TERIPARATIDE

INDEPENDENT REVIEW – PHARMACEUTICAL BENEFITS SCHEME

Generic Name:	Teriparatide (Rbe) Injection
Trade Name:	Forteo ®
Company Sponsor:	Eli Lilly Australia Pty Ltd
Content:	Independent review of matters identified by the sponsor in relation to PBAC decisions not to list Teriparatide for the treatment of severe osteoporosis

The names of reviewers have been withheld, consistent with PBAC procedures

Date: 7 September 2006

TERIPARATIDE

Independent Review of PBAC Submissions by Eli Lilly Australia

Table of contents

EXECUTIVE SUMMARY

1 FORMULATION

The formulation of teriparatide contains 250ug/ml presented in a 3 ml cartridge and supplied in a pre-filled, disposable delivery device (pen). The pen delivers 20ug per dose and contains dosing for 28 treatment days. The recommended dose of teriparatide is 20ug administered once daily. The route of administration is per subcutaneous (SC) injection into the thigh or abdomen.

2 REGISTRATION STATUS

Teriparatide was approved by the TGA for marketing in Australia on May 22, 2003 for "The treatment of postmenopausal osteoporosis and the treatment of primary osteoporosis in men when other agents are considered unsuitable and when there is a high risk of fractures."

3 PROPOSED PBS LISTING

Name, Restriction Manner of administration and form	Max Qty	No of Rpts	Dispensed Price for Max Qty	Proprietary Name and Manufacturer
TERIPARATIDE Injectable 3mL prefilled pen (20 ug/d)	1 pen	5	Not disclosed. The sponsor had proposed a price lower than the current private prescription price.	Forteo Eli Lilly

SECTION 85: Authority required

Treatment by a specialist/consulting physician treating osteoporosis for severe established vertebral osteoporosis in men and postmenopausal women who:

1. have evidence of one or more severe painful osteoporotic vertebral fracture, and
2. have received at least 6 continuous months of anti-resorptive therapy of proven efficacy and safety at the time of the SQ3 vertebral fracture.

A severe vertebral fracture is defined as *(at least)* 40% reduction in height of the anterior or mid portion of a vertebral body relative to the posterior height of that body, or, greater than 40% reduction in any of these heights compared to the vertebral body above or below the affected vertebral body.

Evidence of the fracture/deformity must be demonstrated radiologically and the year of plain x-ray or CT-scan or MRI scan must be included in the authority application.

Antiresorptive therapies for osteoporosis which will be accepted for the purposes of administering this restriction are aledronate sodium 10mg/day or 70mg QW, risedronate sodium 5mg/day or 35mg QW; raloxifene hydrochloride 60mg/day (women only); etidronate 200mg with calcium carbonate 1.25g/day. Patients with 6 months continuous prior treatment with strontium ranelate will also be eligible under the administration of this listing.

If treatment with the above-mentioned drugs is contraindicated according to the relevant TGA-approved Product Information, or intolerance of a severity necessitating permanent treatment withdrawal develops during the relevant period of use, the patient is exempted from the requirement to complete 6 continuous months of therapy with the particular agent or class of agents. Details of the contraindication or intolerance must be provided at the time of application.

Continuing treatment where the patient has previously been issued with an authority prescription for this drug.

Teriparatide is available with a lifetime maximum of 18 months teriparatide therapy (18 pens), a maximum of 18 pens will be reimbursed through the PBS.

4 HISTORY OF SUBMISSIONS TO PBAC

The sponsor filed four submissions to PBAC requesting a listing for teriparatide on the PBS. The first (June 2003) and second (March 2004) submissions positioned teriparatide after failure of antiresorptive therapy. The third submission (July 2005) requested teriparatide for patients with SQ3 vertebral fractures. The fourth submission (March 2006) combined elements of the three prior submissions in requesting a listing for teriparatide in patients who have suffered an SQ3 vertebral fracture despite at least 6 months of antiresorptive treatment.

5 MATTERS FOR REVIEW

In a letter dated 28 April 2006, the sponsor has nominated the following issues on which review is sought:

a) PBAC concern that the indirect data comparison and subgroup analysis does not support the claim of superiority-paragraph 4 of PBAC minutes.
b) PBAC claim that the submission provides no evidence to suggest any biological plausible reason to explain the claimed difference between teriparatide and the comparator-paragraph 9 of PBAC minutes.
c) PBAC concern with regard to continuing use of the same utility values in spite of the sponsor's efforts to address these concerns in its responses-paragraph 4 of PBAC minutes.

6 FINDINGS OF THE REVIEW

a) <u>Indirect comparison and subgroup analysis</u>
   There are two key issues involved in this matter: - (1) the dependence on an indirect data comparison across placebo-controlled trials (rather than head-to-head studies) to infer the superiority of teriparatide over an appropriate comparator treatment, and (2) the scientific validity of using the results of a post hoc subgroup analysis in place of the overall Intention-to-Treat (ITT) results for teriparatide in formulating the clinical conclusion. The two issues are fundamental to the robustness of the sponsor's scientific claim. Although the sponsor has undertaken appropriate statistical measures to optimize the validity of the indirect data comparison, this is a scientifically less rigorous process than the gold standard methodology of head-to-head randomized

controlled studies. In particular, the validity of adjusted indirect comparisons depends on the internal validity and similarity of the trials involved in the comparison. Some concerns remain regarding the internal validity and generalisability of the pivotal GHAC study which include an unanticipated premature termination of the study and the significant loss of evaluable subjects due to various reasons. The result derived from post-hoc subgroup analysis of the GHAC trial indicate an unclear and arguable benefit with teriparatide treatment in patients with more severe forms of vertebral osteoporosis (SQ3) at baseline. The post-hoc subgroup analysis is a major deficiency in the submissions and lacks several measures of scientific rigor including the potential impact of confounding factors. As such, the hypothesis that teriparatide has a superior treatment effect in patients with severe pre-existing vertebral osteoporosis is not adequately supported by the current submissions. This matter should be regarded as a fundamental requirement to the proposed listing.

b) Biological plausibility of difference between teriparatide and comparator
The key issues involved in this matter relate to the biological plausibility of an increased treatment effect with teriparatide compared with antiresorptive therapy (in particular, aledronate) on the basis of mechanism of action, as well as an increasing treatment effect over the spectrum of osteoporosis severity. There is sufficient evidence to support the sponsor claim of biological plausibility based on the anabolic mechanism of action of teriparatide and how such an action may be disproportionately effective in severe osteoporosis where there is significant alteration in the pathophysiology of osteoporosis favouring such a therapeutic action being effective. However, such a theoretical claim is not adequately supported by the current clinical outcome dataset.

c) Continued use of utility values
The validity of the incremental cost effectiveness ratios (cost per QALY) used in the submissions for teriparatide depend crucially on the validity of the QALY (utility) weights used in the cost utility analysis (CUA). The key issue in this matter is the validity of the continued use of the utility weights used in the CUA. This issue is not sufficiently justified in the submissions. The base case ICER (cost per QALY) should be based on the utility valuation of separate health states (symptomatic versus non-symptomatic) and an appropriate transition probability of entering that state. The principal criticism of the utility analysis is an inability to assess whether the sponsor's base case cost per QALY allows for the proportion of symptomatic versus asymptomatic vertebral fractures because of a flaw in the design of the utility survey. It follows that the sensitivity analysis, which does allow for the proportion of symptomatic and asymptomatic fractures and the disutility associated with symptomatic fractures, is plausible. The AQoL derived utility weights for vertebral fracture health states are not directly comparable to the utility weights associated with a hip facture as these weights are derived using a different utility measurement technique. A contributory factor to the principal criticism of the utility analysis is the lack of explicit guidelines from the PBAC during the submissions for this drug on the steps to identify, measure and value Quality-of-Life (QoL) outcomes for inclusion in a QALY-based cost utility analysis.

However, the PBAC appears to be addressing this deficiency with the recent posting of draft guidelines on this issue which are open to consultation. It is paramount that these guidelines are detailed and clear.

