Published in final edited form as:

Pharmacoeconomics. 2006 ; 24(12): 1181–1197.

Assessing Generalisability in Model-Based Economic Evaluation Studies: A Structured Review in Osteoporosis

Hege Urdahl¹, Andrea Manca^{2,*}, and Mark J Sculpher²

¹ Merck Sharp and Dohme.

² Centre for Health Economics, University of York, UK

Abstract

Background—To support decision making many countries have now introduced some formal assessment process to evaluate whether health technologies represent good ‘value for money’. These often take the form of decision models which can be used to explore elements of importance to generalisability of study results across clinical settings and jurisdictions. The objectives of the present review were to assess: (i) whether the published studies clearly defined the decision-making audience for the model; (ii) the transparency of the reporting in terms of study question, structure and data inputs; (iii) the relevance of the data inputs used in the model to the stated decision-maker or jurisdiction; and (iv) how fully the robustness of the model's results to variation in data inputs between locations was assessed.

Methods—Articles reporting decision-analytic models in the area of osteoporosis were assessed to establish the extent to which the information provided enabled decision makers in different countries/jurisdictions to fully appreciate the variability of results according to location, and the relevance to their own.

Results—Of the 18 articles included in the review, only three explicitly stated the decision-making audience. It was not possible to infer a decision-making audience in eight studies. Target population was well reported, as was resource and cost data, and clinical data used for estimates of relative risk reduction. However, baseline risk was rarely adapted to the relevant jurisdiction, and when no decision-maker was explicit it was difficult to assess whether the reported cost and resource use data was in fact relevant. A few studies used sensitivity analysis to explore elements of generalisability, such as compliance rates and baseline fracture risk rates, although such analyses were generally restricted to evaluating parameter uncertainty.

Conclusion—This review found that variability in cost-effectiveness across locations is addressed to a varying extent in modelling studies in the field of osteoporosis, limiting their use for decision-makers across different locations. Transparency of reporting is expected to increase as methodology develops, and decision-makers publish “reference case” type guidance.

Corresponding author: Dr Andrea Manca, Centre for Health Economics, University of York, Alcuin A, Heslington, York, YO10 5DD, UK. Email: am126@york.ac.uk; Tel. +441904321430; Fax: +441904321402..

Author responsibility: HU conducted the literature searches, carried out the data abstraction and wrote the first draft of manuscript; AM provided guidance on the appraisal methods, and prepared the manuscript for submission. MS was the Principal Investigator of the original NHS HTA R&D funded project on generalisability and contributed to the final version of the manuscript.

Financial disclosure: HU is currently employee of pharmaceutical company Merck Sharp and Dohme which has no financial interests in the outcome of the paper. HU was PhD student at the University of York when most of the work included in this paper was undertaken. AM is currently recipient of a Wellcome Trust Training Fellowship in Health Services Research. MS is Public Health Career Scientist funded by the UK NHS R&D Programme.

1. Introduction

Healthcare systems are increasingly under financial pressure to optimise the use and allocation of the available resources. To support decision making many countries have now introduced some formal assessment process to evaluate whether health technologies represent good 'value for money' [1]. Decision analytic modelling is widely used in health technology assessment to evaluate the effectiveness and cost-effectiveness of alternative options under conditions of uncertainty [2, 3]. This form of modelling is necessary in situations where a single primary source of data (e.g. a randomised trial) does not wholly satisfy the data needs of a decision problem, and additional data sources and assumptions are needed. These situations include synthesising information when a number of estimates of a particular parameter exist, extending the results of a short-term trial over a long-term time horizon, increasing the range of alternative treatment strategies being compared and adapting the results of a study undertaken in one location to be relevant to another jurisdiction [4]. It is the latter that this paper is concerned with.

The use of decision models to explore issues of variability and generalisability in cost-effectiveness between locations is of interest for two reasons. Firstly, models are often used to make the results of a particular patient-level study, such as a randomised trial, which was undertaken in one location relevant to one or more alternative locations. Secondly, geographical variability can generally be dealt with in published modelling studies, since the general purpose of models is to identify optimal solutions to specific decision problems faced by particular decision makers or jurisdictions. For example, the National Institute for Health and Clinical Excellence's (NICE) technology appraisal process often includes decision models submitted by manufacturers and/or developed by the academic Technology Assessment Team (see www.nice.org.uk). In each case, the decision maker (the NICE Appraisal Committee) and the jurisdiction (NHS) are explicit. An important feature of economic models is that they relate to the policy maker(s) and jurisdiction(s) whose decision the model is designed to inform. In published journal articles these are not always explicitly stated, in spite of such information being paramount to an assessment of the generalisability of study results to other decision makers. Therefore, it is of value to review published modelling studies to assess how clearly the authors have identified the decision making audience of their work, the extent to which the data incorporated into these models are the most appropriate for that decision maker and the degree to which they have assessed the importance of any variability between locations *within* a particular jurisdiction.

In recent years, a significant contributions in the area of generalisability has been published, much of which are in the statistical area of cost-effectiveness analysis.[5-13] There has been little evidence in the area of conceptual analysis and this article attempts to fill this gap. Rather than select a sample of model-based studies across a range of diseases and interventions, the particular clinical area of osteoporosis has been selected for the review. This clinical area has been selected for two reasons. Firstly, a large number of cost-effectiveness models are available since the majority of cost-effectiveness studies in osteoporosis are model-based. In a review of economic evaluations in this field undertaken in 1998, a total of 16 studies were identified, all of which were based on decision analytic models [14]. Secondly, there is likely to be pronounced variability between countries and other locations in many of the parameters going into these models. This variability potentially relates to baseline event rates, resource use, cost and utility [15].

Osteoporosis is characterised by low bone mineral density (BMD) and deterioration of bone tissue, leading to enhanced bone fragility and consequent increase in fracture risk. The disease is manifested in terms of a high occurrence of hip, wrist and vertebral fractures, and is most prominent in post-menopausal women. Development of fractures is a complex

function of osteoporosis, age and other risk factors which evolve over time. The majority of trials in this area have evaluated the impact of osteoporosis treatments on the intermediate endpoint of BMD [16]. Historically, there has been a scarcity of literature evaluating final endpoints [17]. The need to estimate links between intermediate end-points and ultimate measures of health gain, together with the requirement to synthesise results from several studies and extrapolate over a long-term time horizon, are the main reasons for the preponderance of decision models in osteoporosis.

2. Methods

2.1 Aims and objectives

The aim of the review was to assess how published model-based economic evaluations in the field of osteoporosis deal with variability in results by location and decision-maker.

The review set out to address the following specific objectives:

1. To assess whether the study clearly defines the decision-making audience for the model.
2. To establish whether the model was transparently reported in terms of study question, structure and data inputs.
3. To assess the relevance of the data inputs used in the model to the stated decision-maker or jurisdiction.
4. To assess how fully the robustness of the model's results to variation in data inputs between locations was assessed.

