td>£31,753</td><td>-</td><td>£27,534</td><td>-</td></tr>
<tr><td>Age 65</td><td>£18,391</td><td>1</td><td>£15,301</td><td>1</td><td>£14,542</td><td>1</td></tr>
<tr><td>Age 70</td><td>£9,290</td><td>1</td><td>£8,562</td><td>1</td><td>£8,199</td><td>1</td></tr>
<tr><td>Age 75</td><td>£1,060</td><td>1</td><td>£2,172</td><td>1</td><td>£2,084</td><td>1</td></tr>
</table>
</td><td></td></tr>
</table>

Pro-forma field	Comment	DSU response
	<table><tr><th>BMD?</th><th colspan="4">SE disutility = 10</th><th colspan="4">SE disutility = 2</th><th colspan="4">SE disutility = 1</th></tr><tr><th>CRFs</th><th>0</th><th>1</th><th>2</th><th>3</th><th>0</th><th>1</th><th>2</th><th>3</th><th>0</th><th>1</th><th>2</th><th>3</th></tr><tr><td>Age 50</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>Age 55</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td><td>1</td></tr><tr><td>Age 60</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td>1</td><td>1</td><td></td><td></td><td>1</td><td>1</td></tr><tr><td>Age 65</td><td></td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>Age 70</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>Age 75</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr></table>	

Pro-forma field	Comment	DSU response
Issue	Issue 2 British women assumed to be at far less risk of osteoporosis at a given age than shown by the observational data, making identification less cost-effective than is actually the case.	
Description of problem	The proportions of women with low BMD (as estimated by BMD T-score) as input into the NICE model was output graphically and in tabular form and found to be substantially underestimated for England and Wales. The effect of this is to increase costs of identifying those needing treatment because more screening is required for each woman identified for treatment. We could not identify where the grossly elevated BMD T-score distributions came from; we substituted the distribution published by Holt et al (see below) which remains the largest database of T-scores for British women recruited from population registers and therefore as far as possible free from the effect of volunteer bias. Comparison of population distribution by 5-year age-group over femoral neck BMD T-score group in the NICE model versus observed distribution in 5173 British women aged 50-85 years from 7 centres across the UK (Aberdeen, Bath, Cambridge (City), Cambridge (Rural), Harrow, Norfolk, and Truro. [Holt G et al Br J Radiol. 2002 Sep;75(897):736-42]).  	The DSU have checked the T-Scores values used in the model with those in the database underpinning the Holt *et al* paper provided to the assessment group by the data owners in 2004. (Annex 4). The DSU undertook additional analyses using the Holt et al database to test the assumption of normality assumed for T-Score distribution. These analyses show that the assumptions of normality and a standard deviation of 1 are likely to be favourable to treatment (Annex 4).

<table>
<tr><th>Pro-forma field</th><th>Comment</th><th>DSU response</th></tr>
<tr><td>Description of proposed amendment</td><td>Set the population distribution of T-scores for the femoral neck to be the same as those published by Holt et al (and also restore the numbers of women to those actually known to be living in England and Wales in 2007 from the substantial underestimate found in the model)</td><td></td></tr>
<tr><td>Result of amended model or expected impact on the result</td><td>
<table>
<tr><th></th><th colspan="2">SE disutility = 2</th><th colspan="2">SE disutility = 1</th></tr>
<tr><th></th><th>CPQ</th><th>is it CE</th><th>CPQ</th><th>is it CE</th></tr>
<tr><td>Age 50</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td></tr>
<tr><td>Age 55</td><td>£65,686</td><td>-</td><td>£27,470</td><td>-</td></tr>
<tr><td>Age 60</td><td>£18,623</td><td>1</td><td>£17,523</td><td>1</td></tr>
<tr><td>Age 65</td><td>£10,650</td><td>1</td><td>£10,207</td><td>1</td></tr>
<tr><td>Age 70</td><td>£5,975</td><td>1</td><td>£5,716</td><td>1</td></tr>
<tr><td>Age 75</td><td>£668</td><td>1</td><td>£648</td><td>1</td></tr>
</table>
<table>
<tr><th>BMD?</th><th colspan="4">SE disutility = 2</th><th colspan="4">SE disutility = 1</th></tr>
<tr><th>CRFs</th><th>0</th><th>1</th><th>2</th><th>3</th><th>0</th><th>1</th><th>2</th><th>3</th></tr>
<tr><td>Age 50</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Age 55</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td>1</td><td>1</td></tr>
<tr><td>Age 60</td><td></td><td>1</td><td>1</td><td>1</td><td></td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>Age 65</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>Age 70</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>Age 75</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
</table>
</td><td></td></tr>
</table>

Pro-forma field	Comment	DSU response
Issue	Issue 3 Incremental Cost-Effectiveness Ratios assumed to be identical for all subgroups of women in a 5-year age band, irrespective of their BMD-independent risk factors. This excludes women from treatment with non-BMD related higher than average risk	
Description of problem	Use of mean population ICERS at each BMD level to determine whether an age-cohort was eligible for treatment, irrespective of numbers of clinical risk factors additional to a specific BMD level	The DSU believed that the modelling methodology proposed by the Bone Research Society is fundamentally incorrect. Women who may be extremely cost effective to treat should not subsidise women who cannot be treated cost-effectively. Subgroup analysis should not only be used where the mean ICER is above a recommended threshold. Given this, the DSU believe that the modelling methodology employed in the original assessment is correct.
Description of proposed amendment	Where mean ICER shows non-cost effectiveness, proceed to sub-group ICER analysis (as shown in table to right) before excluding subgroups from treatment.	

Pro-forma field	Comment	DSU response
Result of amended model or expected impact on the result	Median ICERs by No of CRFs, age, and BMD T-score assuming SE disutility factor of 1	

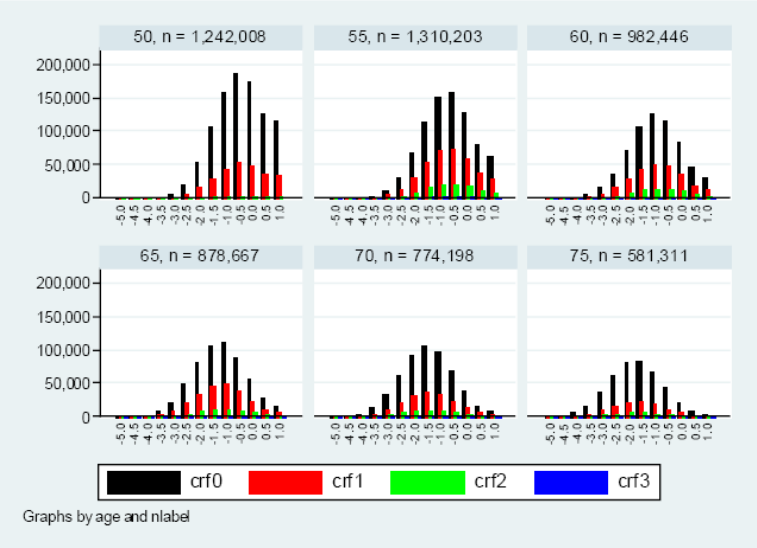
No of CRFs and age	BMD T-score												
	-5	-4.5	-4	-3.5	-3	-2.5	-2	-1.5	-1	-.5	0	.5	
0 CRF													
50	-8,702	-7,051	-4,084	1,078	9,636	23,015	42,565	69,404	104,792	151,418	189,506	223,218	25
55	-9,049	-7,232	-4,282	305	7,040	16,355	28,476	43,507	61,660	80,690	94,927	108,796	12
60	-8,171	-6,445	-3,787	177	5,833	13,599	23,852	36,983	53,542	68,380	82,877	97,911	11
65	-9,392	-7,745	-5,395	-2,129	2,257	8,003	15,338	24,525	34,764	44,137	53,957	64,284	7
70	-9,090	-7,574	-5,651	-3,250	-337	3,179	7,394	12,431	16,976	21,661	26,599	31,846	3
75	-9,731	-8,581	-7,163	-5,422	-3,326	-782	2,309	5,770	9,177	12,986	17,156	21,737	2
1 CRF													
50	-9,035	-7,667	-5,227	-1,028	5,851	16,748	31,450	50,833	74,983	105,336	129,134	149,141	16
55	-9,462	-7,933	-5,440	-1,545	4,028	11,519	21,255	32,758	46,207	59,387	69,675	79,426	8
60	-8,616	-7,157	-4,903	-1,527	3,316	9,561	18,073	28,340	40,529	51,142	61,387	71,737	8
65	-9,877	-8,462	-6,440	-3,719	-103	4,931	11,431	18,900	26,712	33,967	41,438	49,162	5
70	-9,779	-8,491	-6,856	-4,778	-2,111	1,168	5,139	9,274	12,966	16,764	20,720	24,877	2
75	-10,343	-9,356	-8,142	-6,656	-4,834	-2,503	359	3,627	6,679	9,851	13,283	17,010	2
2 CRF													
50	-9,415	-8,376	-6,474	-3,110	2,352	11,063	22,806	35,784	55,118	76,513	92,562	105,594	12
55	-9,933	-8,742	-6,788	-3,628	817	7,617	14,485	22,875	34,134	44,691	52,399	59,553	6
60	-9,122	-7,981	-6,211	-3,463	307	5,817	12,201	19,120	29,300	37,968	45,957	53,852	6
65	-10,527	-9,445	-7,903	-5,756	-2,780	1,332	6,166	11,673	18,380	25,026	31,637	38,370	4
70	-10,860	-9,834	-8,483	-6,729	-4,620	-2,089	964	4,536	7,686	10,948	15,112	19,838	2
75	-11,406	-10,615	-9,614	-8,354	-6,790	-4,946	-2,757	-249	2,342	5,013	7,832	10,855	1
3 CRF													
50	-9,636	-8,803	-7,330	-4,814	-735	5,465	14,199	25,599	39,541	55,939	68,683	79,089	8
55	-10,194	-9,222	-7,669	-5,292	-1,846	2,860	8,895	16,206	24,694	33,285	39,764	45,703	5
60	-9,403	-8,471	-7,063	-4,998	-2,096	1,826	6,908	13,248	20,934	27,889	34,501	40,990	4
65	-11,038	-10,198	-8,993	-7,301	-4,994	-1,933	2,009	6,945	12,528	17,873	23,363	28,929	3
70	-11,692	-10,962	-9,996	-8,737	-7,126	-5,099	-2,584	493	3,658	7,035	10,596	14,308	1
75	-12,446	-11,934	-11,284	-10,460	-9,421	-8,121	-6,495	-4,540	-2,361	163	3,019	6,213	