7 SUMMARY OPINION

The deficiencies in the cost utility analysis are amenable to correction with further work and analysis. However, the correction of the utility analysis will not overcome the inherent problems in the primary clinical outcome analysis, which forms the basis for the cost utility analysis. There is insufficient rigor in the clinical trial data analysis to recommend acceptance of the material presented in the submissions. In particular, the validity and robustness of the post-hoc subgroup analysis is a major deficiency. This results in an interpretation of unclear therapeutic benefit for teriparatide over comparator therapy.

SCOPE FOR REVIEW OF TERIPARATIDE SUBMISSIONS

Details of the Reviewers

The names of reviewers have been withheld, consistent with PBAC procedures

Timing of Review

The principal reviewer commenced the assessment on May 25, 2006. The secondary reviewer was contracted on June 2, 2006 to assist with the evaluation of the matter related to the health utilities. The final report was supplied to the Office of the Convenor on September 8, 2006 in both paper and electronic copy.

Matters for Review

The sponsor nominated the following issues upon which review is sought:

a) PBAC concern that the indirect data comparison and subgroup analysis does not support the claim of superiority-paragraph 4 of PBAC minutes.

b) PBAC claim that the submission provides no evidence to suggest any biological plausible reason to explain the claimed difference between teriparatide and the comparator-paragraph 9 of PBAC minutes.

c) PBAC concern with regard to continuing use of the same utility values in spite of the sponsor's efforts to address these concerns in its responses-paragraph 4 of PBAC minutes.

The following extract from the March 2006 PBAC short minutes provides the context in which the sponsor's request for review has arisen. "The PBAC accepted there is a strong clinical need for an effective treatment for patients with osteoporosis who continue to have symptomatic vertebral fractures whilst receiving a bisphosphonate, with calcium and vitamin D supplementation. The PBAC noted that this re-submission had addressed a number of outstanding issues, including the requested restriction, the appropriate comparator, the toxicity of teriparatide and the uncertainty over the predicted usage. However, some issues remained outstanding, including (a) the reliance on an indirect comparison across placebo-controlled trials to infer the superiority claim for teriparatide over alendronate rather than a head-to-head randomised trial; (b) use of the results of the post hoc sub-group analysis in place of the overall ITT results for teriparatide in the clinical conclusions and the economic model; and (c) the continuing use of the same utilities and disutilities, as previously, in the model where the sensitivity analyses indicate the model is sensitive to the assumptions used to derive the incremental utility estimates from the trial-based outcomes measures.

Furthermore, no evidence was submitted to suggest any biologically plausible reason that teriparatide is more effective in the high risk sub-group than in the overall trial population… As noted previously, the submission again sought to base its biological plausibility arguments on the different mechanisms of action of

the two alternative drugs. The PBAC concluded that although this might be relevant to the question of differential treatment effects across the drugs, it is irrelevant to the question here of the biological plausibility of the one drug (teriparatide) having an increasing relative treatment effect as baseline fracture severity worsens. The PBAC thus considered that the re-submission therefore provides no new basis that could change the PBAC's previous view concerning the invalidity of adopting the results of this group analysis rather than the results of the overall ITT analysis as the basis for deriving an estimate of the effectiveness of teriparatide to compare with alendronate."

Materials considered

The sponsor indicated (via the office of the Convenor for PBS Independent Review) that the review should consider the material provided at the March 2006 and July 2005 meetings of the Pharmaceutical Benefits Advisory Committee (PBAC).

The Convenor provided the following list of documents. The documents contain material provided by the sponsor, the material provided to PBAC, and the PBAC records.

March 2006 PBAC Meeting (labelled by sponsor as November 2005 submission)

Sponsor Submission

1. Cover letter
2. PB 11
3. TGA approval product information
4. Extract from PBAC Minutes July 2005
5. Answers to Key Questions
6. Executive Summary
7. Regulatory Documents
    - 7.1.1 TGA – Clinical Evaluation Report
    - 7.1.2 TGA – Request for pre-ADEC response and Delegates "Request for ADEC Advice"
    - 7.1.3 Eli Lilly – pre-ADEC Response
    - 7.1.4 ADEC Resolution
    - 7.1.5 Eli Lilly letter of 7 March 2003, responding to ADEC resolution
    - 7.1.6 Periodic Safety Update (provided in hard and electronic copy)
8. Main Body of Submission (labelled Volume 4 by sponsor)
9. Appendix to Submission (labelled Volume 4 by sponsor) (provided in hard copy and electronic copy)
10. References (labelled Volumes 5a and 5b by sponsor)
11. Teriparatide Study Reports (provided in electronic copy only except for GHAC)
    - 11.1.1 GHAC
    - 11.1.2 GHBJ
    - 11.1.3 GHAJ
    - 11.1.4 GHAH

11.1.5 GHBM
11.1.6 GHBQ
11.1.7 GHBU

12. Models and spreadsheets used in the submission – (provided in electronic copy only)
13. Previous sponsor submissions to PBAC (Note: did not form part of the March 2006 evaluation by PBAC) (provided in electronic copy only)
    13.1.1 Sponsor submission March 2003 (considered by PBAC June 2003)
    13.1.2 Sponsor submission December 2003 (considered by PBAC March 2004)
    13.1.3 Sponsor submission March 2005 (considered by PBAC July 2005)

PBAC Agenda Papers: Teriparatide – March 2006

14. Secretariat overview and attachments (white pages)
15. Pharmaceutical Evaluation Section (PES) commentary (pink pages)
16. Pre-Subcommittee response (blue pages)
17. Economics Subcommittee (ESC) Advice to PBAC (green pages)
18. Restrictions Working Group (RWG) Advice to PBAC (purple pages)
19. Pre-PBAC response (yellow pages)
20. July 2005 PES Commentary (routinely provided to PBAC Chair and discussants only, available to other members on request).
    (There was no Drug Utilization Subcommittee (DUSC) commentary or advice for this item).

PBAC Minutes

21. Relevant Extract of March 2006 Short Minutes
22. Relevant Extract of ratified PBAC Minutes of March 2006

July 2005 PBAC Meeting (labelled by sponsor as March 2005 submission)

Sponsor Submission

23. Cover letter
24. PB11
25. Answers to Key Questions
26. TGA product information
27. Preface
28. Executive Summary
29. Main Body of Submission (labeled Volume 4 by sponsor)
30. Appendix to Submission (labeled Volume 4 by sponsor) (provided in hard copy and electronic copy)
31. References (labelled Volume 6 by sponsor)
32. Models and spreadsheets used in the submission (provided in electronic copy only)

PBAC Agenda Papers: Teriparatide – July 2005

33. Secretariat overview and attachments (white pages)
34. Pharmaceutical Evaluation Section (PES) commentary
35. Drug Utilization Subcommittee (DUSC) secretariat commentary

36. Pre-Subcommittee response
37. Economics Subcommittee response
38. DUSC Advice to PBAC
39. Restrictions Working Group (RWG) Advice to PBAC
40. Pre-PBAC response
41. March 2004 PES Commentary (routinely provided to PBAC Chair and discussants only, available to other members on request).

PBAC Minutes
42. Relevant Extract of July 2005 Short Minutes
43. Relevant Extract of ratified PBAC Minutes July 2005

Sponsor's nomination of priority of material
The Convenor advised that the sponsor requested that the GHAC clinical study be considered essential and that the other clinical studies should be considered as necessary. Similarly, the sponsor advises that the references in the fourth submission (i.e. the March 2006 PBAC submission) are essential, and that the references in the July 2005 PBAC submission should be reviewed as necessary.

DESCRIPTION OF THE REVIEW PROCESS

Conduct of the review

The review was conducted in accordance with the administrative instructions and guidelines provided by the Convenor. It is understood the Convenor also provided copies of this information to the sponsor. The Convenor and reviewers agreed to regular scheduled time intervals for teleconferences and draft reports.

Attainment of additional material

In addition to the material supplied by the sponsor (outlined previously), a structured literature search was performed by identifying articles in three bibliographic computerized databases, MEDLINE, PubMed and EMBASE. The search was performed during August 2006 and was restricted to studies conducted with human patients and published in English language. The search terms used were “indirect data comparison”, “subgroup analysis”, and “teriparatide-mechanism of action”. Additional references were identified from the reference lists of the published articles.