2.2. Inclusion and exclusion criteria

The inclusion criteria were full economic evaluation models evaluating therapeutic interventions in osteoporosis. Only studies reporting a summary measure of cost-effectiveness (e.g. cost-effectiveness ratio) were included in the review, as these studies combine an estimate of both costs and effectiveness and present an overall assessment of the value for money of the alternative interventions of interest to decision-makers. Since this was a review where methodology rather than results was of primary interest, only studies published in English were included. Economic evaluations which did not describe the structure and assumptions underlying a model were excluded from the assessment, as were simple cost analyses, secondary reviews of economic evaluation models and studies that did not present a summary measure of cost-effectiveness.

2.3 Search strategy

The identification of articles for this review was undertaken in 2001, as part of a NHS HTA R&D funded project on generalisability which commenced at that time. A search strategy was devised to retrieve published papers reporting economic evaluation models of interventions for osteoporosis. The searches were conducted using the following bibliographic databases: MEDLINE, EMBASE, EconLit, Health Economic Evaluation Database (HEED), the internal catalogue of the NHS CRD/CHE information service, the Health Technology Assessment (HTA) Database and the administration version of the NHS Economic Evaluation Database (NHS EED) held at the NHS CRD. The databases were searched from 1980 until 2001, or from the earliest publication in the relevant database after that date, and searches were restricted to English language documents. All the references from the database searches were imported into an Endnote Library and de-duplicated. The search strategies are detailed elsewhere [15]. Searches were extended to bibliographies of retrieved articles, and reference lists of key review articles in the area were scrutinised.

Papers that appeared relevant to the review were retrieved and assessed according to the inclusion criteria.

2.4 Review process

A data extraction tool was developed specifically for the purpose of this review. For included studies, information relevant to the review was extracted into a data extraction form by one of the authors (HU). The information was summarised in data tables, which provided the basis for assessment of the studies. Key characteristics of the models presented in each publication were recorded, as was information on interventions studied and results of the evaluations. The four numbered sections of the data extraction form addressed the four research questions of the review, and further details are provided below.

2.4.1 Definition of target decision-maker or jurisdiction—Being aware of the target decision-making audience for a model is important to a judgement about the appropriateness of the model and its inputs, and the review attempted to elicit the target decision-making audience or jurisdiction. In some instances, where models did not explicitly state the target decision-maker, it was possible to infer a decision-maker from the perspective taken and the data incorporated. In addition, the studies were assessed according to whether the study question was clearly stated or not, or whether the study question could be easily inferred.

2.4.2 Transparent reporting of model specification—Transparent reporting of a model is a prerequisite to understanding the relevance of the model to the target decision-maker, as well as to assessing its generalisability to other decision-makers and jurisdictions. The specification of study setting (e.g. country, and primary or secondary care) and patient population was therefore extracted. In addition, the description and justification of alternative interventions was considered, and an assessment of transparent reporting of the model structure and key assumptions was made.

2.4.3 Relevance of data inputs to target decision-maker or jurisdiction—The ease with which model inputs can be traced, and the relevance of those inputs to the stated decision-maker, will influence the degree to which a model is considered applicable in the target setting. The reporting of sources and the relevance of key data inputs to the model was therefore assessed, ranging from clinical data and their valuation to resource use and unit costs. Models that reported and referenced both baseline risks and risk reductions were considered as having provided sources of clinical data, whereas models that only reported references for risk reduction were assessed as having partially provided sources of clinical data.

2.4.4 Assessment of robustness of model to variation in data inputs within and between jurisdictions—The use of sensitivity analysis to explore the robustness of model results to variation in data inputs that may exist within and between jurisdictions was assessed. It has been extensively argued [15] that resource use estimates and their valuation, as well as health state estimates and their valuation, may vary across settings. This variation may exist between units within a given decision maker's jurisdiction [9] and, to provide information to the decision maker, the implications of such variation should be assessed. The authors may also choose to assess the robustness of their model's results to the level of variation that might be expected between jurisdictions. For example, the average length of stay in hospital for patients with hip fracture has been reported to be 29.6 days in Aberdeen and 41.7 days in Peterborough [18], whereas the national average in Denmark has been reported to be 21 days [19].

Similarly, the estimate of clinical effect incorporated into a model may depend on the study population and target population of the model. The robustness of the model's results under a range of clinical effectiveness estimates from different studies could, therefore, be explored in sensitivity analysis. Models may also undertake adjustments to translate the results of explanatory trials that may not hold in routine clinical practice. For example, compliance is generally acknowledged to be higher within the context of randomised trials than in clinical practice, and this may contribute substantially to the reduction in efficacy when an intervention is used in a non-trial environment [20, 21]. Reduced compliance in a clinical practice setting may result in reduced effectiveness of the drug, so models that evaluate the population-based impact of a strategy in clinical practice may provide a more representative estimate by factoring in the reduced compliance in the analysis [22].

Finally, the articles were reviewed with respect to whether the authors commented that results would be relevant to address the same decision-problem within the different jurisdictions.

3 Results

3.1 Summary of included papers

A total of 18 publications reporting economic evaluation models satisfied the inclusion criteria (Table 1). These included four Markov state transition models [17, 23-25] and four simple decision trees [26-29]. Nine studies were cost utility analyses. Of the studies published in the 1980s, six were from the United States. Eight studies were from European countries, all of which were published in the 1990s. The studies covered the following countries: Australia, Canada, Denmark, UK, US, and Italy. Apart from the Italian study [29], all results were presented in local currencies. Many models based the evaluation on a time horizon spanning the lifetime from the onset of treatment at menopause, usually assumed to be 45 or 50 years of age [17, 19, 24, 25, 30-35]. Others evaluated a more limited time horizon, for example 2 years in Francis *et al.*, [26] 3 years in Rosner *et al.*, [36] 3-4 years in Torgerson *et al.* [28] and 1 year in Visentin *et al.* [29].

Overall, the main base-case results of the studies did not reveal any systematic differences across the studies that might be explained by location (Table 1). Neither systematic variation in cost-effectiveness estimates within countries or systematic changes over time were apparent from the review. Despite focusing on the osteoporosis area, the models compared a range of interventions and presented results using a variety of outcomes. For example, Ankjaer-Jenssen *et al.* [19] reported average cost per hip fracture avoided in screened and unscreened populations, comparing three different interventions. In contrast, Tosteson *et al.* [24] compared costs and QALYs of two different interventions in patient populations with different life expectancy.

3.2 Target decision-maker

None of the studies evaluated cost-effectiveness from a broader perspective than that of the health service sector. The target decision-making audience was only explicitly stated in three (16%) of the included studies (Table 2). Specifically, Coyle *et al.* stated the decision maker to be “a Canadian provincial Ministry of Health”, [37] OTA stated that the report was commissioned by the US Senate Special Committee on Ageing, [17] and Visentin *et al.* commented that the evaluation was targeting the Italian Health Service [29].

It was, however, possible to infer a target decision-maker for many of the remaining studies. For example, Ankjaer-Jenssen *et al.* commented that the analysis was “carried out in a Danish context”; [19] and the study by Cheung *et al.* seemed to target a decision-making body in New South Wales, Australia [30]. Similarly, Daly *et al.* and Torgerson *et al.* appear

to tailor their analyses to be applicable in a British context without explicitly stating this [27, 31]. It was not possible to infer a decision-making audience in eight studies.