Pro-forma field	Comment	DSU response
Issue	Issue 4 Absence of modelling of continuous variables known to the GP that confer risk independently of BMD	
Description of problem	Continuous variables that confer risk independently of BMD are un-modelled (such as lower BMI, eg under 25 which independently increases risk of hip fracture by up to two-fold: de Laet et al 2005 Osteoporos Int 2005 16:1330-8). This disadvantages some high risk subjects	In TA 160/161 BMD is measured in women who receive treatment except in exceptional circumstances ('those women aged 75 years or over if the responsible clinician considers a DXA scan to be clinically inappropriate or unfeasible'). The issue presented is that the model is not sensitive to BMI which is unfavourable to interventions in some women. The DSU has established that it is correct that BMI is set at 26kg/m2. Furthermore, the DSU has explored the impact of BMI on the fracture risk estimated by the WHO algorithm to see if it is correct that using a fixed BMI of 26kg/m2 is unfavourable to interventions. The DSU could not establish why the coefficient of fracture prediction associated with BMI within the WHO algorithm changed markedly when BMD was known as the correlation between the variables in the Holt et al dataset was small. (R2 = 0.079) (Annex 2). Nor was there a strong correlation between BMI and age (R2 = 0.004) Due to the large number of potential permutations of T-Score bands and BMI bands the NICE model did not subdivide women into both T-Score and BMI categories; only T-Score was used to categorise women in addition to age and prior

Pro-forma field	Comment	DSU response
		fracture status (Annex 2). A BMI of 26 kg/m^2,was used for all women in the model, which was the mean value from the Holt et al database. Exploratory analyses of the risk of fracture using BMI values of 20 kg/m^2 and 32 kg/m^2 (which encompass over 85% of women in the Holt et al database) within the WHO algorithm were used to assess the change in fracture risk. The exploratory analyses show that, once BMD is known, BMI values of 20 kg/m^2 and 32 kg/m^2 do not lead to higher fracture risks than a BMI of 26 kg/m^2.
Description of proposed amendment	Risk attributable to various levels of BMI independently of BMD may be modelled by rescaling the currently assumed age-specific absolute fracture risks at a given BMD level by the relative risk appropriate for each BMI level. This would only apply to low BMI values.	
Result of amended model or expected impact on the result	ICERs after scaling the age-specific 1-year hip fracture probabilities by 2.0 to reflect the hip fracture risk of a woman with BMI of 15 kg/m2.	

Pro-forma field	Comment	DSU response
	<table><tr><th></th><th colspan="2">SE disutility = 2</th><th colspan="2">SE disutility = 1</th></tr><tr><th></th><th>CPQ</th><th>is it CE</th><th>CPQ</th><th>is it CE</th></tr><tr><td>Age 50</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td><td>#DIV/0!</td></tr><tr><td>Age 55</td><td>£17,825</td><td>1</td><td>£16,646</td><td>1</td></tr><tr><td>Age 60</td><td>£12,361</td><td>1</td><td>£11,582</td><td>1</td></tr><tr><td>Age 65</td><td>£4,436</td><td>1</td><td>£4,227</td><td>1</td></tr><tr><td>Age 70</td><td>£743</td><td>1</td><td>£736</td><td>1</td></tr><tr><td>Age 75</td><td>-£4,406</td><td>1</td><td>-£4,222</td><td>1</td></tr></table> <table><tr><th>BMD?</th><th colspan="4">SE disutility = 2</th><th colspan="4">SE disutility = 1</th></tr><tr><th>CRFs</th><th>0</th><th>1</th><th>2</th><th>3</th><th>0</th><th>1</th><th>2</th><th>3</th></tr><tr><td>Age 50</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>Age 55</td><td></td><td>1</td><td>1</td><td>1</td><td></td><td>1</td><td>1</td><td>1</td></tr><tr><td>Age 60</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>Age 65</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>Age 70</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr><tr><td>Age 75</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr></table>	

Pro-forma field	Comment	DSU response
Issue:	Issue 5 Distribution of BMD values according to number of CRF	
Description of problem	Distribution of BMD values according to numbers of clinical risk factors as output by the NICE model. Unexpectedly (based on our reading of the evidence) we found that distributions were very similar (see histograms below).	For response to the BMI point please see response to BRS issue 4 above. Additional work was undertaken to explore how the results changed were there a relationship

Pro-forma field	Comment	DSU response
	 We searched through the model spreadsheets for appropriate interactions as detailed in the individual-level meta-analysis of Kanis et al (Osteoporos. Int 2007 18: 1033-46) and found no evidence of their presence in the model in active form	assumed between BMD and the presence of clinical risk factors. None of the sensitivity analyses undertaken altered the recommended strategies, although the number of women who may be applicable for treatment would increase (Annex 4).
Description of proposed amendment	Implement key interactions, such as the one between low BMD and low BMI (which increase risk above that expected for low BMD in presence of a normal BMI, which appears to be 26 in all simulations, whether BMI is 26 or some other figure.	
Result of amended model or expected impact on the result	The minority of very high risk younger women with low BMI and low BMD would get a more appropriate recommendation for alendronate.	

Pro-forma field	Comment	DSU response
Issue	Issue 6 Inadequate documentation of the model	The DSU established that the model has been detailed in an HTA monograph [6] and was subject to peer review both by the reviewers for the HTA and by reviewers for the Journal of Operational Research Society [7]. The DSU has found further correspondence has occurred in Bone [8], with the assertion that the results of the model are relatively similar to those of a contemporary model [9] when the same assumptions are used. Evidence submitted to Osteoporosis International to provide the data to support this argument was not accepted for publication. (Annex 5)
Description of problem	Documentation of the model is sketchy. If another modeller took over from Dr Stevenson, there appears a serious risk of mistakes being made through misunderstanding of the sometimes nonexistent and sometimes ultra-cryptic comment fields.	
Description of proposed amendment	Matt Stevenson should be commissioned to document the model thoroughly, in its final form, assuming that NICE TA 160/1 in final form are based on a revised version of this model. The model should then be subjected to external, independent peer review and published in a high grade scientific journal under the names of the modeller and the commissioning Chair to establish scientific responsibility.	The model provided to consultees and commentators is the model used for the development of the recommendations in TA160/161. To produce another model or a revised version of the existing model with different functionality would be outside the process and inconsistent with the undertaking to the Court.
Result of amended model or expected	Reduction in the risk of serious future errors by up to an order of	

Pro-forma field	Comment	DSU response
impact on the result	magnitude.	

Pro-forma field	Comment	DSU response
Issue	Issue 7 Alcohol intake	
Description of problem	The rationale for the choice of 4 or more units per day intake is not justified anywhere within the NICE documentation. Even if the choice is made to use this threshold, then the coefficient for alcohol intake is incorrect e.g. for hip fracture the coefficient appears to be 1.53, whereas the published literature (Kanis et al, Osteoporos Int. 2005;16: 737-42) demonstrates that the coefficient for 4 units or more should be 2.26-2.39..	The DSU has established that the coefficients for alcohol used in the model was consistent with that in the WHO algorithm, i.e. for >2 units of alcohol intake per day. The DSU has not quantitatively explored the effect of the Committee decision to used a threshold of more than 4 units of alcohol on the estimated fracture risks and subsequent ICERs. . The DSU has estimated that were midpoint values used then it is expected that the risks would increase in women who consumed >4 units of alcohol compared with >2 units of alcohol. The DSU note however that the confidence intervals around the risk ratios are wide and that no data on statistical significance has been provided by the consultee. As such a positive correlation between risk ratio and alcohol consumption may not exist. If this were true then the weighted midpoint may rise only marginally compared with the value of >2 units.
Description of proposed amendment	The alcohol threshold should be modelled at the FRAX threshold of 3 units or more daily and the correct coefficient should be applied	
Result of amended model or expected impact on the result	The ICER will improve	

Pro-forma field	Comment	DSU response
Issue:	Issue 8 Smoking and glucocorticoids	
Description of problem	It is unclear but the spreadsheets appear to suggest that the risks attributable to smoking and glucocorticoid use are included in the identification strategies, but these CRFs are not considered by NICE to be relevant risk factors in the appraisal.	The DSU has established that the coefficients for smoking and glucocorticoids used in the model were consistent with that in the WHO algorithm. The Appraisal Committee decided to not use smoking as a risk factor (section 4.3.8 in TA160/161), and the Institute decided to include glucocorticoid-induced osteoporosis as part of the clinical guideline (Pre-amble in TA160/161). The DSU has not explored this further. The DSU expects, however, that based on the WHO coefficients, the inclusion of glucocorticoids, for example, as a risk factor in the model would favour the interventions. This is because the ICERs for women taking glucocorticoids were calculated in the model and were part of calculating the median ICER per age and T- score group. Because the ICERs for women taking glucocorticoids were relatively lower than the average, this would favour the interventions appraised.
Description of proposed amendment	The model should embrace these risk factors and include the full FRAX algorithm in the strategy for osteoporosis management	
Result of amended model or expected impact on the result		

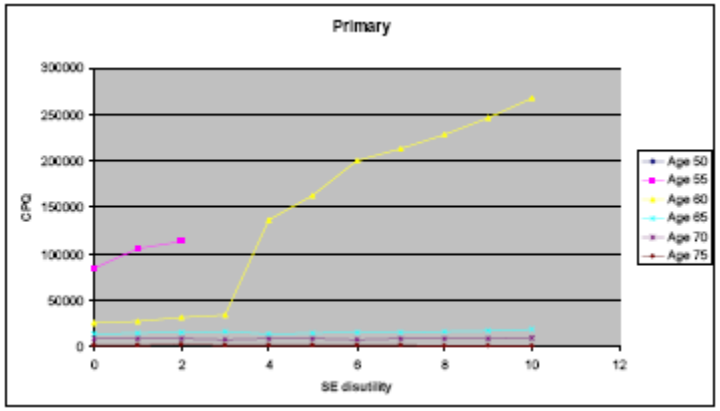
Pro-forma field	Comment	DSU response
Issue	Issue 9 Lack of interactions between risk factors in the model	
Description of problem	There is compelling evidence of significant interactions between several of the risk factors that impact on risk assessment. These interactions are incorporated within FRAX but not within the NICE model and will have an adverse effect on cost-effectiveness especially at younger ages. For example a prior fracture has greater significance at younger ages than in the more elderly population.	The DSU confirm that the interactions referred to -prior fracture:age and BMD:age- have been incorporated within the NICE model. The fracture risks used within the model are contained in the age 50 through to age 75 worksheets of the CPQ Calc Est and CPQ Calc Prev sheets and have been directly calculated from the WHO algorithm using all interactions. These interactions were incorporated into the fracture risk calculated using the regression formula from the WHO algorithm and the resulting risk values pasted in the executable models. As only the executable model was exempt of the AIC, the regression formula calculating the risks could not be released to C&Cs.
Description of proposed amendment	The NICE model should be adapted to accommodate interactions such as BMD and fracture, BMD and BMI etc.	
Result of amended model or expected impact on the result	The ICER at younger ages will be improved	

Comments received from National osteoporosis society

Pro-forma field	Comment	DSU response
Issue	Issue 1 Clarity of the model	The DSU established that the model has been detailed in an HTA monograph [6] and was subject to peer review both by the reviewers for the HTA and by reviewers for the Journal of Operational Research Society [7]. The DSU has found further correspondence has occurred in Bone [8], with the assertion that the results of the model are relatively similar to those of a contemporary model [9] when the same assumptions are used. Evidence submitted to Osteoporosis International to provide the data to support this argument was not accepted for publication. (Annex 5)
Description of problem	The instructions provided and comments within the spreadsheets of the model fall well short of transparency. The information provided is extremely limited and forms a substantial barrier to the charity providing meaningful comment on the economic model.	
Description of proposed amendment	As a matter of record, the model needs to be fully documented and interpretable by external users.	
Result of amended model or expected impact on the result	N/A	

Pro-forma field	Comment	DSU response
Issue	Issue 2 Population Data	
Description of problem	In correspondence with NICE during the consultation period the National Osteoporosis Society asked for further information with regards the population data. We were subsequently provided with the original individual patient simulation model but informed that it was not used in formulating current TA 160/161 guidance. It appears that the distribution of BMD in the NICE model differs quite markedly from published data within the UK. The source of the population data is unclear.	The DSU could not establish why the NOS claims that the BMD in the model differs quite markedly for the BMD distribution in the UK. The distributions used were based on the Holt et al study, with raw data provided by the data owners to the Assessment Group. The DSU have checked these BMD distributions used within the NICE model to establish whether the assumption of normality and standard deviation of 1 T score are appropriate (Annex 4). For some age bands this assumption did not hold. However, in these cases the most appropriate log-normal distribution would result in ICERs less favourable to the interventions.
Description of proposed amendment	Provide more information on the patient simulation model used within TA 106/161 to allow us to fully execute the model. Adjust the population distributions of BMD to accurately reflect the observed distribution in the UK	
Result of amended model or expected impact on the result	The ICERs will improve.	