The reviewers did not consider any new information or other information directly concerning the drug (teriparatide) that was not available to the PBAC. In particular, no unpublished clinical or scientific data was requested from the sponsor.

Consultation

No external consultation was agreed with the Convenor during the conduct of the review. The reviewers concurred that external consultation would not yield additional beneficial information to assist in the review process and was an inappropriate methodology to undertake if the independent nature of the review was to be maintained.

Recruitment and management of secondary reviewer

At the request of the principal reviewer, the Convenor was requested to recruit a secondary reviewer as the third matter of utility assessment was outside the expertise of the principal reviewer. The secondary reviewer was contracted on June 2, 2006 to assist with the evaluation of the matter related to the health utilities. The secondary reviewer received the same material from the sponsor (via the Office of the Convenor) as the principal reviewer.

The principal and secondary reviewer discussed the health utility matter (background information, facts and opinions) by weekly teleconference during July and August 2006. Interim written reports were also discussed.

MATTER 1 – INDIRECT DATA COMPARISON AND SUBGROUP ANALYSIS

Background

As no head-to-head clinical trials have been performed to evaluate the comparative efficacy of teriparatide and aledronate for the requested PBS listing (i.e. SQ3 patient subgroup who have sustained a painful fracture while being treated with an antiresorptive agent), a common comparator (i.e. placebo) and modelled evaluation were supplied by the sponsor to predict the relative outcomes and cost-effectiveness of the two therapies in this high risk patient group.

The data supporting teriparatide is derived from the GHAC trial which was a randomised, placebo-controlled trial of two doses of teriparatide (20ug/day and 40ug/day) using morphometric vertebral fracture as the primary efficacy outcome. Only data from patients on the approved dose of teriparatide (i.e. 20ug/day) and placebo was evaluated in the submissions. Subgroup analyses based on risk factors for future fracture were included as *apriori* analyses. The sponsor identified patients with SQ3 grade fractures using a combination of baseline risk factors, and then performed treatment by subgroup interaction analyses (post-hoc). The ITT results of the complete trial population have been published in a peer-reviewed medical journal (New England Journal of Medicine). In addition, the results from the post-hoc subgroup analysis of patients with prevalent vertebral fractures from the GHAC study have been published by Gallagher et al in The Journal of Clinical Endocrinology and Metabolism.

The data supporting aledronate in patients with established osteoporosis was derived from the FIT-VFA, an arm of the Fracture Intervention Trial which included a subgroup of patients with postmenopausal osteoporosis and a prior vertebral fracture. The trial has also been published in a peer-reviewed journal (The Lancet). Pre-planned subgroup analyses of the FIT-VFA suggest that the effect of aledronate is not attenuated by baseline risk factors. Consequently, the sponsor decided to use the ITT analysis in the common comparator evaluation. However, the submissions to PBAC by Eli Lilly used the full trial data for aledronate on the basis of a published report by Ensrud et al (1997) which was published in the Archives of Internal Medicine journal. The study concluded that aledronate compared with placebo resulted in an overall 47% reduction in the risk of new radiographic vertebral fractures. Treatment effect was uniform across a variety of predefined subgroups with all p-values >0.3 for the interaction between treatment (i.e. aledronate) and subgroup. The subgroups were defined at baseline by femoral neck bone mineral density, number of existing vertebral fractures, and history of postmenopausal fracture.

The choice of comparator had been an issue of conjecture in earlier submissions to the PBAC but appears to have been agreed upon in the final (fourth) submission. It is beyond the scope of this review to directly comment on the choice of an appropriate comparator. However, a brief relevant background to this issue is provided because the validity of the indirect data comparison is dependent on the choice of the comparator and how this choice was derived. Guidelines for submission to the PBAC of new drug listings indicate that if the proposed drug is in a new therapeutic class for which there are other drugs

widely used to treat that indication, then the main comparator will usually be the drug most widely prescribed on the PBS to treat that indication. As such, it was recommended by the PBAC and agreed to by the sponsor to have aledronate as the comparator in the fourth submission. However, in terms of available scientific data, the only agent for osteoporosis that has published evidence of efficacy in patients with severe (SQ3) vertebral fractures is raloxifene. This data is derived from a post-hoc analysis of the Multiples Outcomes of Raloxifene Evaluation (MORE) trial. In summary, both methods with respect to the appropriate choice of a comparator drug have merit and may be a potential subject of difference in opinion.

A summary of the key efficacy data derived from the GHAC study and the FIT-VFA trial which was presented in the fourth submission is shown in Table 1 (page 15). This data was also presented in the economic modelling evaluation.

The indirect data comparison using a common comparator is presented in Table 2 (page 16). As stated previously, there have been no clinical trials of an antiresorptive agent specifically designed and conducted in a SQ3 subgroup of patients. However, Delmas et al have conducted a post-hoc sub-group analysis of the MORE trial by semiquantitive grades. For the purpose of showing a complete comparative dataset, the result is compared with the GHAC study SQ3 subgroup and is also shown in Table 2. This information was used in a sensitivity analysis for the modelled economic evaluation.

Both tables were presented in Section 2.9 and Section 3 of the November 2006 sponsor submission and have been inserted into this review unchanged from that submission.

Table 1 Incidence of new Vertebral Fractures in Postmenopausal Women with osteoporotic fractures. Treatment with Teriparatide for 19 Months (ITT and SQ3 subgroup) and treatment with alendronate for 2 years 9 months.

	Placebo	Active	RR 95% CI	NNT
	n (%)	n (%)		
Teriparatide	N=448	N=444		
ITT analysis				
Number of patients with ≥ 1 new VFx (%)	64 (14.3%)	22 (5.0%)	0.347 (0.22, 0.55) p<0.001	11
Subgroup analyses				
BMD T-score ≤-3.0	41(21.24%)	12 (5.85%)	0.276 (0.15, 0.51)	6.50
BMD T-score ≤-2.5	50 (18.59%)	14 (4.98%)	0.268 (0.15, 0.47)	7.35
BMD T-score ≤-2.0	56 (16.14%)	18 (5.13%)	0.318 (0.19, 0.53)	9.08
Baseline VFx = 0	2 (4.0%)	0 (0.0%)		25.0
Baseline VFx = 1	9 (6.8%)	5 (3.4%)	0.502 (0.17, 1.46)	29.5
Baseline VFx ≥ 1	59 (15.01%)	22 (5.51%)	0.367 (0.23, 0.59)	10.5
Baseline VFx ≥ 2	34 (21.38%	11 (7.38%)	0.345 (0.18, 0.66)	7.14
VFx≥ 2 and T ≤-3.0	22 (27.16%)	6 (7.89%)	0.291 (0.12, 0.68)	5.19
VFx≥ 2 and T ≤-2.5	29 (27.62%)	7 (6.60%)	0.239 (0.11, 0.52)	4.76
VFx≥ 2 and T ≤-2.0	33 (24.81%)	10 (7.75%)	0.312 (0.16, 0.61)	5.86
SQ3 VFx	27/95 (28.42%)	5/86 (5.81%)	0.205 (0.08, 0.51)	4.42
Alendronate	N=1005	N=1022		
ITT analysis				
Number of patients with ≥ 1 new VFx (%)	145 (15%)	78 (8%)	0.53 (0.41 – 0.68)	14.29
Subgroup analyses				
≥ 2 VFx baseline BMD <0.59 g/cm2	28.3% (86)	14.6% (42)	0.52 (0.37 – 0.72) 0.54 (0.40-0.72)	7.3

Abbreviations and Notes: VFx= vertebral fracture; CI = confidence interval; SQ = semi-quantitative grading: SQ0 = Normal <20% reduction in vertebral height, SQ1= mild 20 – 25% reduction in vertebral height, SQ2 =moderate 25 - 40% reduction in vertebral height, SQ3 = severe ≥40% reduction in vertebral height

In response to the data displayed in Table 1, the sponsor claims that for the patient population proposed in the final PBS listing restriction, it is warranted to claim that teriparatide treatment results in a treatment effect of 0.20 (95% CI: 0.08, 0.51; p<0.001), or an 80% reduction in vertebral fractures in a severe (SQ3) subgroup of patients.