The research question was explicitly stated in seven studies (39%) (Table 2). For example, Coyle *et al.* expressed the economic study question as "...to assess the cost-effectiveness of nasal calcitonin compared with no therapy, alendronate or etidronate in the treatment of post-menopausal women with previous osteoporotic fracture" [37]. For those studies that explicitly stated the strategies under comparison, and the outcome, at the outset, it was possible to infer the study question. For example, Torgerson *et al.* did not explicitly state a research question but compared the (average) cost-effectiveness of screening followed by HRT treatment versus no screening and universal treatment [27].

3.3 Transparency of reporting

The vast majority of the studies specified that the target population was women living in the community, though one study specifically considered women in nursing homes [27], and the target country could be inferred for all studies. The target populations were also indicated for all models. Gender, post-menopausal status or age above 50 were used in most studies to identify the study population. Susceptibility to osteoporosis, either via previous fracture [37] or low bone mass density, [28] were also used as descriptive factors. Some models evaluated hysterectomised women separately from those with uterus *in situ* [23, 27, 31, 32, 35]. Two studies restricted the analysis to apply to caucasian women only [23, 26] due to an underlying difference in baseline hip fracture risk between ethnic groups.

The model structure was described in most of the studies through an outline of all clinical outcomes incorporated into the structure or, where relevant, health states and transitions. One study did not present the structure adequately [29]: it was, for example, unclear how this model estimated hip fracture risk reduction as a consequence of treatment.

The main assumptions were clearly stated and justified in most studies. For example, Coyle *et al.*, [37] OTA [17] and Tosteson *et al.* [25] clearly presented all assumptions in the model as well as omissions from the models. The study by Visentin *et al.* [29] was not entirely transparent - for example, it was not clear which trials were used to populate the model.

3.4 Model input and relevance to stated decision-maker

3.4.1 Clinical data and health state valuation—A common feature of the models in this review was the use of epidemiological studies to estimate the relative hip fracture risk reduction in the treated populations. Three models based the hip fracture risk reduction estimates on individual clinical trial data [26, 28, 36], whereas two models based the effect estimates for one of the therapies on meta-analyses of several trials [19, 37]. Of these studies, only two [26, 37] provided information about the clinical characteristics of the population of the trials

The remaining studies based the effect estimate primarily on observational studies. One of these studies [17] provided details of patient characteristics in the studies on which the effectiveness estimate was based, and two further studies presented patient characteristics only in terms of age range [24, 36].

Since the majority of studies provided only limited patient information, the scope for assessing the applicability of the results to the target population - and indeed other populations - was limited. Only two studies failed to provide any basis for the assumption on clinical effect [27, 30]. In spite of the fact that limited information was provided on the population sample, the majority of the studies provided references to primary studies that are likely to have given a more comprehensive description of the relevant patient sample. Only

one study adjusted the risks measured in the trials to the target population for the modelling exercise [36]. In spite of this, most studies appeared to use the best available data relevant to the stated or inferred decision-maker (Table 3).

Of the nine cost-utility studies that were included in the review, only two used utilities derived from patients [32, 37]. Rosner *et al.* [36] used a Canadian Delphi panel and the Health Utilities Index as a basis for the utilities used in their model. All the other authors based their utility weights on those assumed by Weinstein [34] either implicitly or explicitly. With the exception of the three studies mentioned above which used sample-based utility weights, it was difficult to assess the relevance of the utilities assumed by Weinstein [34] to particular health care decision-makers or jurisdictions (Table 3).

3.4.2 Resource use data and unit costs—The sources of resource use were explicit for the majority of the studies included in the review, only four studies omitted reporting this information. For example, resource use estimates and assumptions were explicit in the study reported by Ankjaer-Jensen *et al.*; [19] and Rosner *et al.* used a Delphi panel to estimate resource use [36]. Some studies included drug costs only in the estimate of resource use [26-29]. Sixteen studies in the review reported most sources of unit costs incorporated in the analyses whereas the remaining two studies [27, 29] did not report any sources for unit cost data.

It was difficult, if not impossible, to judge whether the estimates of resource use and unit costs were relevant to the decision-maker for those studies that did not explicitly state the decision-making audience for the study (Table 3). For those studies for which a decision-maker was explicit or could otherwise be inferred, the sources of resource use were largely judged to be relevant. The target decision-maker was not clear in the studies by Tosteson *et al.* [24, 25] and Weinstein [34]; however, the resource use would have been relevant provided that the target decision-maker was Medicare. It was not clear whether resource use was relevant to the decision-maker in four studies [26, 27, 29, 33].

3.5 Sensitivity analysis on context-specific factors

3.5.1 Clinical estimate and health state valuation—The studies were also reviewed to assess whether they had considered the implications of variation in input parameters using sensitivity analysis. A total of ten studies (55%) explored alternative assumptions of effect estimate in sensitivity analyses. Ankjaer-Jensen *et al.* [19] used “best case” and “worst case” scenarios in their analysis of effectiveness, and Rosner *et al.* [36] explored the 95% confidence interval boundaries for vertebral fracture rates in the model. Similarly, Cheung *et al.* [30] varied the risk of death from myocardial infarction over a range, and the OTA [17] varied the risk of all clinical parameters (bone loss, cancer and heart attack) in its model. Daly *et al.* [31, 32] varied both the magnitude and the duration of effect estimates in their sensitivity analysis, whereas Geelhoed *et al.* [23] explored the impact of HRT on different body systems (e.g. breast cancer) in the sensitivity analysis. The study by Coyle *et al.* [37] based its effect estimate on a meta-analysis of several trials, and found that the cost-effectiveness estimate was highly sensitive to the inclusion of one particular study.

These sensitivity analyses were largely undertaken to explore parameter uncertainty (e.g. due to sampling uncertainty) rather than explicitly to consider possible variability in clinical effects within or between jurisdictions. In part, this comment also applies to the two models which estimated hip fracture rates from bone mass density (BMD) [24, 38]. However, varying the population baseline hip fracture risk in these studies (for example, baseline hip fracture risk was increased by 100% and decreased by 50% in the Tosteson *et al.* [24] study) would have been of interest to decision makers as adjustment of baseline risk is often used to adapt the results of models between geographical areas (for an example see adaptation of

the WOSCOP study results to Belgium by Caro *et al.* [39] and the work by Palmer *et al.* [40])

Assumptions of compliance and duration of treatment were made without adjustment to clinical practice circumstances in the majority of studies in the review. For example, the OTA [17] assumed 100% compliance over 10, 20, 30 and 40 years; and Daly *et al.* [32] assumed 100% compliance over 5, 10, 15 and 20 years. The eight studies which took this into consideration differed in their definition of lack of compliance, but often it meant simply that patients ‘declined to accept’ therapy [30] or that patients ‘accepted but discontinued’ therapy [31] (Table 5). Generally, the cost-effectiveness estimates were found to be sensitive to the assumption of compliance, but the recommended policy-decision of the studies remained unchanged (Table 5). For example, Tosteson *et al.* [24] assumed 100% compliance over 15 years in the base case model, but varied compliance to 30% in the sensitivity analysis and found that cost-effectiveness estimates were sensitive to assumption of compliance.