Pro-forma field	Comment	DSU response
Issue	Issue 3 Inflation of side effect disutility	
Description of problem	A Side effect disutility factor of 10 has been used in the model. In an evidence based setting, there appears to be a complete lack of evidence to support the use of this assumption. It has a dramatic effect on the ICER within younger women in the prevention setting with a threshold effect at an SE disutility multiplier of 4 (which is still not justifiable from the literature). An explanation of the marked effect beyond a multiplier of 4 at younger ages needs to be provided – it is not apparent from the model why this should be the case.	As this is a comment on an input parameter, previously known to consultees and commentators, and agreed upon by the Committee, and therefore one on which consultees have previously been able to make representations, no DSU response is provided on the general point. The DSU has explored why the consultee believes there are 'break points' at certain SE disutility factors. The DSU believe that these are caused where the incremental QALY changes from being positive to negative. In these circumstances there will be a marked effect on the cost per QALY ratio. Assuming positive incremental costs this would result in a change from a high positive cost per QALY to a high negative cost per QALY (where the intervention is dominated)
Description of proposed amendment	Return the SE disutility factor to the evidence based estimates (i.e. a multiplier of 1)	

Pro-forma field	Comment	DSU response
Result of amended model or expected impact on the result	Primary <table><tr><th>SE Disutility Factor</th><th>10</th><th>1</th></tr><tr><td>Age 50</td><td></td><td></td></tr><tr><td>Age 55</td><td></td><td>£105,301</td></tr><tr><td>Age 60</td><td>£267,460</td><td>£27,534</td></tr><tr><td>Age 65</td><td>£18,391</td><td>£14,542</td></tr><tr><td>Age 70</td><td>£9,290</td><td>£8,199</td></tr><tr><td>Age 75</td><td>£1,060</td><td>£2,084</td></tr></table> 	

Pro-forma field	Comment	DSU response
	Secondary <table><tr><th>SE Disutility Factor</th><th>10</th><th>1</th></tr><tr><td>Age 50</td><td>£ 26,894</td><td>£ 25,424</td></tr><tr><td>Age 55</td><td>£ 18,333</td><td>£ 17,914</td></tr><tr><td>Age 60</td><td>£ 14,806</td><td>£ 13,950</td></tr><tr><td>Age 65</td><td>£ 6,211</td><td>£ 9,142</td></tr><tr><td>Age 70</td><td>£ 3,389</td><td>£ 2,823</td></tr><tr><td>Age 75</td><td>-£ 2,369</td><td>-£ 1,659</td></tr></table>	

Pro-forma field	Comment	DSU response
Issue	Issue 4 Clinical risk factors	
Description of problem	The model uses a number of risk factors (current smoking, corticosteroid use previous or current) that are not included in the guidance.	The DSU has established that the coefficients for smoking and glucocorticoids used in the model were consistent with that in the WHO algorithm. The Appraisal Committee decided to not use smoking as a risk factor (section 4.3.8 in TA160/161), and the Institute decided to include glucocorticoid-induced osteoporosis as part of the clinical guideline (Pre-amble in TA160/161). The DSU has not explored this further. The DSU expects, however, that based on the WHO coefficients, the inclusion of glucocorticoids, for example, as a risk factor in the model would favour the interventions. This is because the ICERs for women taking glucocorticoids were calculated in the model and were part of calculating the median ICER per age and T-score group. Because the ICERs for women taking glucocorticoids were relatively lower than the average, this would favour the interventions appraised
Description of proposed amendment	Incorporate these risk factors into the guidance and/or produce separate guidance for glucocorticoid users	

Pro-forma field	Comment	DSU response
Result of amended model or expected impact on the result	Incorporating smoking as a risk factor in the guidance would acknowledge the increased risk that the 21% of women aged 50-59 and 12% of women aged 60+ who smoke have1. This will ensure that they receive the appropriate treatment commensurate with their fracture risk. 1http://www.statistics.gov.uk/downloads/theme_compendia/GHS07/GHSSmokingandDrinkingAm ongAdults2007.pdf	

Pro-forma field	Comment	DSU response
Issue	Issue 5 Alcohol CRF	
Description of problem	Within the guidance an alcohol intake of 4 or more units per day is used. However the model appears to use an intake of greater than 2 units per day. The coefficient for the latter will be inappropriately low for the 4 unit threshold and underestimate fracture risk. Again, the documentation of the choice of a 4 unit threshold is sadly lacking and is an extremely rare occurrence in post-menopausal women in the UK.	The DSU has established that the coefficients for alcohol used in the model was consistent with that in the WHO algorithm, i.e. for >2 units of alcohol intake per day. The DSU has not quantitatively explored the effect of the Committee decision to used a threshold of more than 4 units of alcohol on the estimated fracture risks and subsequent ICERs. The DSU has estimated that were midpoint values used then it is expected that the risks would increase in women who consumed >4 units of alcohol compared with >2 units of alcohol. The DSU note however that the confidence intervals around the risk ratios are wide and that no data on statistical significance has been provided by the consultee. As such a positive correlation between risk ratio and alcohol consumption may not exist. If this were true then the weighted midpoint may rise only marginally compared with the value of >2 units.

Pro-forma field	Comment	DSU response
Description of proposed amendment	Ensure that there is consistency between the guidance and information used within the economic model. Use appropriate thresholds and their associated coefficients.	
Result of amended model or expected impact on the result	N/A	

Pro-forma field	Comment	DSU response
Issue	Issue 6 Sensitivity Analysis	
Description of problem	The model as supplied does not permit alterations to a number of elements preventing sensitivity analysis 1. Body mass index (BMI) is set at a fixed value. This is not consistent with the construct of FRAX® and the gradient of fracture risk rises dramatically as BMI falls, independent of other risk factors such as age. 2. The NICE model uses predominantly a ten-year time horizon which has a large effect on apparent costeffectiveness. 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 calculations made to remedy the deficit in the model. 3. The model is populated with pre-specified clinical risk factor estimations, so that sensitivity analysis around the assumptions cannot be performed.	The issue presented is that the model is not sensitive to BMI which is unfavourable to interventions in some women. The DSU has established that it is correct that BMI is set at 26kg/m^2. Furthermore, the DSU has explored the impact of BMI on the fracture risk estimated by the WHO algorithm to see if it is correct that using a fixed BMI of 26kg/m^2 is unfavourable to interventions. The DSU could not establish why the coefficient of fracture prediction associated with BMI within the WHO algorithm changed markedly when BMD was known as the correlation between the variables in the Holt et al dataset was small. (R^2 = 0.079) (Annex 2). Nor was there a strong correlation between BMI and age (R^2 = 0.004) Due to the large number of potential permutations of T-Score bands and BMI bands the NICE model did not subdivide women into both T-Score and BMI categories; only T-Score was used to categorise women in addition to age and prior fracture status (Annex 2). A BMI of 26 kg/m^2 ,was used for all women in the model, which was the mean

Pro-forma field	Comment	DSU response
		value from the Holt et al database. Exploratory analyses of the risk of fracture using BMI values of 20 kg/m^2 and 32 kg/m2 (which encompass over 85% of women in the Holt et al database) within the WHO algorithm were used to assess the change in fracture risk. The exploratory analyses show that, once BMD is known, BMI values of 20 kg/m^2 and 32 kg/m^2 do not lead to higher fracture risks than a BMI of 26 kg/m^2. The DSU believe 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 [11].It is noted that the results produced by the 'NICE' model and the reference model are similar when populated with similar input parameters (Annex 5) The baseline risk of fracture was taken from the WHO algorithm and was assumed to be correct. No correlations were provided between the variables within the WHO algorithm, and thus only the midpoint estimates could be used without the risk of sampling algorithms that would not fit the

Pro-forma field	Comment	DSU response
		underlying data. The epidemiology of osteoporosis within the model was driven by the WHO algorithm and the dataset from the Holt et al publication [10] provided to the assessment group; both of these have been taken on trust and assumed to be correct.
Description of proposed amendment	Amend model to allow sensitivity analysis of varying BMI Amend model to allow sensitivity analysis of varying time horizon Amend model to allow sensitivity analysis of varying CRFs	
Result of amended model or expected impact on the result	This will improve the accuracy of all risk estimates (other than the current fixed value of 26kg/m2).	

Comments received from the Society for Endocrinology

Pro-forma field	Comment	DSU response
Issue	Issue 1 Complexity of data	
Description of problem	The information provided was unintelligible to the majority of the commentators. Moreover, in the time that had been allowed to feedback to NICE, it has not been possible for these commentators to consult individuals with relevant expertise.	The DSU established that the model has been detailed in an HTA monograph [6] and was subject to peer review both by the reviewers for the HTA and by reviewers for the Journal of Operational Research Society [7]. The DSU has found further correspondence has occurred in Bone [8], with the assertion that the results of the model are relatively similar to those of a contemporary model [9] when the same assumptions are used. Evidence submitted to Osteoporosis International to provide the data to support this argument was not accepted for publication. (Annex 5)
Description of proposed amendment	-	
Result of amended model or expected impact on the result	-	

Pro-forma field	Comment	DSU response
Issue	Issue 2 Stalled clinical guidance	
Description of problem	Clinical guidance on ibandronic acid and zoledronic acid has been delayed due to issues surrounding TAG160 and 161.	General comment to NICE, not related to model. No DSU response needed.
Description of proposed amendment	Select these compounds under “topics for guidance”.	
Result of amended model or expected impact on the result		

Pro-forma field	Comment	DSU response
Issue	Issue 3 Cost reduction	
Description of problem	Non-generics will be available for prescribing.	General comment to NICE, not related to model. No DSU response needed.
Description of proposed amendment	NICE to consider the resulting reduction in costs.	
Result of amended model or expected impact on the result		

Annex 2

Exploration of the use of Body Mass Index (BMI) within the model

BMI has been set within the model to be a fixed value of 26 kg/m^2.This was the average value derived from the dataset forming the population reviewed in the Holt et al paper [10] supplied to the Assessment Group in 2004 by the owners of the data. The average value was used to simplify the number of patient groups considered within the model. This decision was based on their being no clear relationship between T-Score and BMI or age and BMI. The statistical fits between the data within the Holt et al database [10] are shown in Figure 1 and Figure 2. There is a small positive relationship between T-Score and BMI. It is noted that in neither regression was the R^2 value greater than 0.1.