Additionally, the sponsor states “the improvement in RR is consistent with biological mechanisms of action associated with teriparatide and the pathophysiology of severe vertebral osteoporosis. However, the treatment by subgroup effects for alendronate remains consistent across patient subgroups,

therefore, in the absence of data to suggest otherwise, we suggest that the reduction in risk of new vertebral fracture in SQ3 patients treated with alendronate will be 0.53. A common comparator analysis demonstrates that the difference in RR associated with teriparatide in the SQ3 group and alendronate in an osteoporotic population of postmenopausal women is not zero (p=0.0487)."

Table 2: Indirect comparison of teriparatide and alendronate on incident vertebral fractures using a common comparator

Comparison	Teriparatide	Antiresorptive	Difference		
Vertebral Fx	RR	RR	Difference	RRR	P
Paired radiographs	0.347 (0.22, 0.55)	0.529 (0.14, 0.69)	-0.422 (-0.96, 0.11)	0.66 (0.38, 1.12)	0.1211
All randomised	0.346 (0.22, 0.55)	0.529 (0.41, 0.69)	-0.426 (-0.96, 0.11)	0.65 (0.38, 1.12)	0.1206
BMD T-score < -2.5 SD	0.268 (0.15, 0.47)	0.529 (0.41, 0.69)	-0.680 (-1.31, 0.06)	0.53 (0.28, 1.01)	0.0330
>2 VFx at baseline	0.345 (0.18, 0.66)	0.522 (0.37, 0.73)	-0.414 (-1.14, 0.31)	0.62 (0.30, 1.26)	0.2624
T-score and > 2 VfX	0.239 (0.11, 0.52)	0.529 (0.41, 0.69)	-0.794 (-1.62, 0.03)	0.46 (0.20, 1.08)	0.0583
SQ3 (paired)	0.187 (0.07, 0.48)	0.529 (0.41, 0.69)	-1.041 (-2.02, 0.06)	0.35 (0.13, 0.94)	0.0374
SQ3 (GHAC subgroup) vs ITT	0.205 (0.08, 0.51)	0.529 (0.41, 0.69)	-0.950 (-1.90, 0.01)	0.39 (0.15, 0.99)	0.0487
SQ3 subgroup (MORE)	0.205 (0.08, 0.51)	0.332 (0.26, 0.42)	0.730 (0.548, 0.971)	0.28 (0.11, 0.73)	0.009

The results of the indirect comparison (Table 2) do not indicate a treatment difference by statistical significance between teriparatide and aledronate for the overall treated population. However, when analysed by severity of disease there is evidence to possibly suggest that teriparatide is superior to aledronate in a subgroup of patients with more severe forms of vertebral osteoporosis at baseline. The result reaches statistical significance in the above presented analysis, though, the finding needs to be assessed for validity before acceptance as being clinically meaningful. In general, the statistical analyses are complex and there are several features of the data and analysis that are open to varying interpretation. This will be discussed in further detail later in the review.

As recorded in the minutes of the March 2006 PBAC meeting, the teriparatide results for the paired radiograph population and the SQ3 subgroup are summarized in Table 3. In addition, the respective heterogeneity tests from that document are presented in Table 4 (shown here unchanged).

Table 3: Results for vertebral fractures in paired radiograph population and SQ3 and sub-group of trial GHAC

Patients with a new vertebral fracture	**Teriparatide**	**Placebo**	**RD (95% CI)**	**RR (95% CI)**
Paired radiograph population (n=892)	22/444 (5.0%)	64/448 (14.3%)	-9.3% (-13.3,– 5.6)	0.35 (0.22, 0.55)
SQ3 sub-group population (n=181)	5/86 (5.8%)	27/95 (28.4%)	-22.6% (-34.5, -10.7)	0.20 (0.08, 0.51)

Table 4: Test for heterogeneity for effect sizes of sub-groups

Fracture sub-groups at baseline	Risk difference (95% CI)	Relative risk (95% CI)
SQ3 sub-group	-0.226 (-0.332, -0.122)	0.205 (0.084, 0.484)
Non-SQ3 sub-group (nil + mild + moderate)	-0.057 (-0.098, -0.019)	0.453 (0.261, 0.782)
Test for heterogeneity*	**p = 0.0027**	**p = 0.1433**
Entire trial results (for reference)	-0.093 (-0.133, -0.056)	0.35 (0.22, 0.55)

*Cochran Q test as calculated during the evaluation for the July 2005 PBAC meeting

Background to the matter arising

The sponsor described the matter as "PBAC concern that the indirect comparison and subgroup analysis does not support the claim of superiority – paragraph 4 of PBAC minutes." The relevant excerpt from the minutes read "However, some issues remained outstanding, including (a) the reliance on an indirect comparison across placebo-controlled trials to infer the superiority claim for teriparatide over alendronate rather than a head-to-head randomized trial; (b) use of the results of the post hoc sub-group analysis in place of the overall ITT results for teriparatide in the clinical conclusions and the economic model…"

The discussion in the March 2006 PBAC short minutes is as follows:
"Item 7.8.31 The PBAC noted the submission had appropriately chosen alendronate as the comparator. The Committee further noted that the submission had compared the results of the high risk group for teriparatide with the overall trial population for alendronate, in the absence of the available data for the high risk group for alendronate.

Item 7.8.32 The PBAC also noted that although the tests for interaction between baseline fracture severity and treatment effect indicate that there is a statistically significant difference in absolute risk between the high risk sub-group and the remainder of the trial period, as previously, there was no statistically significant difference for the relative risk (which is used in the modelled economic evaluation) across the two sub-groups. This statistical conclusion is unchanged from that reached previously, namely that the relative risk results of the overall trial population applied to the different baseline risk of the high-risk sub-group should form the basis of any clinical or economic evaluation. The re-submission again relies on a test for trend between the teriparatide and placebo populations in the key trial to support its claim for treatment effect modification. However, this test does not assess treatment effect modification on the relative treatment effect (or multiplicative scale) – which is relied upon in the model, or the absolute treatment effect (or additive scale) – which as concluded above would be expected to vary given that the baseline risk varies across the populations examined in the sub-groups."

Reviewer's understanding of the matter

The key issues involved in this matter appear to be two-fold. Firstly, the dependence upon an indirect comparison across placebo-controlled trials (rather than head-to-head studies) to infer the superiority of teriparatide over an appropriate comparator treatment. Secondly, the scientific validity of using the results of a post hoc subgroup analysis in place of the overall ITT results for teriparatide in the clinical conclusion. The two issues are fundamental to the robustness of the sponsor's scientific claim. Although the two issues are inter-related, at this stage of the review they will be discussed separately to optimize clarity of understanding.

Materials considered

All materials considered by the PBAC relating to the indirect data comparison and subgroup analysis were reviewed. In addition, key reference studies listed on pages 48-49 were used in forming an opinion.

1(A) – VALIDITY OF INDIRECT DATA COMPARISON BETWEEN THERAPIES

Additional background to the matter arising

In the March 2006 pre PBAC response, the sponsor stated "while we agree that indirect comparisons are not ideal, the indirect comparison method is not without benefits in cases where a head-to-head randomised controlled trial is not feasible. To ignore this comparison as providing only weak evidence ignores the following points:

- The FIT-VFA and GHAC trials have the same design
- The FIT-VFA and GHAC trials had the same inclusion/exclusion criteria
- The FIT-VFA and GHAC trials used the same primary endpoint to ascertain efficacy
- The patient populations in the FIT-VFA and GHAC trials were matched on all major risk factors for fracture – age, BMD, prior fractures."

In addition to tabled material already presented, Table 5 displays a summary of the key results for the indirect data comparison derived from the GHAC trial (full trial data and SQ3 subgroup for teriparatide) and the FIT-VFA trial (for aledronate). This data has been reproduced without amendment from the agenda papers of the March 2006 PBAC meeting.