One study [37] explored the sensitivity of health state valuation. Alternative utilities that assigned a 0.1 higher utility weight to the post-fracture health state for women receiving nasal calcitonin were identified. The estimates of ICER increased as a result of the alternative utilities. Again, however, this sensitivity analysis was associated more with parameter uncertainty than geographical variation.

Only five studies contrasted their findings with other economic evaluation studies in the area (Table 4). For example, the report by the OTA [17] provided a comprehensive discussion of methodology, costs, clinical assumptions and results in relation to other cost-effectiveness analyses. In principle, this would have allowed decision makers to assess whether other studies in the field had incorporated more appropriate data inputs for their jurisdiction.

3.5.2 Resource use and unit costs—Most studies applied unit cost data to the analysis relevant to the country for which the study was targeted (Table 8). Three studies used regional cost estimates [23, 30, 36]. None of the models accommodated differences in treatment patterns within regions or across countries. Furthermore, none of the investigators attempted to use cost estimates applicable to a broader audience of decision-makers within or between countries by using a range of costs or treatment patterns representing geographical differences. The sensitivity of the study results to national variation in cost per fracture was only explored in the study reported by Coyle *et al.* [37] who found that their cost-effectiveness estimate was markedly reduced when using cost calculations from an alternative cost-of-illness study.

3.6 Authors comments on generalisability

Overall, four studies explicitly commented on the issue of the generalisability of their analysis to other settings in their presentation or discussion of results (Table 6). [19, 27, 35, 37] For example, Coyle *et al.* [37] commented that the results were sensitive to the baseline population hip fracture risk, and for that reason the generalisability of the results were unclear.

4. Discussion

This paper has reviewed the use of decision analytic cost-effectiveness models, in a specific clinical area, to assess a range of factors associated with the potential relevance of an analysis to a particular decision maker, and the extent to which its results might be transferable to other jurisdictions. More general issues of good methods in decision analytic

cost-effectiveness modelling have been considered elsewhere [41-43], and a general critical review of cost-effectiveness models in osteoporosis has also previously been published [14].

We found that only 3 (16%) of the studies in this review provided details of jurisdiction and target decision-maker explicitly, although it was possible to infer the decision context from other information provided in some studies. Similarly, authors frequently report the methods and results of studies for which a firm research question has not been stated. To aid decision making, models also need to be clear about the decision problem(s) being addressed. The majority of reports in the review defined the study setting, patient population, model structure and key assumptions in a transparent manner.

Once the target decision-maker/jurisdiction and decision problem have been established, the former will need to decide whether the data inputs and assumptions in the model are the best available for their context. The majority of the reports provided sources for clinical and economic data, and their valuation. The data inputs to those models for which a decision-maker was stated or could be inferred appeared to be relevant and, as far as could be judged from only the published article, these were the 'best available' for the decision context. There is often an implicit assumption in models that estimates of clinical effectiveness are transferable between locations in a way that resource use and cost data are not. Perhaps reflecting this assumption, the papers in the review generally made more effort to ensure (and to be seen to ensure) that their cost inputs (at least unit costs) were specific to their target jurisdiction. Most studies were prepared to use clinical data from studies undertaken outside the context of stated decision-maker. As in other clinical areas, an exception to this assumption of the transferability of clinical data is the adjustment of baseline risks to make them specific to a particular jurisdiction (usually country) whilst assuming that the *relative* treatment effect is exchangeable geographically. [39] This adjustment of baseline risks was rarely undertaken in the sample of papers reviewed,[24] although the effect of variation in these parameters for the generalisability of analyses was discussed in another paper [37].

Little attempt was made to justify the particular utilities used with respect to the target decision-maker. This may reflect the limited amount of utility data available relating to osteoporosis. In other words, the authors often used any utility data which were available – in the majority of studies this was the assumption (rather than empirical elicitation) made by Weinstein [34]. Increasingly, decision makers will be specific about the type of health state utility data they wish to see in economic evaluations. For example, NICE has indicated that it wishes to see the use of public preferences relating to the British population in cost-utility analyses submitted to its technology assessment programme [43]. There are two aspects to this decision-maker preference. The first is the position a given decision-maker takes on the most appropriate utility estimation methods (e.g. use of patient or public preferences). The second is the issue of whether the preferences of individuals outside the particular jurisdiction of interest are acceptable.

One aspect of the review was to assess the extent to which studies had assessed the impact of variability in parameter estimates associated with location using sensitivity analysis. In principle, this sort of analysis might be undertaken for two reasons. Firstly, there may be variability in clinical and/or cost parameters *within* a given jurisdiction. If there was good reason to think that this level of variability might impact on the conclusions of the analysis, sensitivity analysis would be imperative. The second way sensitivity analysis could be used is to assess whether the results of the model, as they apply to the target jurisdiction, would also apply to other locations by appropriate variation in parameter values. If it is accepted that the purpose of decision models is to address decision problems for particular jurisdiction/decision makers, then this process of generalisation should probably not be seen as an essential element of model-based economic analysis.

Although most of the studies in the review undertake extensive sensitivity analysis, few explicitly do this to explore variation between locations/jurisdictions. Only one [17] of the models is probabilistic; that is, reflecting the uncertainty in parameters as random distributions and propagating that uncertainty through the model, to be jointly reflected in the results, using Monte Carlo simulation [44]. Therefore, most of the standard one-way or two-way sensitivity analysis was undertaken to assess the importance of parameter uncertainty to the results, rather than variability between locations. The failure to assess variability within and between locations may have reflected the view that parameter uncertainty dwarfed variability within the target jurisdiction, and generalising their results across locations was not a primary concern.

One area of interest in the review was how the models dealt with the issue of patients' compliance with therapy. Compliance is expected to vary between research (e.g. trial) settings and routine practice and this is therefore important when assessing generalisability to clinical practice. Incorporating compliance rates in economic analyses, for example through changing of clinical benefit parameters or deviations in cost input, have previously been demonstrated to impact the results of the analyses (Ref 14), [22].. Different definitions of compliance were used by the authors in this review, but, for the most part, the assumption of compliance was based on the proportion of patients that initiates therapy. The overall results of the analysis were sometimes substantially influenced by the assumption of patient compliance, and more research is needed in order to determine the optimum manner of generalising from study to clinical practice.

Whilst development of a set of recommendations for authors would be valuable, it was beyond the scope of this journal article to provide specific guidance over and above that provided by others.[45] However, this review has highlighted important aspects to consider in reporting results of modelling studies so that decision-makers can determine generalisability, and relevance of the results in their location. For example, resource use data and costs are key determinants of relevance. Could authors in their reporting distinguish those factors that prove specific to the stated decision-problem (and decision-making audience) from those that are generalisable to other locations?

This review was limited to articles published in the English language and limited to articles published up to, and including, the year 2001 only. One may assume that single-country assessments published in a local language have a higher degree of generalisability, or relevance, to the local environment, and as such this review may have captured a biased sample of studies. . Given that the purpose of this review was to provide a conceptual analysis of an under-investigated issue in the HTA literature, i.e. the extent to which the information produced in a cost-effectiveness analysis report allows jurisdiction specific decision makers to establish the validity of the study results to their own setting, we do not believe the lack of inclusion of more recent studies reduces the importance of the findings. However, the review was undertaken before agencies such as NICE [43] published their "Reference Case" for conducting and reporting economic models, and one would expect the standard of reporting to increase following such influential guidance. The review was also primarily limited to published journal articles where limited word-length naturally restricts the amount of information available to the reader.