Figure 1: The relationship between T-Score and BMI

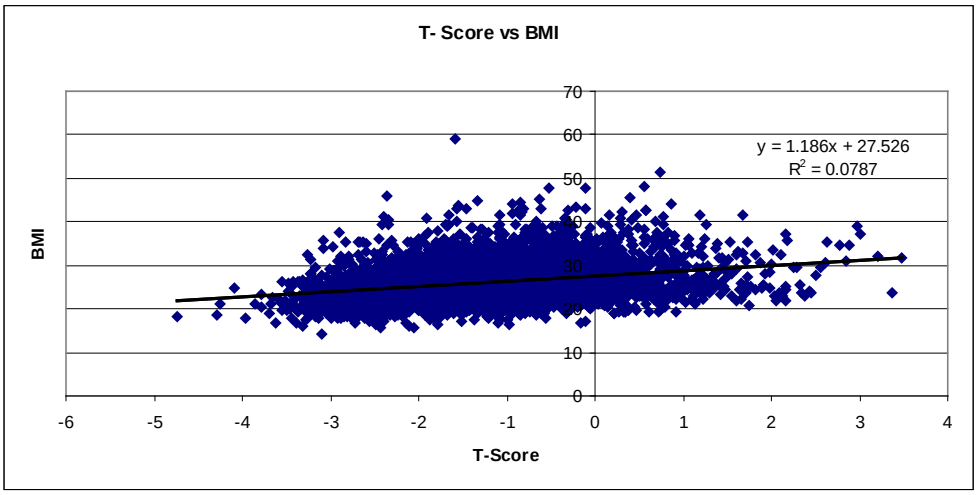
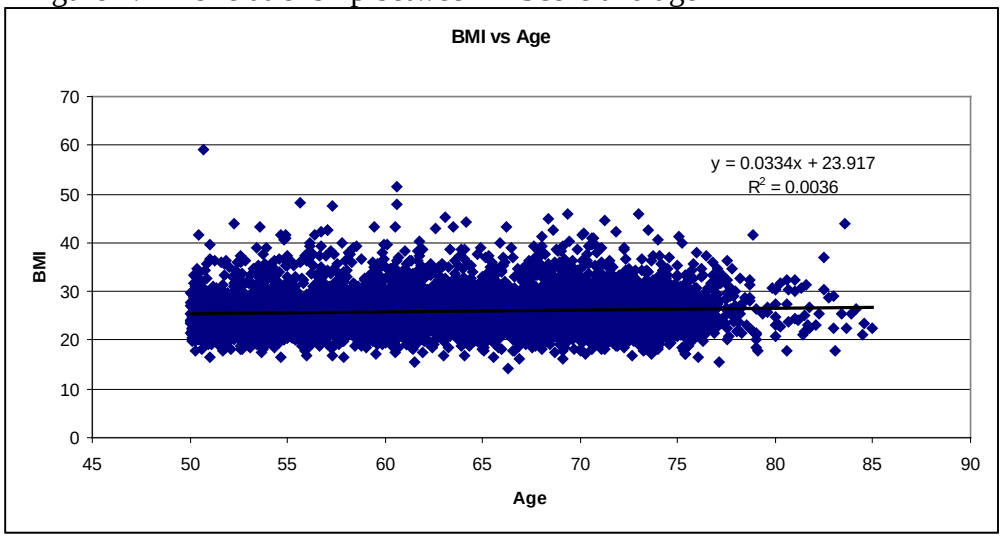


Figure 2: The relationship between T-Score and age



The Appraisal Committee concluded that it is important to establish the BMD before a treatment decision is made for most women. Only in circumstance where the likelihood of a low BMD was high, the Committee recommended that a DXA scan may not be required if the responsible clinician considers it to be clinically inappropriate or unfeasible. These exceptional circumstances were described in the guidance from secondary prevention as being applicable to 'women aged 75 years or over if the responsible clinician considers a DXA scan to be clinically inappropriate or unfeasible'. Thus risks derived from the WHO algorithm would be used in the model where BMD is already known.

Consultees have claimed that using a fixed BMI of 26kg/m^2 is unfavourable to interventions and also quote a paper indicating that a lower BMI should be associated with a greater risk of fracture [13]. The DSU has explored the impact of BMI on the fracture risk estimated by the WHO algorithm to test the consultees' claim. In order to do so, exploratory analyses were performed by exploring the risk of fracture predicted by the WHO algorithm when BMI was set to 20 kg/m^2 and when the value was set to 32 kg/m^2, and when BMD is known.

Figure 3: The effects of reducing the BMI of a woman on hip fracture. [confidential]

Figure 4: The effects of reducing the BMI of a woman on non-hip fracture. [confidential]

Therefore, these exploratory analyses show that, once BMD is known, BMI values of 20 kg/m^2 and 32 kg/m^2 do not lead to higher fracture risks than a BMI of 26 kg/m^2 and that the relationship in the publication quoted [13] is not exhibited in the WHO algorithm provided to the Assessment Group.

ANNEX 3

Clarification of the methods used in the bolt-ons in the model - mortality beyond the 10 year time horizon and mortality attributable to hip, vertebral and wrist fractures

Consultees have commented that the estimation of the 'bolt-on' cost consequences which are included in the model is not sufficiently transparent.

The DSU confirms that an explanation of the derivation of the additional QALYs that may be gained by treatment after the 10-year time horizon is contained in the Strontium Ranelate Assessment Report [1] and the HTA report [6]. The explanation is relatively limited in its description.

The DSU has also established that following an update based on more recent evidence, this methodology had been amended for use in the modelling from 2006 onwards and incorporated into the appraisal from thereafter. However, this had not been captured clearly in the Assessment Reports from 2006 onwards. The DSU provide a clarification of the 'bolt-on' methodology in the first part of this Annex. The DSU confirm that the updates to the mortality modelling described here were included in the model from 2006 onwards, that is, they are included in the model used for the development of TA160/161.

The model was constructed covering a time period of 10 years, which covered the assumed treatment duration of 5 years and a residual waning of efficacy over a 5-year period. The 10-year time horizon was necessitated by the desire to reduce the computational time required originally by running a large number of individual patient models. Restricting the time horizon to 10 years would underestimate the quality adjusted life years (QALYs) gained by preventing mortality compared with a lifetime horizon. This underestimation would be more pronounced where the mortality avoided occurred in younger women. Additionally, if women avoided a fracture with long-term disutility consequences in the individual patient model then the QALYs gained would also be underestimated as this benefit would be expected to continue beyond the 10 year time horizon.

In order to address these issues, additional calculations were performed to determine the likely magnitude of the underestimation of the QALYs gained. The DSU notes the following points:

1) At the time of construction of the original individual patient model there were no robust data to correlate vertebral or proximal humerus fractures with an increased risk of mortality. Subsequent data, detailed in the Strontium Ranelate Assessment Report p28 [1], have shown a mortality risk, attributable to these fractures, is likely. The results of the model were adjusted to take into account attributable mortality using the following methodology for all modeling from 2006 onwards:
2) The model used a mortality rate attributable to hip fracture that was greater than that predicted by more recent data. Therefore, the QALYs gained through

avoidance of hip fracture needs to be decreased as undertaken using the methodology described in the forthcoming pages.

3) In fitting the meta-model to the data produced by the individual patient model the residual benefit of treatment after cessation (the 'fall time') was problematic and the model was fitted without this parameter, with the assumption that the fall time was set to zero years. The likely gain in QALYs and potential cost savings within the fall time was estimated using the following methodology. It was assumed that the incremental QALYs and incremental costs saved (excluding drug acquisition costs) predicted by the meta-model were solely accrued in the initial 5-year treatment period, and that (assuming no discounting nor mortality) that there would be an additional 50% of QALYs gained and cost saved during the fall time which would be associated with the average benefit over the 5-year period. Discounting and mortality were incorporated and affected the values which ranged from approximately 48% increase in QALYs and 37% increase in costs savings at age 50-54 years through to a 38% increase in QALYs and 32% increase in costs savings at age 70-74 years. This methodology is likely to be favourable to the interventions as the majority of the individual patient runs had 3 or 5 years fall time rather than the zero assumed in the meta-model, but could be unfavourable to treatments were there a significant rise in fracture risk during years 5-10 of the model, which would not be considered in the individual patient runs where the fall time was zero. The DSU believe that on balance the methodology is likely to slightly favour the interventions.

Description of how the model took account of the QALYs associated with mortality attributable to hip fracture occurring after 10 years

In order to calculate the likely degree of underestimation of QALYs gained due to the 10-year time horizon of the model, an estimation of the QALYs accrued by women following the termination of the model was required. This was achieved by multiplying the standard mortality rate for women [14] by a factor of 1.22 per Z-Score [15] assuming that women were at the threshold for osteoporosis. During the period between the end of the 10-year modelling horizon and death it was assumed that a woman would have a utility associated with the average woman at that age [16]. This assumption is not unfavorable to interventions as it does not include any assumptions about the disutility associated with a previous fracture. Utility values were discounted at 1.5% per annum, which was the prevailing recommended discount rate at the start of the project.

These calculations led to the expected QALYs gained beyond the time horizon of the model shown in Table 1. As expected, these values decrease as the age of the woman in the model increases. It is noted that these values may overestimate the QALYs gained as these women may suffer a further fracture that causes mortality.

Table 1. The estimated discounted QALYs gained by women alive at the end of the model

Age (years)	Estimated QALYs 10 years after treatment initiation
50	11.12
55	8.11
60	6.74
65	4.92
70	3.33
75	2.07

Thus it is anticipated that the QALYs gained through preventing a death attributable to hip fracture are underestimated by 11.12 in woman aged 50 years at treatment initiation and 3.33 in women aged 70 years.

The model did not record the number of mortalities attributable solely to hip fracture so this was estimated using the calculated risks of hip fracture over the 5-year treatment period and the assumed risk of mortality used in the individual patient model.

At 50 years of age there were expected to be 1.22 hip fractures per 100 women over the 5-year treatment period and a mortality rate of 2%, which results in an estimated 0.024 deaths attributable to hip fracture per woman. The corresponding values for individuals aged 70 years were 1.22 hip fractures per 100 women, a mortality rate of 6% and an estimated 0.118 deaths per 100 women.