Table 5: Indirect analysis of morphometric vertebral fracture rates using placebo as common reference

	Teriparatide data			Alendronate data		
	Active	**Placebo**	**RR (95% CI)**	**Active**	**Placebo**	**RR (95% CI)**
1 – Teriparatide & alendronate: all randomised populations						
	22/541	64/544	0.346 (0.22,0.55)	78/1022	145/1005	0.529 (0.41, 0.69)
Risk difference (indirect comparison) = -0.426 (-0.96, 0.11)						
Relative risk reduction (indirect comparison) = 0.65 (0.38, 1.12), p=0.1206						
2 – Teriparatide & alendronate: paired radiograph populations						
	22/444	64/448	0.347 (0.22, 0.55)	78/981	145/965	0.529 (0.41,0.69)
Risk difference (indirect comparison) = -0.422 (-0.96, 0.11)						
Relative risk reduction (indirect comparison) = 0.66 (0.38, 1.12), p=0.1211						
3 – Teriparatide: SQ3 sub-group population; alendronate: paired radiograph population						
	5/86	27/95	0.205 (0.08, 0.51)	78/981	145/965	0.529 (0.41, 0.69)
Risk difference (indirect comparison) = -0.950 (-1.90, 0.01)						
Relative risk reduction (indirect comparison) = 0.39 (0.15, 0.99) p=0.0487						

NB: Comment included in the March 2006 PBAC short minutes "The calculation of the risk differences for the indirect comparison of teriparatide and alendronate could not be reproduced and the values are implausibly large."

Reviewer's opinion

The preferred method to assess the comparative effects of medical interventions is head-to-head randomized controlled trials (RCT). Concerns have been expressed over the use of indirect comparisons of treatments. The Cochrane Collaboration's guidance to authors states that indirect comparisons are not randomized, but are "observational studies across trials, and may suffer the biases of observational studies, for example confounding." Some investigators suggest that indirect comparisons may systemically over-estimate the effects of treatments as "randomization is not sufficient for comparability".

However, in the absence of direct comparator data, other types of data evaluations may be considered but these methods are scientifically less robust than head-to-head RCTs. There is emerging evidence in the literature that when no head-to-head evidence is available, the method of "adjusted indirect comparisons" may be useful to estimate the relative efficacy of competing interventions. Bucher and colleagues developed a model for making adjusted indirect comparisons of the magnitude of treatment effects that produces an unbiased estimate of relative efficacy of the treatments.

In fact, several recent papers have demonstrated that results from adjusted indirect comparisons usually, but not always, agree with the results from head-to-head (direct) comparisons. For example, a recent paper by Song et al, demonstrated that only 3 of 44 treatment comparisons showed significant discrepancy ($p<0.05$) between the direct results and the adjusted indirect estimate. The categories of patients involved in these studies covered a diverse range of medical conditions including those with an increased risk of vascular occlusion, HIV infection, chronic Hepatitis C virus infection, gastro-oesophageal reflux disease, post-operative pain, heart failure, and cigarette smoking.

However, the methodological process involved in the indirect comparison (i.e. when two or more interventions are compared through their relative effect versus a common comparator) is crucial. Some authors have used a naïve (unadjusted) indirect comparison, in which the results of individual treatment arms between different trials are compared as if they were from a single trial. Simulation studies and empirical evidence indicates that the naïve indirect comparison is liable to bias and produces over-precise estimates. As such, the naïve indirect comparison should be avoided. However, suitable statistical methods for comparing multiple treatments that fully respect randomization have been available for some time. They have not been widely used, although their application is increasing in USA based medical journals and in medical decision making. In this particular matter, the sponsor states it has applied the methods described by Bucher et al and Song et al to the common comparator analysis of the GHAC and FIT-VFA trials in deriving the data on the primary and secondary fracture endpoints. The workings of this analysis were not located in the submissions but can be assumed to be performed correctly.

In addition to the statistical process, for the adjusted indirect comparison method to be valid an assessment of study characteristics (in particular, similarity and internal validity) that are related to the exchangeability of results across trials, such as patient characteristics, methodological quality, endpoint definitions,

outcome measures, and adherence rates need to be considered. In general, the trial designs, primary endpoint, patient inclusion and exclusion criteria, patient demographic and baseline characteristics, and fracture rates in the placebo groups are comparable between the GHAC and FIT-VFA studies. This provides a reasonable level of assurance that the results are valid when compared via a common comparator analysis. However, there are some issues with the GHAC trial that detract from both the internal validity and generalisability of the data. The GHAC study was terminated after a median of 19 months of treatment (mean 18 months; maximal treatment period 2 years and 20 days) because a long term carcinogenicity study in rats revealed the occurrence of skeletal proliferative lesions, including osteosarcoma. This finding was later determined to be unlikely to have significant predictive ability in humans but nonetheless, may introduce a bias into the comparison. The study had originally been designed to continue for 3 years. The sponsor believes that comparing teriparatide where the active treatment phase for teriparatide is much shorter (19 month median) to comparator (mean follow-up of 2.9 years for aledronate in the FIT-VFA study) is biased against teriparatide. This interpretation is probably true.

However, other potential sources of bias related to patient disposition, trial design, and analysis were observed. Vertebral fracture analyses were based on 801 (placebo=398; teriparatide 20ug/day=403) of the 1085 (73.8%) women who were originally randomized to the two treatment arms of interest in the GHAC trial. There were multiple reasons for loss of evaluable subjects. A significant proportion (17.7% in total; 193/1085) of patients (placebo=96; teriparatide=97) lacked an adequate baseline or follow-up radiograph. Some placebo treated (n=50) and teriparatide treated (n=41) subjects were subsequently judged by the central reader to not have prevalent vertebral fractures. The study protocol required that patients have one or more prevalent vertebral fractures at screening as assessed by the investigative site but clearly a discrepancy arose in some patients for this assessment. Because the vertebral fracture status of these patients was unclear their data was excluded from the analysis. The sponsor does not believe these missing values impacted upon the results as baseline characteristics were not statistically different between placebo and teriparatide treated subjects. Moreover, prior to the sponsor's decision to terminate the GHAC trial, 18.9% (205/1085) patients discontinued from the study. In addition, the central radiograph readers were not blinded to radiograph temporal sequence but were unaware of patient treatment assignment. It is also noteworthy that neither study (GHAC or FIT-VFA) limited enrolment to patients with SQ3 grade (severe) vertebral fractures at baseline which would be a desirable characteristic of the supporting clinical evidence.

Furthermore, because the severe vertebral fracture subgroup is such a limited subpopulation in total number (n=86 for teriparatide, n=95 for placebo) and the absolute event rate of new morphometric fractures is relatively small (n=5 for teriparatide, n=27 for placebo) this result may be prone to over-interpretation in the setting of low statistical power.

Reviewer's summary

The sponsor's use of an indirect data comparison across placebo controlled trials to infer the superiority of teriparatide treatment over aledronate is scientifically less robust than the gold standard methodology of head-to-head randomized controlled studies. The sponsor has undertaken the appropriate statistical measures to optimize the validity of the indirect data comparison. However, the validity of the adjusted indirect comparisons depends on the internal validity and similarity of the trials involved. Some concerns remain regarding the internal validity and generalisability of the pivotal GHAC study which include an unanticipated premature termination of the study, and the significant loss of evaluable subjects due to patient drop-out and the lack of appropriate paired radiographs. As such, the data is open to varying interpretations which clearly detract from the scientific robustness of the treatment claim.

1(B) – VALIDITY OF SUBGROUP ANALYSIS IN PIVOTAL STUDY

Additional background to the matter arising

The sponsor has said "The analysis in the table (presented on page 7.8. ESC ADV 4) is the same post-hoc analysis that was presented in the previous submission. In relative terms, the results are no longer different between the SQ3 sub-group and the non-SQ3 sub-group (7.8. ESC ADV 4). We are again frustrated that a post-hoc analysis completed during the PES evaluation of the previous submission which was judged to be ambiguous during that evaluation is again presented in the ESC advice as unequivocal. This analysis is based on summary data. We have not replicated this analysis in the submission as we have already presented a treatment effect analysis conducted using the patient level data. These data presented in both the presented and previous submission, also presented during the hearing for the previous submission, demonstrated a significant treatment effect.