5. Conclusions

The review found that most studies either stated their target decision-maker/jurisdiction or provided sufficient information from which this could be inferred. There was a greater tendency to ensure that cost (resource use and unit cost) inputs were specific to the target jurisdiction, rather than clinical parameters. Although there was extensive sensitivity

analysis undertaken in most studies to assess parameter uncertainty, there was little use of these methods to explore the implications of variability within or between locations in parameter inputs. Only four studies explicitly commented on the issue of the generalisability of their analysis to other settings in their presentation or discussion of results. In spite of some limitations, this review found that variability in cost-effectiveness across locations is addressed to a varying extent in modelling studies in the field of osteoporosis, limiting their use for decision-makers. Further guidance is needed to those publishing results of economic model, supporting the reporting of design, conduct and results of models, however, transparency of reporting is expected to increase as methodology develops, and decision-makers publish “reference case” type guidance.

Acknowledgments

This work was developed as part of a project on generalisability in economic evaluation studies in health care (98/22/05) funded by the NHS Health Technology Assessment Programme.

References

1. Hjelmgren J, Berggren F, Andersson F. Health economic guidelines - similarities, differences and some implications. *Value in Health*. 2001; 4:225–250. [PubMed: 11705185]
2. Weinstein, MC.; Fineberg, HV. *Clinical Decision Analysis*. Philadelphia: 1980.
3. Hunink, M.; Glaziou, P.; Siegel, J.; Weeks, J.; Pliskin, J.; Elstein, A., et al. *Integrating Evidence and Values*. Cambridge: Cambridge University Press; 2001. *Decision Making in Health and Medicine*.
4. Sculpher, MJ.; Claxton, K.; Akerhurst, R. It's just evaluation for decision making: recent developments in, and challenges for, costeffectiveness research. In: Sculpher, MJ., editor. *Health Policy and Economics: opportunities and challenges*. Maidenhead, Berkshire, England: Oxford University Press; 2005.
5. Willke RJ, Glick HA, Polsky D, Schulman KA. Estimating country-specific cost-effectiveness from multinational clinical trials. *Health Economics*. 1998; 7:481–493. [PubMed: 9809707]
6. Willan AR, Pinto EM, O'Brien BJ, Kaul P, Goeree R, Lynd L, et al. Country specific cost comparisons from multinational clinical trials using empirical Bayesian shrinkage estimation: the Canadian ASSENT-3 economic analysis. *Health Economics*. 2005; 14:327–338. [PubMed: 15685652]
7. Thompson SG, Nixon R, Grieve R. Addressing the issues that arise in multicentre cost data, with application to a multinational study. *Journal of Health Economics*. 2006 forthcoming.
8. Pinto EM, Willan AR, O'Brien BJ. Cost-effectiveness analysis for multinational clinical trials. *Statistics in Medicine*. 2005; 24:1965–1982. [PubMed: 15803442]
9. Manca A, Rice N, Sculpher MJ, Briggs AH. Assessing Generalisability by Location in Trial-Based Cost-Effectiveness Analysis: the Use of Multilevel Models. *Health Economics*. 2005; 14(5):471–85. [PubMed: 15386662]
10. Manca A, Lambert PC, Sculpher MJ, Rice N. Cost effectiveness analysis using data from multinational trials: The use of bivariate hierarchical modelling. 2006 submitted. [PubMed: 17641141]
11. Koopmanschap MA, Touw KCR, Rutten FFH. Analysis of costs and cost-effectiveness in multinational trials. *Health Policy*. 2001; 58(2):175–186. [PubMed: 11551665]
12. Grieve R, Nixon R, Thompson SG, Normand C. Using multilevel models for assessing the variability of multinational resource use and cost data. *Health Econ*. 2005; 14(2):185–96. [PubMed: 15386660]
13. Coyle D, Drummond MF. Analyzing differences in the costs of treatment across centers within economic evaluations. *International Journal of Technology Assessment of Health Care*. 2001; 17(2):155–63. [PubMed: 11446127]
14. Sculpher, MJ.; Torgerson, D.; Goeree, R.; O'Brien, B. *A Critical Structured Review of Economic Evaluations of Interventions for the Prevention and Treatment of Osteoporosis*. University of York: CHE; 1999. Centre for Health Economics Discussion Paper No. 169

15. Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Davies LM, et al. Generalisability in Economic Evaluation Studies in Health Care: a Review and Case-Studies. *Health Technology Assesment*. 2004; 8(49):1–206. [PubMed: 15544708]
16. Royal College of Physicians of London. *Osteoporosis: Clinical Guidelines for Prevention and Treatment*. London: Royal College of Physicians of London; 1999.
17. Office of Technology assessment (OTA) US Congress. *Effectiveness and Costs of Osteoporosis Screening and Hormone Replacement Therapy, Volume I: Cost effectiveness analysis*. Washington, DC: US Government Printing Office; 1995. OTA-BP-H-160
18. French F, Torgerson DJ, Porter RW. Overestimation of cost savings in prevention of osteoporosis (letter). *BMJ*. 1994; 308:138. [PubMed: 8298410]
19. Ankjaer-Jensen A, Johnell O. Prevention of osteoporosis: cost effectiveness of different pharmaceutical treatments. *Osteoporosis International*. 1996; 6:265–275. [PubMed: 8883114]
20. Revicki DA, Frank L. Pharmacoeconomic evaluation in the real world: Effectiveness versus efficacy studies. *Pharmacoeconomics*. 1999; 15(5):423–434. [PubMed: 10537960]
21. Bombardier C, Maetzel A. Pharmacoeconomic evaluation of new treatments: efficacy versus effectiveness studies? *Annals of Rheumatic Diseases*. 1999; 58:182–185. [PubMed: 10577979]
22. Hughes DA, Bagust A, Haycox A, Walley T. The impact of non-compliance on the cost-effectiveness of pharmaceuticals. A review of the literature. *Health Economics*. 2001; 10:601–615. [PubMed: 11747044]
23. Geelhoed E, Harris A, Prince R. Cost-effectiveness analysis of hormone replacement therapy and lifestyle intervention for hip fracture. *Australian Journal of Public Health*. 1994; 18(2):153–60. [PubMed: 7948331]
24. Tosteson A, Rosenthal D, Melton L, Weinstein M. Cost effectiveness of screening perimenopausal white women for osteoporosis: bone densitometry and hormone replacement therapy. *Annals of Internal Medicine*. 1990; 113(8):598–603. [PubMed: 2119161]
25. Tosteson A, Weinstein M. Cost-effectiveness of hormone replacement therapy after the menopause. *Baillieres Clinical Obstetrics and Gynaecology*. 1991; 5:943–959. [PubMed: 1822828]
26. Francis R, Anderson F, Torgerson D. A comparison of the effectiveness and cost of treatment for vertebral fractures in women. *British Journal of Rheumatology*. 1995; 34:1167–1171. [PubMed: 8608360]
27. Torgerson D, Reid D. Osteoporosis prevention through screening: will it be cost effective? *Baillieres Clinical Rheumatology*. 1993; 7(3):603–622. [PubMed: 8293493]
28. Torgerson D, Kanis J. Cost effectiveness of preventing hip fractures in the elderly population using vitamin D and calcium. *Quarterly Journal of Medicine*. 1995; 88:135–139. [PubMed: 7704564]
29. Visentin A, Ciravegna R, Fabris F. The cost per avoided hip fracture by osteoporosis treatment in Italy. *Maturitas*. 1997; 26:185–192. [PubMed: 9147350]
30. Cheung AP, Wren BG. A cost-effectiveness analysis of hormone replacement therapy in the menopause. *Medical Journal of Australia*. 1992; 156:312–316. [PubMed: 1588862]
31. Daly E, Roche M, Barlow D, Gray A, McPherson K, Vessey M. HRT: An analysis of benefits, risks and costs. *British Medical Bulletin*. 1992; 48(2):368–400. [PubMed: 1450876]
32. Daly E, Vessey M, Barlow A, Gray K, McPherson K, Roche M. Hormone replacement therapy in a risk-benefit perspective. *Maturitas*. 1996; 23:247–259. [PubMed: 8735363]
33. Garton M, Cooper C, Reid D. Perimenopausal bone density screening - will it help prevent osteoporosis? *Maturitas*. 1997; 26:35–43. [PubMed: 9032745]
34. Weinstein M. Estrogen use in postmenopausal women - costs, risks and benefits. *New England Journal of Medicine*. 1980; 303:308–16. [PubMed: 6770270]
35. Weinstein M, Schiff I. Cost-effectiveness of hormone replacement therapy in the menopause. *Obstetrical and Gynaecological Survey*. 1983; 38:445–454. [PubMed: 6310460]
36. Rosner A, Grinna D, Torrance G, Bradley C, Adachi J, Sebaldt R, et al. Cost effectiveness of multi-therapy treatment strategies in the prevention of vertebral fractures in postmenopausal women with osteoporosis. *Pharmacoeconomics*. 1998; 14(5):559–573. [PubMed: 10344918]