The estimated numbers of deaths attributable to hip fracture were then multiplied by the estimated QALYs that were not gained due to the model being constrained to a 10-year time period (Table 1). Thus it was estimated that the model would underestimate the QALYs associated with treatment in the 50-54 year old age group by 0.272 (0.024 *11.12) per 100 women and 0.392 in women aged 70-74 years in the event where all hip fractures were prevented by the intervention. These values were then modified by the actual relative risk of fracture associated with each intervention. For example, if the relative risk of an intervention for hip fracture were 0.5, then the QALYs would be underestimated by 0.136 (0.272/2) per 100 women aged 50 years and 0.196 per 100 women aged 70 years.

When calculating the likely underestimation of QALYs women were assumed to be at the fracture risks associated with a T-Score of –2.5SD. These differ from those estimated by the WHO algorithm (which incorporated clinical risk factors). The underestimation in QALYs was assumed to be proportional to fracture risk. For example, if the WHO algorithm predicted that for a specific patient the fracture risks were twice that assumed for a woman with a T-Score of -2.5, then if the QALYs gained for a woman with a T-Score of -2.5SD were predicted to be underestimated by 0.13, an additional 0.26 QALYs would be added.

Description of how the model took account of the overestimation of QALYs associated with all mortality following hip fracture being assumed attributable to the fracture.

Data on the rates of mortality following hip fracture were taken from a UK audit [17], which recorded mortality at 90 days post hip fracture. These values are provided in Table 6 (p27) of the Strontium Ranelate Assessment Report [1]. When the model was run it was assumed that all deaths in the 90 days subsequent to a hip fracture were attributable to the fracture. However, subsequent data have shown this was incorrect, with only 25% directly related to the fracture, 42% possibly related and the remaining 33% not related [18].

Deaths may, however occur after 90 days, with this value appearing to be approximately 40% of that within the initial 90 days [18]. These data indicate that the estimate of mortality associated with hip fracture would lie between 35% (25%*1.4) and 94% (67%*1.4) of the value at 90 days. For the analyses undertaken for NICE a value of 50% was used. As the model assumed 100% of mortality was attributable to the hip fracture the QALYs gained through intervention will have been overestimated. Adjustments were undertaken to provide an indication of the likely error using the following methodology.

The estimated number of hip fractures within the treatment period was calculated and for simplicity it was assumed that all fractures occurred at year 3. The discounted QALYs lost due to a death in year 3 of treatment compared with a woman of average life expectancy were calculated and are shown in Table 2. Note that these are greater than the values in Table 1 which were calculated from the end of the 10-year modelling horizon, rather than 3 years within the modeling period.

Table 2. The estimated discounted QALYs that are lost when a women is assumed to have died from a hip fracture

Age (years)	Estimated discounted QALYs 10 years after treatment initiation	Estimated discounted QALYs 10 years after treatment initiation incorporating the likely long-term disutility of prior fracture
50	16.63	15.39
55	14.09	13.03
60	11.56	10.70
65	9.24	8.55
70	7.14	6.65
75	5.40	5.00

In the model, these values were then multiplied by 0.925 to take into consideration that women who did not die following a hip fracture are likely to be living with a previous fracture that would affect their long-term disutility. The value of 0.925 was chosen as a reasonable estimate of the average long-term disutility.

At 50 years of age there were expected to be 1.22 hip fractures per 100 women over the 5-year treatment period. Assuming that 50% of the mortality rate was attributable to the fracture rather than 100% the mortality rate would reduce from 2% to 1%, which results in an overestimation of mortality by 0.0122 (0.0244-0.0122) deaths per

100 women. This would equate to an overestimation of QALYs gained of 0.188 (15.39*0.0122) per 100 women. Using the same methodology for 70-year old women resulted in an expected overestimation of QALYs of 0.392.

When calculating the likely overestimation of QALYs women were assumed to be at the fracture risks associated with a T-Score of –2.5SD. These differ from those estimated by the WHO algorithm (which incorporated clinical risk factors) and the overestimation in QALYs was assumed to be proportional to the underlying fracture risk. For example, if the WHO algorithm predicted that for a specific patient the risks for fracture were twice that assumed for a woman with a T-Score of -2.5, then if the QALYs gained for a woman with a T-Score of -2.5SD were predicted to be overestimated by 0.24, an additional 0.48 QALYs would be subtracted for the woman evaluated within the WHO algorithm.

Description of how the model took account of combining the expected overestimation and underestimation of the QALYs gained

As detailed earlier, limitations within the original model have led to assumptions that have both under and overestimated the QALYs gained associated with hip fracture. These values were combined to produce an estimate of the likely underestimation of QALYs associated with hip fracture. These are provided in Table 3, with a negative number denoting that the QALYs gained have been overestimated.

Table 3. The expected underestimation of QALYs associated with hip fracture mortality per 100 women

Age (years)	The underestimation of QALYs associated with hip fracture mortality
50	0.084
55	0.042
60	0.117
65	0.063
70	0.000
75	-0.077

Women were assumed to have the risks associated with a T-Score of –2.5SD, which differ from those that would be produced by the WHO algorithm, which required adjustments to the calculated underestimation of QALYs. For example, if the QALYs gained for a woman with a T-Score of -2.5SD were predicted to be underestimated by 0.10, but that the WHO algorithm predicted that the risk of hip fracture was twice that assumed for a woman with a T-Score of –2.5SD, an additional 0.20 QALYs would be added.

Description of how the model took account of the effect of mortality following vertebral fracture

The original model developed before 2005 did not include any mortality attributable to vertebral fractures. However subsequently available data show that vertebral fracture can be associated with mortality, with an assumed increase in the mortality rate by a factor of 4.4 [19], even when only 28% of mortalities were assumed to be

attributable to the fracture [20]. The following describes how mortality following vertebral fracture was incorporated into the model for TA160/161.

Assuming that all fractures occurred in year 3 of treatment, the expected QALYs lost if a woman died from a vertebral fracture are equal to those if the woman dies following a hip fracture (Table 2). The absolute underestimation of QALYs is calculated by multiplying the expected number of vertebral fractures in the treatment period by the increased mortality risk and by the QALYs lost due to death. These values are provided in Table 4.

For illustration, at 50 years of age there were an expected 1.10 vertebral fractures per 100 women in the treatment period. The standard mortality rate was 0.24% [14] which would imply an increased mortality rate following fracture of 0.82% (0.24*3.3), of which 0.22% (0.82%*28%) would be attributable to the fracture. The expected number of mortalities attributable to vertebral fracture would thus be 0.0025 (1.10*0.22%) per 100 women. Given an expected QALY loss of 15.39 per mortality (Table 2) this equates to an expected QALY loss of 0.038 (0.0025*15.39) per 100 women.

Table 4. The expected underestimation of QALYs gained per 100 women through treatment due to the omission of mortality attributable to vertebral fracture in the model

Age (years)	Deaths estimated to be attributable to vertebral fracture (per 100 women)	Expected underestimation in the number of QALYs gained
50	0.002	0.038
55	0.007	0.089
60	0.009	0.097
65	0.021	0.184
70	0.051	0.339
75	0.079	0.397

Women were assumed to have the risks associated with a T-Score of –2.5SD, which differ from those that would be produced by the WHO algorithm, which required adjustments to the calculated underestimation of QALYs. For example, if the QALYs gained for a woman with a T-Score of -2.5SD were predicted to be underestimated by 0.15, but that the WHO algorithm predicted that the risk of hip fracture was twice that assumed for a woman with a T-Score of –2.5SD, an additional 0.30 QALYs would be added.

Description of how the model took account of the effect of mortality following proximal humerus fracture

The original model developed before 2005 did not include any mortality attributable to proximal humerus fractures. However subsequently available data show that proximal humerus fracture can be associated with mortality. It was conservatively assumed that the mortality rate would double in the year following a proximal humerus fracture [6] although only 28% of deaths were assumed attributable to the

fracture [20]. The following describes how mortality following proximal humerus fracture was incoroporated into the model for TA160/161.

An methodology identical to the one used for vertebral fracture mortality was used to calculate the expected underestimation in QALYs associated with not including mortality attributable to proximal humerus fractures within the model. The expected values are given in Table 5.

Table 5. The expected underestimation of QALYs gained per 100 women through treatment due to the omission of mortality attributable to proximal humerus fracture in the model

Age (years)	Deaths estimated to be attributable to proximal humerus fracture (per 100 women)	Expected underestimation in the number of QALYs gained
50	0.000	0.007
55	0.001	0.012
60	0.002	0.023
65	0.003	0.024
70	0.007	0.048
75	0.012	0.063

When calculating the likely underestimation of QALYs women were assumed to be at the fracture risks associated with a T-Score of –2.5SD. These differ from those estimated by the WHO algorithm (which incorporated clinical risk factors) and the underestimation in QALYs was assumed to be proportional to the underlying fracture risk. For example, if the WHO algorithm predicted that for a specific patient the risks for fracture were twice that assumed for a woman with a T-Score of -2.5, then if the QALYs gained for a woman with a T-Score of -2.5SD were predicted to be underestimated by 0.02, an additional 0.04 QALYs would be added for the woman evaluated within the WHO algorithm.

Sensitivity analysis 1: Exploring the impact of different mortality assumptions on the cost effectiveness of treatment strategies

In order to test the consultees' claim that mortality has been underestimated in the model, the DSU have carried out exploratory calculations to test the sensitivity of the model outputs to mortality assumptions.

In order to test the robustness of the results to changes in assumptions for mortality after the 10 year time horizon and for incorporating mortality associated with vertebral and proximal humerus fractures, the DSU multiplied the above basecase mortality values by 2.

Analyses were undertaken for generic alendronate and strontium ranelate. Estimated underestimation in QALYs gained are shown in Table 6. These sensitivity analyses were undertaken by doubling the values in B3:D3 of the boltons worksheets in the executable model.

Table 6: The maximum expected underestimation in QALYs gained (per 100 women) used in the sensitivity analyses

	The underestimation in QALYs used in the sensitivity analyses. (per 100 women)					
	Hip fracture		Vertebral fracture		Proximal Humerus fracture	
Age (years)	Risk	Doubled Risk	Risk	Doubled Risk	Risk	Doubled Risk
50	0.084	0.17	0.038	0.08	0.007	0.01
55	0.042	0.08	0.089	0.18	0.012	0.02
60	0.117	0.23	0.097	0.19	0.023	0.05
65	0.063	0.13	0.184	0.37	0.024	0.05
70	0.000	0.00	0.339	0.68	0.048	0.10
75	-0.077	-0.15	0.397	0.79	0.063	0.13

It is stressed that these values are the maximum QALYs that would be obtained if an intervention eliminated the possibility of a fracture, i.e. having a relative risk of fracture of 0. Interventions that had a relative risk of 0.8 would only accrue 20% of the values given in Table 6.

The impact of these sensitivity analyses on the modelling outputs are presented in Table 7 for self-identifying women and Table 8 for women who would be opportunistically assessed assuming that the underestimation in QALYs should be double that used in the base-case.

For alendronate, the age and number of clinical risk factors that are required for BMD testing and subsequent treatment to be cost effective are presented, for strontium ranelate the T-Score threshold at which the treatment becomes cost effective (assuming a cost per QALY threshold of £30,000 per QALY for self-identifying women and £20,000 per QALY for women without a fracture) is presented.