The Cochran-Armitage test for trend demonstrated that baseline severity of fracture was a predictor of subsequent fracture in the placebo group, this effect was not significant for the teriparatide treated group demonstrating that teriparatide breaks the cycle of accelerating fracture risk associated with fracture severity. For teriparatide 20 ug/day, the Armitage Trend Test Statistic = 0.7909, p-value = 0.429, for patients treated with placebo the Armitage Trend Test Statistic = -3.6764, p-value = 0.002."

In the re-submission for consideration at the March 2006 meeting of the PBAC, the sponsor states "In discussions with the Chair of the PBAC following the rejection of the previous submission, it was suggested that the PBAC appreciate the lack of ideal comparative data for the proposed listing, but would consider a scientific and clinically valid chain of logic for the proposed positioning and using alendronate as the comparator. This logic is critiqued in the PES commentary at 7.8 PES COM.19 and therefore this response focuses primarily on the response to this critique.

1. The sub-group analysis was not pre-specified in the statistical plan and the statistical evidence for interaction is weak. For instance, an earlier submission (March 2004) presented data on teriparatide treatment effects in the full trial population stratified by baseline vertebral fractures. The evidence for a trend to greater protective effect higher numbers of baseline fractures is unconvincing and if anything the trend is in the opposite direction when non-vertebral fractures are the outcome."

The following is the sponsor response "Eli Lilly Australia agrees that the analysis of fractures by baseline severity was not pre-specified in the statistical analysis plan written in 1994. The GHAC trial was essentially developed to fulfill regulatory purposes. Semiquantitive (SQ) analysis of fractures was not common clinical practice nor was it required as standard by the regulatory agencies. However, SQ grading was included in all protocols for osteoporosis agents (including alendronate, raloxifene and risedronate) but to date only analysed and published for raloxifene. Due to evidence gained since the study of the new

treatments, SQ grading of fractures has become more common in research and clinical practice and while low BMD and number of baseline fractures are recognized as independent risk factors for new fracture, evidence from analyses of the placebo group from the Multiple Outcomes of Raloxifene Evaluation (MORE) study demonstrated that baseline fracture severity is the most important predictor of incident fracture."

Our analysis of the baseline characteristics of SQ3 patients in both the GHAC and MORE studies, as well as the characteristics of SQ3 fracture patients in a large scale European epidemiological study, demonstrates that these patients can be characterized as having both low BMD (within the osteoporotic range) and multiple vertebral fractures at baseline (see Section 2.5.3 of the submission). In the GHAC trial, these two distinct analyses were pre-specified and demonstrate that as the severity of disease increases (based on the risk factors of BMD and number of vertebral fractures), the reduction in risk of new vertebral fracture associated with teriparatide is improved.

Furthermore, the incidence of new moderate (SQ2) or severe (SQ3) fractures was a pre-specified analysis. No new SQ3 fractures were recorded during the GHAC trial, the reduction in risk of new moderate or severe fractures for the ITT sample of the GHAC trial is 0.096 (0.035, 0.266) (see Table 37 of the submission). That is, a 90% reduction in risk of new SQ3 fractures.

The proposed listing is for patients with severe vertebral fractures and focuses on these outcomes. The GHAC study was conducted over a median 19 months compared with 36 months for the FIT-VFA. Due to the low incidence of non-vertebral fractures over the 19 months of the trial the GHAC is not adequately powered for measuring trend in non-vertebral fractures by SQ sub-groups."

Reviewer's understanding of the matter involved

A further issue of concern identified in the July 2005 and March 2006 submissions was the validity of the subgroup analysis in the pivotal teriparatide trial, GHAC, that consists of a patient subpopulation who had a prevalent baseline vertebral fracture that is severe (SQ3), meaning at least a 40% vertebral height reduction. The subgroup of patients defined as having a severe prevalent vertebral fracture in the GHAC trial comprised 20.3% (181 of 892 patients) of the total evaluable (i.e. those patients with paired radiographs) trial cohort. Subgroup analyses based on risk factors for future fracture were included as *apriori* analyses. The sponsor identified patients with SQ3 grade fractures using a combination of baseline risk factors, and then performed treatment by subgroup interaction analyses (post-hoc). The criteria for patient identification are objective and can be documented using radiological evaluation. The results from the post-hoc subgroup analysis of patients from the GHAC study have been published by Gallagher et al (2005).

An additional issue in the fourth submission is the disparity between the proposed listing for teriparatide and the direct trial data from the patient subgroup analysis. The proposed listing states that the qualifying fracture needs to be painful and appears to infer that all severe vertebral fractures are symptomatic.

However, the GHAC trial used new morphometric (not necessarily clinically apparent) vertebral fractures as its primary outcome. The sponsor states in its latest submission that this patient subgroup with severe osteoporosis is readily identifiable from a clinical perspective, presenting with characteristic compression, crush or wedge fractures of the spine, and presumably pain. However, this assumption is not supported by data that accurately quantifies the proportion of clinically apparent patients in this particular subgroup. In the sensitivity analysis, the sponsor has given a figure of up to 70% of patients with vertebral fractures being asymptomatic. This is accurate for a patient population with any grade of vertebral fracture but is not specific to patients with SQ3 grade vertebral fractures. The methods used to adjust for the proportion of painful fractures (as per the requested listing) in the sensitivity analysis is discussed at a later stage of this review (refer to Matter 3-page 34). In addition, it remains unclear what proportion of patients in the severe subgroup would be accurately identified on radiological grounds. In the GHAC study itself, there was discrepancy between the investigative site and central radiographic reader on the presence of any grade of prevalent vertebral fracture in 8.4% (91/1085) patients. It is reasonable to expect that outside of the controlled environment of a formal trial, discrepancy rates in the correct identification of baseline radiological criteria would at best be equivalent in the non-trial clinical setting.

Reviewer's opinion

Subgroup analyses attempt to identify a subset of patients who derive a greater or lesser benefit from therapies than does the average trial patient. Because the sponsor in this matter is requesting a restricted listing in a patient subgroup thought to derive the most benefit from teriparatide, then it is appropriate for a subgroup analysis to be considered as part of the submissions. Moreover, I believe it is justifiable because of the potential plausibility of an increased treatment effect with severity of disease and stage in the natural history of osteoporosis. This particular issue of potential heterogeneity of treatment effect in relation to pathophysiology will be discussed further in Matter 2.

However, the methodology of the subgroup analysis needs to be scrutinized to ensure the validity of the process, appropriate interpretation, and application to everyday clinical settings. Several prominent authors on analysis of clinical trials (Oxman and Guyatt-page 49) have proposed criteria that should be satisfied before accepting results from subgroup analyses. The criteria are outlined below and are relevant in assessing this particular matter:-

- *Clinically significant*: the magnitude of the differences between treatment groups are clinically important and would lead to different decisions for different subgroups.
- *Statistically significant*: the differences remain statistically significant after formally testing for treatment-subgroup interactions using appropriate statistical methods.
- *A priori hypothesis*: the hypothesis of subgroup differences preceded rather than followed the analysis (i.e. a priori hypothesis pre-specified in the trial protocol, not a discovery made from post-hoc analyses).
- *Limited number of comparisons*: the subgroup analysis in question was one of a small number of hypotheses tested to minimize the number of

seemingly significant differences (i.e. interactions) that could simply occur by chance.

- *Within study comparisons*: subgroup differences were suggested by comparisons within studies (i.e. direct comparisons) rather than between studies (indirect comparisons)
- *Reproducibility*: the subgroup difference is reproducible in other studies that have adequate power and are of similar design in terms of patient characteristics, co-interventions and outcome measures.
- *Supporting evidence*: the subgroup difference is biologically compelling and consistent with current understanding of biologic mechanisms of disease.

When the data presented in the submissions derived from the GHAC trial is evaluated against the above criteria, the subgroup analysis arguably satisfies some but clearly not all of the seven elements. Without doubt the criteria of *apriori* hypothesis and reproducibility have not been met and these criterion should be fundamental to the scientific strength of the data supporting the proposed PBS listing.