37. Coyle D, Cranney A, Lee KM, Welch V, Tugwell P. Cost-effectiveness of nasal calcitonin in postmenopausal women. Using Cochrane collaboration methods for meta analysis in economic evaluation. *Pharmacoeconomics*. 2001; 19:565–575. [PubMed: 11465301]
38. Weinstein M, Tosteson A. Cost-effectiveness of hormone replacement. *Annals of the New York Academy Science*. 1990; 592:162–173. [PubMed: 2115756]
39. Caro JJ, Huybrechts KF, De Backer G, De Bacquer D, Closon MC. Are the WOSCOPS clinical and economic findings generalizable to other populations? A case study for Belgium. *Acta Cardiologica*. 2000; 55(4):239–246. [PubMed: 11041122]
40. Palmer, S.; Asseburg, C.; Bravo-Vergel, Y.; Fenwick, E.; Sculpher, MJ. The cost-effectiveness of primary angioplasty for acute myocardial infarction: an application of bayesian evidence synthesis using decision modelling. In: SMDM. , editor. Society for Medical Decision Making 27th Annual Meeting; San Francisco. 2005; 2005.
41. Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models: a suggested framework and example of application. *Pharmacoeconomics*. 2000; 17(5): 461–477. [PubMed: 10977388]
42. International Society for Pharmacoeconomics and Outcomes Research. Principles for Good Practice in Decision Analytic Modelling in Health Care Evaluation. 2001A Statement from the ISPOR Taskforce on Modelling(<http://www.ispor.org/workpaper/healthscience/TFModeling.pdf>)
43. National Institute for Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal. London: National Institute for Health and Clinical Excellence (NICE); 2004.
44. Briggs AH, Goeree R, Blackhouse G, O'Brien BJ. Probabilistic analysis of cost-effectiveness models: choosing between treatment strategies for gastroesophageal reflux disease. *Medical Decision Making*. 2002; 22:290–308. [PubMed: 12150595]
45. Drummond MF, Manca A, Sculpher MJ. Increasing the Generalisability of Economic Evaluations. Recommendations for the Design, Analysis and Reporting of Studies. *International Journal of Technology Assessment in Health Care*. 2005; 21(2):165–71. [PubMed: 15921055]

Table 1
Studies included in the review and the base-case results of the evaluations

Study (year)	Interventions evaluated	Incremental cost-effectiveness ratio	Incremental cost per QALY
Ankjaer-Jensen 1996[19]	Screening or not followed by calcium supplementation, etidronate, calcitonin or HRT.	Cost per hip fracture avoided: Population based vs. screening: Calcitonin: 1,136,200 DKK Etidronate: 77,567 DKK Calcium: 522,300 DKK	NA
Cheung 1992[30]	Oestrogen replacement therapy alone vs. no treatment and oestrogen and progestin vs. no treatment	NA	10 years treatment ORT: \$ 9,500 – 17,500, symptomatic women ORT: \$ 45,800 – dominated, asymptomatic women ORT: \$ 6,510 – 1 020,000, hysterectomy CRT: \$ 26,100 – 1 450,000, asymptomatic women CRT: \$ 9,820 – 34,700, symptomatic women
Coyle 2001[37]	Calcitonin was compared with alendronate and etidronate.	Cost per life-year gained, 5-year treatment of 65-year old: Alendronate dominant strategy. ICER for nasal calcitonin vs no therapy: \$Can 46,500. ICER for nasal calcitonin vs etidronate: \$Can 32,600.	5 year treatment in 65-year old: Alendronate dominant strategy. ICER for nasal calcitonin vs no therapy: \$Can 46,500. ICER for nasal calcitonin vs etidronate: \$Can 32,600.
Daly 1992[31]	Oestrogen only vs. no treatment (no uterus) and oestrogen vs oestrogen + progestin (uterus)	Cost per life-year gained (10 years treatment): ORT: £2,900 in women with no uterus ORT: £8,300 in women with uterus CRT: £14,400 in women with uterus	10 years treatment in mildly symptomatic women ORT: £1,700 in women with no uterus ORT: £4,400 in women with uterus CRT: £6,200 in women with uterus
Daly 1996[32]	Oestrogen only vs. no treatment (no uterus) and oestrogen vs oestrogen + progestin (uterus) vs. no treatment.	NA	Mildly symptomatic women ORT: £310 (5 years treatment) to £660 (20 years treatment) CRT: £550 (5 year treatment) to £1,250 (20 years treatment)
Francis 1996[26]	Calcitonin vs. no treatment, oestrogen + progestin vs no treatment, etidronate vs. no treatment.	Cost per vertebral fracture avoided, treatment vs. no therapy: HRT: £138 – 680 Etidronate: £1,880 Calcitonin: £9,075 – 25,013	NA
Garton 1997[33]	Interventions (screening vs no	Cost per fracture prevented HRT vs. no therapy:	NA