Table 7. Sensitivity analyses 1: The effect of doubling the base case mortality values for mortality after the 10 year time horizon and for incorporating mortality associated with vertebral and proximal humerus fractures - self-identifying women

Age (years)	No of CRF	Base Case Mortality Assumptions		Base Case Mortality Assumptions *2	
		Alendronate	Strontium Ranelate	Alendronate	Strontium Ranelate
50-54	0	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	1	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	2	BMD and treat with T-Score < -2.0SD	T-Score < -3.5SD	BMD and treat with T-Score < -2.0SD	T-Score < -3.5SD
	3	BMD and treat with T-Score < -1.5SD	T-Score < -3.0SD	BMD and treat with T-Score < -1.0SD	T-Score < -3.0SD
55-59	0	BMD and treat with T-Score < -2.5SD	T-Score < -4.0SD	BMD and treat with T-Score < -2.5SD	T-Score < -4.0SD
	1	BMD and treat with T-Score < -2.0SD	T-Score < -3.5SD	BMD and treat with T-Score < -2.0SD	T-Score < -3.5SD
	2	BMD and treat with T-Score < -1.5SD	T-Score < -3.5SD	BMD and treat with T-Score < -1.5SD	T-Score < -3.5SD
	3	BMD and treat with T-Score < -1.5SD	T-Score < -3.0SD	BMD and treat with T-Score < -1.0SD	T-Score < -3.0SD
60-64	0	BMD and treat with T-Score < -2.5SD	T-Score < -4.0SD	BMD and treat with T-Score < -2.0SD	T-Score < -4.0SD
	1	BMD and treat with T-Score < -2.0SD	T-Score < -3.5SD	BMD and treat with T-Score < -2.0SD	T-Score < -3.5SD
	2	BMD and treat with T-Score < -1.5SD	T-Score < -3.5SD	BMD and treat with T-Score < -1.5SD	T-Score < -3.0SD
	3	BMD and treat with T-Score < -1.0SD	T-Score < -3.0SD	BMD and treat with T-Score < -1.0SD	T-Score < -3.0SD
65-69	0	BMD and treat with T-Score < -2.0SD	T-Score < -4.0SD	BMD and treat with T-Score < -2.0SD	T-Score < -3.5SD
	1	BMD and treat with T-Score < -1.5SD	T-Score < -3.5SD	BMD and treat with T-Score < -1.5SD	T-Score < -3.5SD
	2	BMD and treat with T-Score < -1.0SD	T-Score < -3.0SD	BMD and treat with T-Score < -1.0SD	T-Score < -3.0SD
	3	BMD and treat with T-Score < -0.5SD	T-Score < -2.5SD	BMD and treat with T-Score < 0.0SD	T-Score < -2.5SD
70-74	0	BMD and treat with T-Score < -1.0SD	T-Score < -3.0SD	BMD and treat with T-Score < -0.5SD	T-Score < -3.0SD
	1	BMD and treat with T-Score < -0.5SD	T-Score < -3.0SD	BMD and treat with T-Score < 0.5SD	T-Score < -2.5SD
	2	BMD and treat with T-Score < 0.5SD	T-Score < -2.0SD	BMD and treat with T-Score < 1.0SD	T-Score < -2.0SD
	3	BMD and treat with T-Score < 1.0SD	T-Score < -2.0SD	BMD and treat with T-Score < 1.0SD	T-Score < -1.5SD
75 and older	0	BMD and treat with T-Score < 0.0SD	T-Score < -3.0SD	BMD and treat with T-Score < 0.0SD	T-Score < -2.5SD
	1	BMD and treat with T-Score < 0.5SD	T-Score < -2.5SD	BMD and treat with T-Score < 1.0SD	T-Score < -2.5SD
	2	BMD and treat with T-Score < 1.0SD	T-Score < -2.0SD	BMD and treat with T-Score < 1.0SD	T-Score < -1.5SD
	3	BMD and treat with T-Score < 1.0SD	T-Score < -1.0SD	BMD and treat with T-Score < 1.0SD	T-Score < -1.0SD

Table 8 Sensitivity analyses 1: The effect of doubling the base case mortality values for mortality after the 10 year time horizon and for incorporating mortality associated with vertebral and proximal humerus fractures - opportunistically assessed women

Age (years)	No of CRF	Base Case Mortality Assumptions		Base Case Mortality Assumptions *2	
		Alendronate	Strontium Ranelate	Alendronate	Strontium Ranelate
50-54	0	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	1	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	2	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	3	Do not BMD	Do not BMD	Do not BMD	Do not BMD
55-59	0	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	1	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	2	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	3	Do not BMD	Do not BMD	Do not BMD	Do not BMD
60-64	0	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	1	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	2	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	3	Do not BMD	Do not BMD	Do not BMD	Do not BMD
65-69	0	Do not BMD	Do Not BMD	Do not BMD	Do Not BMD
	1	BMD and treat with T-Score < -2.5SD	T-Score < -4.5SD	BMD and treat with T-Score < -2.5SD	T-Score < -4.5SD
	2	BMD and treat with T-Score < -2.0SD	T-Score < -4.0SD	BMD and treat with T-Score < -2.0SD	T-Score < -4.0SD
	3	BMD and treat with T-Score < -1.5SD	T-Score < -3.5SD	BMD and treat with T-Score < -1.5SD	T-Score < -3.5SD
70-74	0	BMD and treat with T-Score < -2.0SD	T-Score < -4.5SD	BMD and treat with T-Score < -2.0SD	T-Score < -4.5SD
	1	BMD and treat with T-Score < -2.0SD	T-Score < -4.0SD	BMD and treat with T-Score < -2.0SD	T-Score < -4.0SD
	2	BMD and treat with T-Score < -1.0SD	T-Score < -3.5SD	BMD and treat with T-Score < -1.0SD	T-Score < -3.5SD
	3	BMD and treat with T-Score < -0.5SD	T-Score < -3.0SD	BMD and treat with T-Score < -0.5SD	T-Score < -3.0SD
75 and older	0	BMD and treat with T-Score < -1.5SD	T-Score < -4.0SD	BMD and treat with T-Score < -1.5SD	T-Score < -4.0SD
	1	BMD and treat with T-Score < -1.0SD	T-Score < -4.0SD	BMD and treat with T-Score < -1.0SD	T-Score < -3.5SD
	2	BMD and treat with T-Score < -0.5SD	T-Score < -3.0SD	BMD and treat with T-Score < 0.0SD	T-Score < -3.0SD
	3	BMD and treat with T-Score < 0.5SD	T-Score < -2.5SD	BMD and treat with T-Score < 0.5SD	T-Score < -2.5SD

Sensitivity analysis 2: Exploring the impact of different mortality assumptions on the cost effectiveness of treatment strategies

The DSU undertook a further analysis to assess the impact of changing the percentage of deaths that are assumed attributable to hip fracture. In the base case a value of 50% was used, and this value was set to 100% for the current sensitivity analyses. Analyses were undertaken for generic alendronate and strontium ranelate. A summary of the results is provided in Table 9 together with the value were all mortality following hip fracture attributable to the fracture. These sensitivity analyses were conducted by altering cell C3 of the boltons worksheet in the executable model.

Table 9: QALYs gained (per 100 women) used in the sensitivity analyses on % of mortality attributable to hip fracture

	Difference in QALYs gained depending on % of mortality attributable to hip fracture	
Age (years)	50% of mortality attributable to hip fracture	100% of mortality attributable to hip fracture
50	0.084	0.272
55	0.042	0.212
60	0.117	0.568
65	0.063	0.479
70	0.000	0.392
75	-0.077	0.374

It is stressed that these values are the maximum QALYs that would be obtained if an intervention eliminated the possibility of a fracture, i.e. having a relative risk of fracture of 0. Interventions that had a relative risk of 0.8 would only accrue 20% of the values given in Table 9.

Results from the sensitivity analyses are presented in Table 10 for self-identifying women and Table 11 for women who would be opportunistically assessed, assuming 50 or 100% of mortality attributable to hip fracture.

For alendronate, the age and number of clinical risk factors that are required for BMD testing and subsequent treatment to be cost effective are presented, for strontium ranelate the T-Score threshold at which the treatment becomes cost effective (assuming a cost per QALY threshold of £30,000 per QALY for self-identifying women and £20,000 per QALY for women without a fracture) is presented.

Table 10. Sensitivity analyses on the percentage of mortality attributable to hip fracture - self-identifying women

Age (years)	No of CRF	Base Case Assumption on mortality attributable to hip fracture (50%)		Alternative Assumption on mortality attributable to hip fracture (100%)	
		Alendronate	Strontium Ranelate	Alendronate	Strontium Ranelate
50-54	0	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	1	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	2	BMD and treat with T-Score < -2.0SD	T-Score < -3.5SD	BMD and treat with T-Score < -2.0SD	T-Score < -3.5SD
	3	BMD and treat with T-Score < -1.5SD	T-Score < -3.0SD	BMD and treat with T-Score < -1.5SD	T-Score < -3.0SD
55-59	0	BMD and treat with T-Score < -2.5SD	T-Score < -4.0SD	BMD and treat with T-Score < -2.5SD	T-Score < -4.0SD
	1	BMD and treat with T-Score < -2.0SD	T-Score < -3.5SD	BMD and treat with T-Score < -2.0SD	T-Score < -3.5SD
	2	BMD and treat with T-Score < -1.5SD	T-Score < -3.5SD	BMD and treat with T-Score < -1.5SD	T-Score < -3.5SD
	3	BMD and treat with T-Score < -1.5SD	T-Score < -3.0SD	BMD and treat with T-Score < -1.5SD	T-Score < -3.0SD
60-64	0	BMD and treat with T-Score < -2.5SD	T-Score < -4.0SD	BMD and treat with T-Score < -2.0SD	T-Score < -3.5SD
	1	BMD and treat with T-Score < -2.0SD	T-Score < -3.5SD	BMD and treat with T-Score < -2.0SD	T-Score < -3.5SD
	2	BMD and treat with T-Score < -1.5SD	T-Score < -3.5SD	BMD and treat with T-Score < -1.5SD	T-Score < -3.0SD
	3	BMD and treat with T-Score < -1.0SD	T-Score < -3.0SD	BMD and treat with T-Score < -1.0SD	T-Score < -2.5SD
65-69	0	BMD and treat with T-Score < -2.0SD	T-Score < -4.0SD	BMD and treat with T-Score < -2.0SD	T-Score < -3.5SD
	1	BMD and treat with T-Score < -1.5SD	T-Score < -3.5SD	BMD and treat with T-Score < -1.5SD	T-Score < -3.5SD
	2	BMD and treat with T-Score < -1.0SD	T-Score < -3.0SD	BMD and treat with T-Score < -1.0SD	T-Score < -3.0SD
	3	BMD and treat with T-Score < -0.5SD	T-Score < -2.5SD	BMD and treat with T-Score < -0.5SD	T-Score < -2.5SD
70-74	0	BMD and treat with T-Score < -1.0SD	T-Score < -3.0SD	BMD and treat with T-Score < -0.5SD	T-Score < -3.0SD
	1	BMD and treat with T-Score < -0.5SD	T-Score < -3.0SD	BMD and treat with T-Score < 0.0SD	T-Score < -3.0SD
	2	BMD and treat with T-Score < 0.5SD	T-Score < -2.0SD	BMD and treat with T-Score < 0.5SD	T-Score < -2.0SD
	3	BMD and treat with T-Score < 1.0SD	T-Score < -2.0SD	BMD and treat with T-Score < 1.0SD	T-Score < -1.5SD
75 and older	0	BMD and treat with T-Score < 0.0SD	T-Score < -3.0SD	BMD and treat with T-Score < 0.0SD	T-Score < -3.0SD
	1	BMD and treat with T-Score < 0.5SD	T-Score < -2.5SD	BMD and treat with T-Score < 0.5SD	T-Score < -2.5SD
	2	BMD and treat with T-Score < 1.0SD	T-Score < -2.0SD	BMD and treat with T-Score < 1.0SD	T-Score < -1.5SD
	3	BMD and treat with T-Score < 1.0SD	T-Score < -1.0SD	BMD and treat with T-Score < 1.0SD	T-Score < -1.0SD