The sponsor's submissions of July 2005 and March 2006 use a post-hoc subgroup analysis of the GHAC trial data. Although the PBAC does not currently have submission guidelines that transparently represent their opinion weight on the validity of certain types of clinical data analyses, it is clear from the medical literature that post-hoc subgroup analyses, in general, are relatively poor options in terms of scientific rigor and validity. Post-hoc analyses of certain subgroups are potentially fraught with multiple hazards, especially the play of chance and uncontrollable confounders. They are particularly prone to false findings and indeed, the medical literature has numerous examples from randomized trials in which an apparently important differential response to therapy suggested by subgroup analysis generated a hypothesis that was subsequently refuted in a trial designed to test that hypothesis (references such as Parker, Pfeffer, Rothwell and Scott-page 49). In addition, respected authors on the topic of subgroup analysis such as Rothwell report that simulations of randomized controlled trials powered to determine the overall effect of treatment suggest that false subgroup treatment effects occur by chance in 7%-21% of analyses. The same author states "post-hoc observations should be treated with skepticism irrespective of their significance" and that "no test of significance is reliable in this situation."

The sponsor did perform an appropriate statistical test (i.e. to test for a subgroup-treatment effect interaction) when presenting the subgroup data. However, the best test for validity of subgroup-treatment effect interactions is not statistical significance but reproducibility in other trials. The Cochran-Armitage trend test is sensitive to linearity between response and covariates. It tests for trends in binomial proportions across levels of either a single factor or a covariate. In the subgroup data analysis, the sponsor applied this test to determine the relationship between prevalent vertebral fracture grade and incident vertebral fractures in both the placebo and teriparatide treated groups. In the submissions, the sponsor refers to this analysis (i.e. a test for trend) to support its claim for treatment effect modification with teriparatide.

The data for the paired radiograph population and the SQ3 subgroup cohort of the GHAC trial is summarized in Table 3 (page 16) and the test for heterogeneity for effect sizes of subgroups is presented in Table 4 (page 17). In particular, it is noteworthy that the PBAC tested for heterogeneity (treatment effect modification) by performing a Cochran Q test. Statistical tests for heterogeneity seek to determine whether or not there are genuine differences underlying the results of studies. This analysis is usually evaluated by Cochran's Q test.

Treatment effect in terms of new vertebral fractures within the subgroups was assessed by the sponsor using a likelihood ratio test based on a logistic regression model. In the fourth submission, the sponsor reports the analysis of the GHAC data demonstrating the following treatment interaction effects via logistic regression:-

- Treatment (20ug/day vs placebo), SQ grade (1,2,3), and the interaction (saturated model): p=0.2239 (i.e. not statistically significant)
- Treatment (20ug/day vs placebo), SQ grade (1 and 2 combined vs 3), and the interaction (saturated model): p=0.0865 (i.e. "significant at the 0.10 level")

The statistical analyses are complex and some aspects of both the analysis and results are open to varying interpretation. However, in my opinion, the analysis of the data presented in Table 4 (page 17) is the most scientifically valid means of assessing a treatment-subgroup interaction. Testing for heterogeneity in terms of treatment effect modification is optimally evaluated by the Cochran Q test, but all such tests are statistically underpowered in this subgroup analysis dataset. As such, any results suggesting a treatment effect modification are only weakly supported by the data. For the SQ3 subgroup, the test for heterogeneity on the risk difference is statistically significant in terms of further fractures (i.e. 22.6% versus 5.7% respectively, p=0.0027). Hence, there is a statistically significant difference in absolute risk between the high risk subgroup and the remainder of the trial population. This is an expected finding given the different background prognostic factors for future fracture events between the subgroup populations. There is convincing evidence in the medical literature that prior osteoporotic fractures (both number and severity) independently predict the risk for new vertebral fractures. Moreover, the background (i.e. placebo treatment group) rates of fracture in the SQ3 subgroup populations are quite different for the two studies. For the placebo treated SQ3 subgroup in the GHAC study, the rate of incident vertebral fracture was 28.4% (27/95) and for the placebo treated SQ3 subgroup in the FIT-VFA, the rate of incident vertebral fracture was 15% (145/965) -refer to section 3 of Table 5 (page 19).

However, the key result for data interpretation of treatment effect was the absence of a statistically significant difference (p=0.1433) for the relative risk between the two baseline fracture subgroups of severe (SQ3) and the "non-severe" (non-SQ3) subgroups on the formal test for heterogeneity.

Guidelines on the analysis and reporting of subgroup analyses recommend reporting both absolute and relative risk reductions, with tests for subgroup-treatment effect interaction also to be performed. In this particular matter, given the heterogeneity of treatment effect is likely to be plausibly related to varying

absolute risks (with or without treatment) within the different sub-populations at baseline, relative risk reductions with treatment is the most informative means of assessing treatment effect across populations. As such, in my opinion, the relative risk reduction of the overall trial population should be applied to the different baseline risk of the SQ3 subgroup in deriving the clinical outcome data.

In addition, potential sources of bias related to methodology and patient disposition were also observed as per the study report published by Gallagher et al. These have already been outlined in Matter 1A (page 19).

The pivotal GHAC trial was conducted in 1996-1997 and the analysis of fractures by baseline severity was not pre-specified in the statistical analysis plan which was written in 1994. Unfortunately there have been no randomized controlled trials of teriparatide performed since that time to directly assess the validity of the proposed listing. The medical literature indicates that genuine unanticipated subgroup-treatment effect interactions are uncommon. In addition, part of the expected process in designing clinical trials for regulatory purposes is to consult expert clinical opinion in the development phase with the aim of pre-defining potentially important subgroups. In addition, if relative treatment effect is likely to be related to baseline fracture risk, then the analysis plan should include stratification by predicted risk and be statistically powered to determine any such treatment effect.

Reviewer's summary The result derived from post-hoc subgroup analysis of the GHAC trial suggesting a possible but highly contestable and unclear benefit with teriparatide treatment in patients with more severe forms of vertebral osteoporosis (SQ3) at baseline should be interpreted as informative and hypothesis-generating. The current dataset and analysis is deficient in several measures of scientific rigor that are open to varying interpretation and may be impacted by confounding factors. As such, the hypothesis that teriparatide has a superior treatment effect in patients with severe pre-existing vertebral osteoporosis is not adequately supported by the current submissions. This matter is a fundamental requirement to the proposed listing. Confirmation in subsequent well designed clinical trials is recommended before the proposed listing is accepted on the basis of a scientifically robust treatment effect.

MATTER 2 – BIOLOGICAL PLAUSIBILITY OF DIFFERENCE BETWEEN TERIPARATIDE AND COMPARATOR

Background to the matter arising

The sponsor raised the following in its requesting letter for review "PBAC claim that the submission provides no evidence to suggest any biologically plausible reasons to explain the claimed difference between Teriparatide and the comparator – paragraph 9 of the PBAC minutes."

The relevant excerpt from the March 2006 PBAC short minutes is contained within paragraph 5, (not paragraph 9), and reads "Furthermore, no evidence was submitted to suggest any biologically plausible reason that teriparatide is more effective in the high risk sub-group than in the overall trial population. Although a biologically plausible argument could be mounted that alendronate may be less effective in the high risk sub-group because of a lower bone turnover associated with loss of bone architecture in this sub-group, as noted above, the submission did not rely on using the results of a sub-group analysis in place of the overall trial population results for alendronate in its modelled economic evaluation. As noted previously, the submission again sought to base its biological plausibility arguments on the difference mechanisms of action of the two alternative drugs. The PBAC concluded that although this might be relevant to the question of differential treatment effects across the drugs, it is irrelevant to the question here of the biological plausibility of the one drug (teriparatide) having an increasing relative treatment effect as baseline fracture severity worsens. The PBAC thus considered that the re-submission therefore provides no new basis that could change the PBAC's previous view concerning the invalidity of adopting the results of this sub-group analysis rather than the results of the overall ITT analysis as the basis for deriving an estimate of the effectiveness of teriparatide to compare with alendronate."