Study (year)	Interventions evaluated	Incremental cost-effectiveness ratio	Incremental cost per QALY
	screening) clear but not justified. A strategy of oestrogen + progestin with screening vs a strategy of oestrogen + progestin without screening (universal)	Screening: £1,000 – 4,200 Universal treatment: £1,200	
Geelhoed 1994[23]	Oestrogen vs. no treatment and lifestyle regime (exercise) vs. no treatment.	NA	Intervention vs. no treatment: ORT for life from 50 years: \$8,810 ORT for 15 years: \$16,500 ORT from age 65: \$8,450 Calcium and exercise: \$28,480
OTA 1995[17]	Oestrogen (ERT) vs. oestrogen + progestin (PERT) following different screening + treatment strategies	Mean cost per Life Year Gained ORT (Screening): \$22,431 – \$151,392 depending on duration of therapy and screening threshold. ORT (Population-based): \$23,334 – \$126,876 depending on duration of therapy. Life-long therapy most cost-effective.	NA
Rosner et al 1998[36]	Oestrogen, oestrogen + progestin, alendronate, calcium combinations and no intervention	ICER: Cost per averted vertebral fracture for four non-dominated strategies in order: (i) calcium – no therapy (ii) HRT – calcium – no therapy: ICER = \$Can166 (iii) HRT – etidronate – calcium – no therapy: ICER = \$Can 2,331 (vi) HRT – alendronate – calcium – no therapy: ICER = \$Can 40,965	NA
Torgerson 1993[27]	Compares a strategy of oestrogen only in high-risk patients (screening) vs. oestrogen + progestin in population based treatment.	Cost per averted hip fracture Universal treatment for 10 years: £40,080 Screening, then treatment: £34,971	NA
Torgerson 1995[28]	Vitamin D and calcium vs. no treatment and vitamin D vs. no treatment	Cost per averted hip fracture Oral vitamin D + Calcium: £17,379 in community (low BMI); £4,735 in nursing homes; cost saving overall in nursing homes (low BMI). Vitamin D injection: all options cost saving overall.	NA

Study (year)	Interventions evaluated	Incremental cost-effectiveness ratio	Incremental cost per QALY
Tosteson 1990[24]	Screening vs. no screening. No intervention, screening followed by HRT (oestrogen + progestin) and universal HRT (oestrogen + progestin).	NA	Screening and then treatment (15 years): \$4,200 – 37,800 depending on treatment threshold Universal treatment (15 years): \$144,000
Tosteson 1991[25]	oestrogen vs. no intervention (intact uteri), oestrogen + progestin vs. no intervention (hysterectomy).	NA	ORT (in women without uterus): US\$7,010 – \$9,020; without quality of life side effects 9,930 – 14,940 CRT (in women with uterus): US\$32,660 – \$33,780; without quality of life side effects US\$ > 150,000
Visentin 1997[29]	Calcitonin compared to no treatment.	Cost per averted hip fracture: Calcitonin vs. no treatment: 2,367,987 Population based approach: US\$ 2,367,987 Screening and treatment of high-risk patients US\$838,120	NA
Weinstein 1980[34]	Oestrogen vs no treatment.	NA	ORT in women with uterus: \$7,420-18,160 ORT in women with osteoporosis: \$ 5,460 – 15,100
Weinstein 1983[35]	Oestrogen vs. no treatment and oestrogen + progestin vs no treatment.	NA	ORT: \$ 130,000 vs. CRT CRT: \$ 42,000 (5 years treatment) to 24,000 (10 years treatment) vs. no treatment
Weinstein 1990[38]	Oestrogen vs. no treatment and oestrogen + progestin vs no treatment.	NA	ORT for 5 years: US\$ 72,100 in asymptomatic women ORT for 5 years: US\$ 12,600 – 33,100 in symptomatic women depending on symptom relief. CRT for 5 years dominates ORT for 5 years assuming quality of life differences are minimal CRT for 15 years: US\$ 22,650

CRT = oestrogen only therapy

HRT = Hormone replacement therapy

ORT = oestrogen and progestin therapy

NA = Not applicable

Table 2

Target decision-maker and study question

Study (year)	Q1.1 Decision-maker	Q1.2 Economic study question
Ankjaer-Jensen 1996[19]	Inferred. "the analysis was carried out in a Danish context".	Inferred. Cost-effectiveness of population-based treatment programme compared to CE of screening programme.
Cheung 1992[30]	Inferred. Study was carried out in New South Wales, Australia.	Explicit.
Coyle 2001[37]	Explicit. The target decision-maker of the model was clearly stated as "a Canadian provincial ministry of health".	Explicit. The economic study question was clearly outlined as "to assess the cost-effectiveness of nasal calcitonin compared with no therapy, alendronate or etidronate in the treatment of postmenopausal women with previous osteoporotic fracture".
Daly 1992[31]	Inferred. The authors state that "an assessment (...) in the British context is timely".	Inferred. Three explicit treatment strategies were compared.
Daly 1996[32]	Inferred. "costs to the NHS..."	Explicit.
Francis 1996[26]	No.	No.
Garton 1997[33]	No.	Inferred. Bone mass measurement compared to no measurement (i.e. screening)
Geelhoed 1994[23]	Inferred. Western Australian health service.	Inferred. Four strategies were compared.
OTA 1995[17]	Yes. Report commissioned by the US Senate Special Committee on Aging.	Yes. Strategies of screening and HRT treatment compared.
Rosner et al 1998[36]	Inferred. Clinical decision-makers in Canada.	Yes. Four explicit strategies compared.
Torgerson 1993[27]	Inferred. The NHS in the UK.	Inferred. Screening vs. no screening followed by HRT.
Torgerson 1995[28]	No.	Yes. Estimate cost per averted fracture for two interventions.
Tosteson 1990[24]	Inferred.	Yes.
Tosteson 1991[25]	Inferred.	Inferred.
Visentin 1997[29]	Yes.	Inferred.

Study (year)	Q1.1 Decision-maker	Q1.2 Economic study question
Weinstein 1980[34]	Italian Health Service.	Inferred.
Weinstein 1983[35]	No.	Yes. "to synthesise the available evidence in comparing the costs, risks and benefits of oestrogen-progestin therapy and oestrogen-only therapy in post-menopausal women."
Weinstein 1990[38]	No.	Inferred.

Table 3
Relevance of the model input to the stated decision-maker

Study (year)	Q3.2a Clinical data sources relevant to decision-maker?	Q3.2b Resource use data relevant to decision-maker?	Q3.2c Unit costs relevant to decision-maker?	Q3.2d Preferences / utilities relevant to decision-maker?
Ankjaer-Jensen 1996[19]	Yes.	Yes.	Yes.	NA
Cheung 1992[30]	No.	Yes.	Yes.	Not clear.
Coyle 2001[37]	Yes.	Yes.	Yes.	Yes.
Daly 1992[31]	Yes.	Yes.	Yes.	Not clear.
Daly 1996[32]	Yes.	Yes.	Yes.	Yes.
Francis 1996[26]	Not clear.	Not clear.	Not clear.	NA.
Garton 1997[33]	Not clear.	Not clear.	Not clear.	NA
Geelhoed 1994[23]	Yes.	Yes.	Yes.	Not clear.
OTA 1995[17]	Yes.	Yes.	Yes.	NA
Rosner et al 1998[36]	Yes.	Yes.	Yes.	Not clear.
Torgerson 1993[27]	Not clear.	Not clear.	Not clear.	NA
Torgerson 1995[28]	Not clear.	Not clear.	Yes.	NA
Tosteson 1990[24]	Not clear.	Yes.	Yes.	Not clear.
Tosteson 1991[25]	Not clear.	Yes.	Yes.	Not clear.
Visentin 1997[29]	In part.	Not clear.	Not clear.	NA
Weinstein 1980[34]	Yes.	Not clear.	Not clear.	Not clear.
Weinstein 1983[35]	Yes.	Yes.	Yes.	Not clear.
Weinstein 1990[38]	Yes.	Yes.	Yes.	Not clear.