Table 11. Sensitivity analyses on the percentage of mortality attributable to hip fracture - opportunistically assessed women

Age (years)	No of CRF	Base Case Mortality Assumptions		Base Case Mortality Assumptions *2	
		Alendronate	Strontium Ranelate	Alendronate	Strontium Ranelate
50-54	0	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	1	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	2	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	3	Do not BMD	Do not BMD	Do not BMD	Do not BMD
55-59	0	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	1	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	2	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	3	Do not BMD	Do not BMD	Do not BMD	Do not BMD
60-64	0	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	1	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	2	Do not BMD	Do not BMD	Do not BMD	Do not BMD
	3	Do not BMD	Do not BMD	Do not BMD	Do not BMD
65-69	0	Do not BMD	Do Not BMD	Do not BMD	Do Not BMD
	1	BMD and treat with T-Score < -2.5SD	T-Score < -4.5SD	BMD and treat with T-Score < -2.5SD	T-Score < -4.5SD
	2	BMD and treat with T-Score < -2.0SD	T-Score < -4.0SD	BMD and treat with T-Score < -2.0SD	T-Score < -4.0SD
	3	BMD and treat with T-Score < -1.5SD	T-Score < -3.5SD	BMD and treat with T-Score < -1.5SD	T-Score < -3.5SD
70-74	0	BMD and treat with T-Score < -2.0SD	T-Score < -4.5SD	BMD and treat with T-Score < -2.0SD	T-Score < -4.5SD
	1	BMD and treat with T-Score < -2.0SD	T-Score < -4.0SD	BMD and treat with T-Score < -2.0SD	T-Score < -4.0SD
	2	BMD and treat with T-Score < -1.0SD	T-Score < -3.5SD	BMD and treat with T-Score < -1.0SD	T-Score < -3.5SD
	3	BMD and treat with T-Score < -0.5SD	T-Score < -3.0SD	BMD and treat with T-Score < -0.5SD	T-Score < -3.0SD
75 and older	0	BMD and treat with T-Score < -1.5SD	T-Score < -4.0SD	BMD and treat with T-Score < -1.5SD	T-Score < -4.0SD
	1	BMD and treat with T-Score < -1.0SD	T-Score < -4.0SD	BMD and treat with T-Score < -1.0SD	T-Score < -3.5SD
	2	BMD and treat with T-Score < -0.5SD	T-Score < -3.0SD	BMD and treat with T-Score < 0.0SD	T-Score < -3.0SD
	3	BMD and treat with T-Score < 0.5SD	T-Score < -2.5SD	BMD and treat with T-Score < 0.5SD	T-Score < -2.5SD

Summary of the sensitivity of the results to the changes in mortality assumptions

It is seen that there are only a small number of T-Score thresholds that would change (by 0.5SD in favour of intervention) when the values used in the base case for the underestimation of QALYs gained are doubled (Tables 7 and 8), and that there are only a small number of changes when 100% of all mortality following hip fracture is assumed attributable to the fracture (Table 10). These results show that the results of the model are not sensitive to these assumptions.

ANNEX 4

The distribution of T-Scores (at the femoral neck) in women aged 50-80 as reported in the database of patients reported in the Holt *et al* paper [10] that was provided to the Assessment Group.

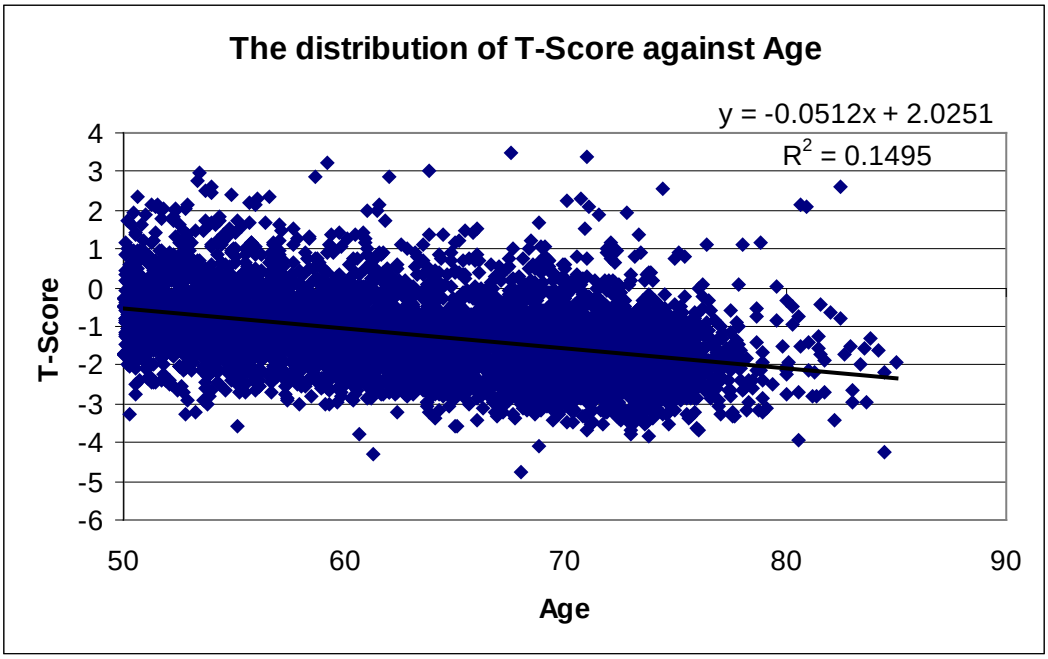
Consultees commented on the use of the Holt *et al* data [10] as a basis for the population data used in this appraisal. The DSU confirm that this data has been described in all appraisal documents since December 2003, including the 2005 Strontium Ranelate Assessment Report (pp16-17) [1], and that these were the most appropriate data to use in a UK context. The DSU review confirms that the data in the model is consistent with the data underpinning the Holt *et al* paper. [10].

The comments also focus on the distribution of BMD T-scores in the model compared with data underpinning the publication by Holt *et al* [10]. The DSU have reviewed the T-Score values used in the model with those in the database underpinning the Holt *et al* paper [10] provided to the Assessment Group by the data owners in 2004, and have undertaken the following additional analyses using this data [10] to test the assumption of normality assumed for T-Score distribution.

In order to be consistent with the WHO algorithm, it was necessary to use the T-Score at the femoral neck. The data in Figure 1 were taken from the database provided to the ScHARR assessment team that was used in the Holt *et al* publication [10] by the owners of the data. It is seen that there is a negative relationship between age and T-Score. There are a small number of outliers with high T-Scores, however, ignoring these, the variance around the linear regression fit appears to remain relatively stable across all ages.

The linear regression was used in the model as it utilised the full Holt *et al* dataset [10[and there was no reason to believe that there would be marked differences between the age bands (for instance a 64 year old would have a similar T-Score on average to a 65 year old).

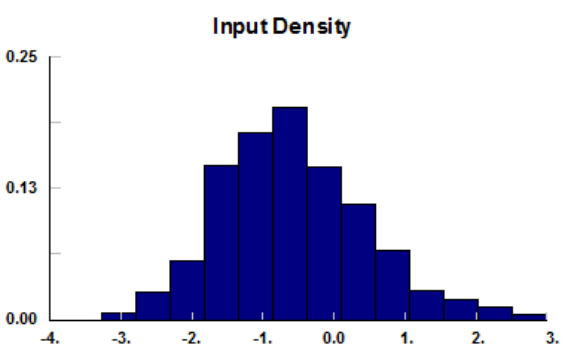
Figure 1. The distribution of T-Score against age



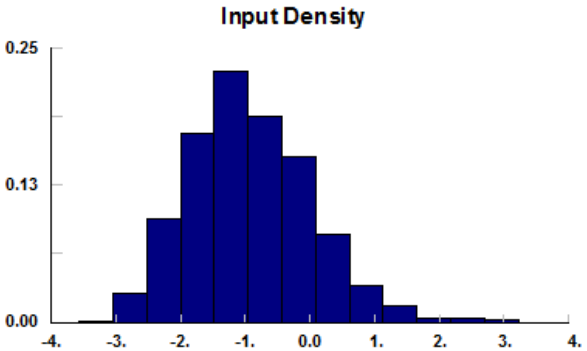
The T-Scores for women within each age band used in the model were graphed. These are shown in Figure 2.

Figure 2. The distribution of T-Score by Age band.

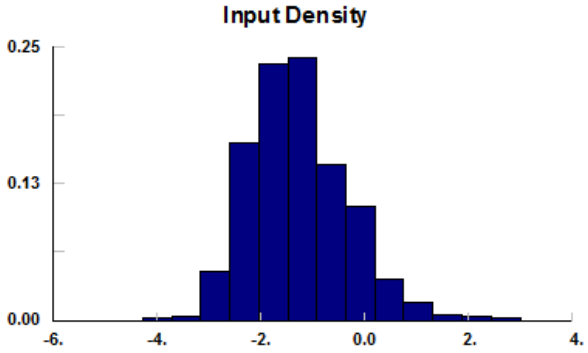
In all graphs the x-axis is the T-Score value, whilst the y-axis is the relative frequency within defined T-Score bands.



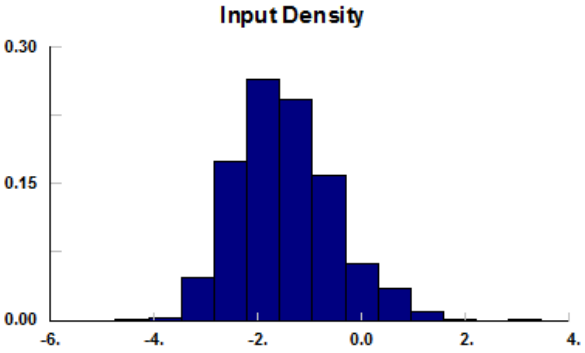
Age 50-54yrs



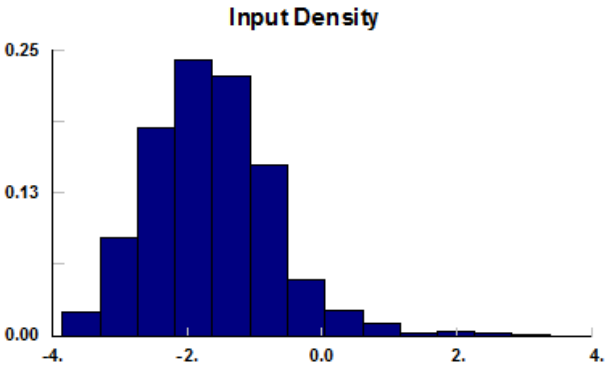
Age 55-59yrs



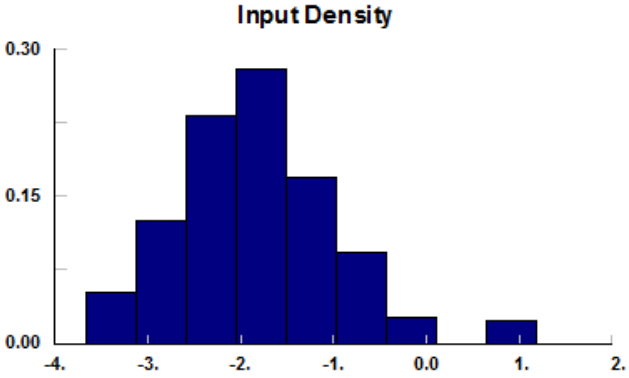
Age 60-64



Age 65-69



Age 70-74



Age 75-79

Tests for normality were conducted on these data. These analyses are provided in Table 1.

Table 1: Testing the assumption of normality that was used in the model for the distribution of BMD values within each age band

Age Band	Mean Age within the group (years)	Best fitting normal distribution (mean, SD)	Could the hypothesis that the data came from this normal distribution be rejected?	The normal distribution used within the model assuming mid point age.
50-54	52.5	-0.58, 1.020	No	-0.66, 1.000
55-59	57.4	-0.93, 0.945	Yes	-0.92, 1.000
60-64	62.3	-1.27, 0.940	Yes	-1.17, 1.000
65-69	67.5	-1.49, 0.934	Yes	-1.43, 1.000
70-74	71.8	-1.66, 0.928	No	-1.69, 1.000
75-79	76.4	-1.81, 0.845	No	-1.94, 1.000

A normal distribution appeared appropriate for ages 50-54 years and 70-79 years. In this latter group the mean age was lower in the Holt *et al* database (71.8 years and 76.8 years compared with 72.5 years and 77.5 years) than that used within the model. It is thus not surprising that the T-Score distributions used in the model have lower mean T-Scores (-1.69 and –2.94 SD compared with –1.66 and –1.81SD). The standard deviation is slightly wider in the distribution used within the model. This is thought to be favourable to an intervention as within the model there will be more women with lower T-Scores and it is in these women that treatment is more cost effective. This is generally not counterbalanced by the greater number of women with higher T-Scores, as at younger ages women with below average T-Scores are unlikely to be able to be treated cost-effectively. For women aged between 50 and 54 years it appears that using the mean from the linear regression may be favourable to treatment at 50 years of age compared with using the raw data.

For ages 55-69 years the normal distribution did not fit the data primarily because of outliers with high T-Scores. For each of these distributions a lognormal distribution was fitted as detailed in Table 2. In all cases the lognormal distribution selected could not be statistically rejected as being the source of the data.

Table 2. The best fitting lognormal distribution for the T-Score data for women aged 55-69 years

Age Band (years)	Best fitting lognormal distribution (minimum value, mean (on a log scale), SD (on a log scale))
55-59	-6.73, 1.74, 0.162
60-64	-6.49, 1.64, 0.177
65-69	-7.71, 1.82, 0.149

Exploratory calculations were undertaken to compare the number of women in each T-Score band from the assumed normal distribution used in the model and from the lognormal distributions (Table 3). In all cases the normal distribution used appears to be more favourable to interventions than the lognormal distributions as the normal distribution estimates a greater proportion of women to be osteoporotic. The number of women estimated to be osteoporotic within an age band was never estimated to be greater than 15% regardless of methodology used.

Table 3. The distribution between T-Score band between the normal and lognormal distributions for osteoporotic women only.

		T-Score						
Age Band (years)	Distn	<-5.0	-5.0 to -4.5	-4.5 to -4.0	-4.0 to -3.5	-3.5 to -3.0	-3.0 to -2.5	> -2.5
55 -59	Norm	0.00%	0.01%	0.09%	0.39%	1.38%	3.83%	94.29%
	LN	0.00%	0.00%	0.00%	0.02%	0.42%	2.85%	96.70%
60 – 64	Norm	0.01%	0.04%	0.19%	0.76%	2.37%	5.81%	90.82%
	LN	0.00%	0.00%	0.00%	0.10%	1.27%	6.01%	92.61%
65 -69	Norm	0.02%	0.09%	0.40%	1.41%	3.90%	8.41%	85.77%
	LN	0.00%	0.00%	0.03%	0.48%	2.97%	9.29%	87.22%

NB: Norm denotes the normal distribution used in the modeling; LN denotes the lognormal fit to the raw data

In conclusion, these exploratory analyses show that any assumptions about normality and a standard deviation of 1, as used in the model, are likely to be favourable to the treatments appraised.

The relationship between T-Score and CRF

Consultees have commented that the distribution of BMD values according to the number of clinical risk factors was not captured correctly in the model.

DSU has investigated whether any relationship between T-score and clinical risk factor would affect the results of the model. In order to do so, the DSU have carried out an exploratory analysis based on the raw data underpinning the WHO algorithm provided to the Assessment Group.

Data used to populate the WHO algorithm which was provided as academic-in-confidence are shown in Figure 3. It is commented that in general these T-Scores are lower than those for the UK population contained within the Holt *et al* dataset [10], which was used to populate the model. However the relative difference between women with and without a CRF was assumed applicable to the UK context in this sensitivity analysis.

Figure 3. The relationship between T-Score and CRF
[confidential]

Figure 4. The average difference in T-Score between a woman with no CRFs and a similar woman with a single CRF
[confidential]

Figure 5. Comparing the T-Score between women with 1 or multiple CRFs
[confidential]

Additional sensitivity analyses

[redacted]

Table 4. The distribution amongst T-Score bands were it assumed that women with a clinical risk factor had a T-Score 0.12SD lower than the average.
[confidential]

Impact of the sensitivity analyses

The additional exploratory analyses carried out to respond to the consultee's comments on the model showed that the average T-Score for women with a CRF is only slightly lower than for a similar woman with no CRF. For both women with and without fracture, the results of the model did not change when the T-Score of women with a CRF was assumed to be [redacted] lower than the average. This is not surprising, as this reduction does not affect whether treatment is cost effective for a certain T-Score band, but only the number of women within each band. The change in distribution (provided in Table 4) was not large enough to result in any scenarios previously judged not to be cost effective (due to the costs of BMD scanning and questioning time for those women opportunistically assessed) becoming cost-effective under the new assumptions.

The DSU notes that the number of women receiving treatment would rise, however due to the change in T-Score distribution explored in the sensitivity analyses, the criteria for receiving treatment does not change.

ANNEX 5

5.1. LETTER SUBMITTED TO OSTEOPOROSIS INTERNATIONAL AND SUBSEQUENT DECISION

Evidence that the structure of the 'NICE' model does not produce substantially different answers from that of a recently published model.

Dear Sir,

I write in response to the recent editorial provided by Kanis and Compston [21]. As previously stated [8], the results of the 'NICE' model and that of Kanis *et al* [9] are similar when the same assumptions and parameters are used. However, the validity of this remark was questioned in the editorial [21], as the numerical evidence has not been published. In response, I detail the results produced from the 'NICE' model using the assumptions and parameter values from Kanis *et al*. [9].

These results are presented as the cost per QALY ratios for alendronate treatment in women aged 50, 55, 60, 65, 70 or 75 years with a T-Score of –2.5 SD. Simulated women are further divided into those without a prior fracture and those with a prior fracture. For the 'NICE' model the results were also presented in relation to the number of clinical risk factors possessed by the woman in addition to a previous fracture, where appropriate. Clinical risk factors were defined as parental history of hip fracture, exposure to systemic glucocorticoids, current smoking, high intake of alcohol (>2 units daily on average) and the presence of rheumatoid arthritis.

The presence of clinical risk factors increases the risk of subsequent fracture and hence improves the cost-effectiveness of the intervention. As the majority of women have no other additional risk factors (Prof Kanis. Personal Communication) we would expect the results produced by Kanis *et al* [9], which are not reported by risk factors, to fall between the values for 0 and 1 additional clinical risk factors, were the models identical. This occurs on 9 of 12 of the age and prior fracture status combinations. On the three occasions where this does not happen, the 'NICE' model is more favourable to treatment on two occasions and unfavourable on one occasion, indicating no systematic bias. We conclude that the models produce similar cost-effectiveness results when populated with the same data and using similar assumptions. It is also noted that the adaptations made to the model to allow for effects beyond the initial 10-year time horizon, as previously reported [7] appear appropriate.

Table 1: The cost per QALY (£'000) results for women at a T-Score of –2.5 SD without a previous fracture

		'NICE' Model using the same assumptions of Kanis *et al*.[9]	
Age	Kanis *et al.* [9]	0 clinical risk factors	1 additional clinical risk factor.
50	14.7	26.0	15.7
55	16.2	21.0	11.8
60	14.3	17.7	10.0
65	7.0	14.0	6.5
70	3.7	6.1	1.3
75	3.0	1.7	Dominated

Dominated means producing more QALYs for a lower cost than the comparator

Table 2: The cost per QALY (£'000) results for women at a T-Score of –2.5 SD with a previous fracture

		'NICE' Model using the same assumptions of Kanis *et al.* [9]	
Age	Kanis *et al.* [9]	0 clinical risk factors	1 additional clinical risk factor.
50	6.7	8.5	2.8
55	7.3	7.4	2.5
60	7.3	6.6	1.9
65	2.9	5.0	0.0
70	0.8	1.4	Dominated
75	Dominated	Dominated	Dominated

Dominated means producing more QALYs for a lower cost than the comparator

The guidance produced by NICE [22] and that which would be inferred from Kanis *et al.* [9] are clearly different. However, it is seen that the modelling structure is not a key driver of this dispute, as is implied within the editorial [21]. Instead, as previously stated [8], these disparities are caused by the different assumptions and parameter values used by NICE and Kanis *et al*.[9].

Yours sincerely

Dr Matt Stevenson

5.2. RESPONSE FROM OSTEOPOROSIS INTERNATIONAL

Manuscript ID: OI-2008-06-0274
Title: The structure of the 'NICE' model does not produce substantially different answers from that of a recently published model.

Dear Dr. Stevenson

Thank you for submitting your Letter to the Editor to Osteoporosis International.

It has been reviewed by all our four associate editors. Based on their opinion, I am sorry to inform you that we will not be able to accept your letter for publication but thank you for your interest in Osteoporosis International.

With kind regards.

Yours sincerely,
Fina Liu, PhD
Managing Editor, Osteoporosis International