Table 4
Sensitivity analysis of estimates of clinical effectiveness and health state valuation

Study (year)	Q4.1a Was the robustness of the effect estimate explored?	Q4.1b Did the model accommodate for potential difference in compliance rates?	Q4.1c Did the model accommodate for differences in utilities?	Q4.1d External consistency: Were the results compared to other relevant studies?
Ankjaer-Jensen 1996[19]	Yes.	Yes.	NA	Yes.
Cheung 1992[30]	Yes.	No.	No.	No.
Coyle 2001[37]	Yes.	Yes.	Yes.	No.
Daly 1992[31]	Yes.	Yes.	No.	Yes.
Daly 1996[32]	Yes.	No.	No.	No.
Francis 1996[26]	No.	No.	NA	No.
Garton 1997[33]	No.	Yes.	NA	Yes.
Geelhoed 1994[23]	Yes.	Yes.	No.	No.
OTA 1995[17]	Yes.	No.	NA	Yes.
Rosner et al 1998[36]	Yes.	Yes.	No.	No.
Torgerson 1993[27]	No.	Yes	NA	Yes.
Torgerson 1995[28]	No.	No.	NA	No.
Tosteson 1990[24]	Yes.	Yes.	No.	No.
Tosteson 1991[25]	No.	No.	No.	No.
Visentin 1997[29]	No.	No.	NA	Yes.
Weinstein 1980[34]	No.	No.	No.	NA.
Weinstein 1983[35]	No.	No.	No.	No.
Weinstein 1990[38]	Yes.	No.	No.	No.

Table 5
Definition of patient compliance and impact of compliance on the results of the cost-effectiveness analysis. Only studies explicitly considering compliance are included in the table.

Study	Definition of Compliance	Result of analysis
Ankjaer-Jensen 1996[19]	Compliance was % of patients attending bone mass density (BMD) screening. 'Compliance' of 100% and 50% was explored.	Screening was more cost-effective than population based approach under both assumptions of 'compliance' (attendance).
Cheung 1992[30]	Compliance was % of patients completing 15 years of treatment. Model base-case was 90% compliance for high-risk patients and 70% compliance for mid-risk patients. Explored impact of reduced compliance at 10% intervals	The estimate of cost savings was more sensitive to compliance in high-risk group. "Break even" (cost of savings = cost of intervention) was reached at compliance of 77% in high-risk only, or 70% in high-risk and 40% in mid-risk group.
Coyle 2001[37]	Compliance was proportion of patients taking medication after 1 year. Base-case was 50% after 1 year. The impact of 25% and 75% compliance was tested as well as 10% superior compliance with calcitonin.	Analysis was sensitive to the assumption about compliance with therapy. However, the overall results of the analysis did not change.
Daly 1992[31]	Compliance was % of patients continuously taking the drug. Compliance of 100% in first 5 years falling to 50% (oestrogen only, ORT) and 67% falling to 33% (oestrogen and progestin)	Cost-effectiveness increased with reduced compliance over all treatment strategies. The overall results of the analysis did not change.
Garton 1997[33]	Compliance rate used implied % patients initiating therapy and continuing beyond year 1 to complete a course of 10 years of treatment. Compliance of 10, 30 and 50% were explored.	Universal HRT was more cost-effective than screening strategy, both under high and low compliance. However, if screening could increase compliance then screening could prove more cost-effective.
Geelhoed 1994[23]	Assumed that patients who fill prescriptions but do not take the drug incur costs but gain no benefits. Explored scenario where 70% of prescriptions are filled but only 30% are taken as directed.	The net cost per QALY would increase ('more than double') under a scenario of reduced compliance.
Rosner et al 1998[36]	Different rates of willingness of patients to initiate (WTI) and continue (WTC) treatment were incorporated in the model ranging from 18.1% to 100%. These were based on epidemiological studies.	The model was "moderately sensitive" to changes in WTI and WTC. One strategy was particularly sensitive but it remained cost-effective under the assumption that public willingness to pay for a QALY is more than \$100,000
Torgerson 1993[27]	Compliance rates were defined as 'willingness to initiate' therapy. Once therapy initiated, 100% compliance assumed in 30 years. Different rates of willingness of patients to initiate and continue treatment are varied from 18.1% to 100% and based on epidemiological studies.	The outcome of the evaluation was sensitive to the assumption of compliance. More than 50% need to initiate therapy ('be compliant') in order for targeted intervention to be cost-effective.
Tosteson 1990[24]	Assumed 100% compliance with treatment varied from 5 year to lifetime use (baseline model 15-year). Explored 30% compliance in sensitivity analysis.	Results were "sensitive", but compliance did not change the main conclusion of the model: screening remained more cost-effective than universal treatment.

Table 6
Authors' comments on the generalisability of the results

Study (year)	Did the authors explicitly address the issue of transferability of the results to other jurisdictions?
Ankjaer-Jensen 1996[19]	Yes. The authors commented that the results are limited to a Danish context, considering wide variability in drug costs across countries. For example, they comment that drugs cost 50% less in Sweden and that this would influence the results.
Cheung 1992[30]	No.
Coyle 2001[37]	Yes. The authors commented that the results are sensitive to baseline risks of fracture in the population and that the generalisability to other jurisdictions was, therefore, unclear.
Daly 1992[31]	No.
Daly 1996[32]	No.
Francis 1996[26]	No.
Garton 1997[33]	No.
Geelhoed 1994[23]	No.
OTA 1995[17]	No.
Rosner et al 1998[36]	No.
Torgerson 1993[27]	No.
Torgerson 1995[28]	Yes. The authors comment that the cost of hip fracture was excluded because the hospital length of stay varies between UK regions and therefore an estimate based on a region with long stay would not apply to an estimate in a region with short stay.
Tosteson 1990[24]	No. However, the authors acknowledge that baseline fracture risk in the population is important to the CE ratio.
Tosteson 1991[25]	No.
Visentin 1997[29]	No. However, the authors note that the high CE ratio may have been due to relatively low incidence of hip fracture in the Italian population compared to other populations.
Weinstein 1980[34]	No.
Weinstein 1983[35]	Yes. The authors state that the cost advantage of the oestrogen-progestin alternative would have been even greater if British recommendations on more frequent endometrial monitoring had been adopted.
Weinstein 1990[38]	No.