The relevant extract from the sponsor response to the ESC and RWG advice is: "We are somewhat frustrated by the lack of acknowledgement given in the ESC advice relating to this issue. The submission includes reference to a number of clinical, pre-clinical, histomorphological, biological studies and modelling studies presented to support the plausibility that an anabolic agent which stimulates osteoblasts and thus builds new bone would produce superior outcomes to an anticatabolic agent, and that this effect may be moderated by the existing microarchitecture of the bone. We provided independent references relating to the action of anabolic and anti-catabolic drugs on bone, references relating to the structure of bone in mild, moderate and severe patients and evidence demonstrating the differential effects of teriparatide on depleted bone in humans and animals. We have sought independent advice on this issue prior to the submission and believe that this has been addressed "beyond re-stating" that the drugs have a different mechanism of action. We therefore urge the PBAC to discuss this with members of the scientific and medical community. Without further explanation from the evaluator as to why these multiple studies are so easily rejected we cannot further address this issue."

Reviewer's understanding of the key issues

The key issues involved in this matter appear to relate to the biological plausibility of an increased treatment effect with teriparatide compared with antiresorptive therapy (in particular, aledronate) on the basis of mechanism of action, increasing treatment effect over the spectrum of osteoporosis severity, and the presence of clinical data to support such a claim.

Materials considered

All materials considered by the PBAC relating to the assessment of biological plausibility were reviewed. In addition, key reference studies listed on page 49 were used in forming an opinion.

Reviewer's opinion

An issue repeatedly identified in the July 2005 and March 2006 submissions was the biological plausibility of an increasing treatment effect with teriparatide over comparator in patients with more severe forms of vertebral osteoporosis. The sponsor claims that vertebral fracture severity is a treatment effect modifier and that an agent such as teriparatide which stimulates osteoblasts as opposed to antiresorptive agents which maintain existing bone is likely to have an enhanced treatment effect in patients with more severe forms of osteoporosis.

Pharmacological therapies for osteoporosis may be classified as either anti-catabolic (antiresorptive) or anabolic depending on their mode of action. Most currently available drug treatments for osteoporosis are anti-catabolic in action and include hormone therapy with oestrogen receptor modulating effects (e.g. raloxifene) and the bisphosphonates (e.g. aledronate and risedronate). Anti-catabolic drugs reverse the negative bone balance seen in osteoporosis by decreasing the associated increase in bone remodelling. They do this by decreasing the number of bone multicellular units (BMU), and hence preferentially decrease bone resorption. The reduction in bone remodelling by anti-catabolic drugs may increase bone mass moderately during the interval in which previously activated BMUs are still completing mineralization. As such, in the period of therapy before bone formation decreases, bone balance becomes positive, but thereafter remains in equilibrium explaining why bone mineral density increases are typically largest during the first two years of therapy with anti-resorptive agents.

The skeletal effects of parathyroid hormone (PTH) depend upon the pattern of systemic exposure. When used in an intermittent or cyclical manner, PTH can be classified as an anabolic therapy for osteoporosis. In contrast, a continual excess of endogenous PTH, as seen in hyperparathyroidism, may be detrimental to the skeleton as bone resorption may be stimulated more than bone formation. Teriparatide is the active fragment (1-34) of endogenous human PTH manufactured using recombinant DNA technology. It is the most extensively studied form of PTH in patients with osteoporosis. The intermittent administration of PTH acts in two principal ways to induce a positive bone balance that persists in the medium term. Firstly, activation of the BMU results in a preferential increase in bone formation over bone resorption (i.e. augmentation of osteoblast function). Secondly, there is also increased conversion of bone lining cells to osteoblasts. This response has been demonstrated in cell culture systems, in

intact animals, and in humans. It is this property of the proposed administration of teriparatide that increases bone formation and results in a positive bone balance. In addition, teriparatide induces renewed modelling which is an efficient way to increase bone mass, because the formation of the BMU is not preceded by a resorption phase. Other effects of teriparatide include improvements in trabecular micro-architecture, and cortical bone geometry and thickness. In contrast, anti-catabolic agents do not increase bone tissue mass or trabecular bone volume. Instead, these agents increase the tissue mineral content of density present in the existing bone tissue mass.

The different mechanisms of action between the two drug classes raise the theoretical possibility of an enhanced treatment effect in favour of anabolic agents in patients with more severe forms of osteoporosis, particularly involving skeletal sites comprised of predominately high-turnover trabecular bone. In healthy individuals, bone is constantly remodelled through the process of active formation and resorption which occurs in a close temporal sequence. The hallmark of osteoporosis is a reduction in skeletal mass with an associated micro-architectural disturbance caused by an imbalance between bone resorption and bone formation. Loss of gonadal function and ageing are the two most important elements contributing to this imbalance. Bone is made up of two types: cortical and trabecular. Cortical bone comprises 80% of all bone and is found mainly in long bones. Bone loss that occurs primarily in cortical bone results in fractures of the proximal femur and humerus. Trabecular bone is predominately found in flat bones (vertebrae, sternum, and pelvis) and the metaphyses of long bones. Thus, when trabecular bone is most affected by bone loss, vertebral, pelvic and Colles' fractures of the forearm are manifestations of this pattern of osteoporosis. In healthy adults, 25% of trabecular bone is resorbed and replaced every year, compared with 3% of cortical bone, which indicates the rate of remodeling and hence susceptibility to the pathophysiological changes of osteoporosis is different between the different types of bone. Vertebrae are principally composed of trabecular bone encased in a thin shell of cortical bone. Some underlying medical conditions or risk factors predispose certain individuals to one particular pattern of osteoporosis. For example, the use of corticosteroids may be associated with rapid bone loss, especially in high-turnover trabecular bone, and may lead to fractures in predominately trabecular bone sites such as the vertebrae and ribs. However, for most individuals with osteoporosis there is no differential loss of bone between the two types of bone, i.e. osteoporosis is a systemic skeletal disease and the process affects all bony sites.

The pathophysiology of osteoporosis does appear to alter with increasing severity of osteoporosis with bone density (quantity) and bone architecture (quality) becoming disproportionately disturbed. Whereas it has been shown that bone mass is the major determinant of bone strength, the mass-based concept does not fully account for the contribution of bone microarchitecture to mechanical efficiency and ability to withstand fracture at normal physiological stress levels. The pivotal paper by Parkinson and Fazzalari on this topic was included by the sponsor in the fourth submission. This study showed that changes to cancellous bone structure are bone volume-dependent in a nonlinear manner. Specifically, at low bone volume (<15%), structural parameters of

cancellous bone such as trabecular thickness and trabecular separation change at a much greater rate than at higher bone volume. This indicates that the structural integrity of cancellous bone may become rapidly compromised when bone volume falls below a critical value.

Anti-catabolic therapies such as bisphosphonates may be less effective in the high risk subgroup of patients with severe bone loss, particularly in predominately higher-turnover trabecular bone sites such as vertebral body osteoporosis. This hypothesis appears to be supported by in vivo histomorphometric data favoring the differential effect of teriparatide compared to aledronate in postmenopausal women. This theory (supported by scientific data) is relevant in considering a differential treatment effect across the two drug classes used in osteoporosis. Anti-catabolic drugs work by inhibiting bone resorption and reducing the chance that trabecular plates will perforate. In patients with more severe grades of osteoporosis, bone structure is lost, thus abnormal stress is placed on the remaining structures which increases the possibility of damage. In response, bone turnover is stimulated, but osteoblastic bone formation increases at an insufficient rate, so overall bone mass decreases and trabecular architecture diminishes. Anti-catabolic drugs can slow this cycle and try to maintain bone strength. However, in patients with severe vertebral osteoporosis with fractures, the bone structure is damaged to the threshold where maintenance of the existing bone is insufficient and bone formation is required.

Because of the mechanism of action of teriparatide, it is conceivable that this drug may have an increasing relative treatment effect as baseline severity risk of fracture worsens. There are several studies (supplied by the sponsor) that show that teriparatide induces beneficial changes in bone quality, structural architecture and geometry. These attributes are particularly advantageous in more severe forms of vertebral osteoporosis. As such, it is theoretically conceivable that teriparatide has an increasing relative treatment effect as baseline severity risk of fracture worsens because of the beneficial inter-relationship between the anabolic mechanism of action of the drug and the pathophysiology of severe vertebral osteoporosis. Consequently, on the basis of biological plausibility, I would disagree with the PBAC statement that the different mechanisms of action of the two alternative drug classes is irrelevant to whether or not a single drug (i.e. teriparatide) has an increasing relative treatment effect as baseline severity risk of fracture worsens.

Despite the theoretical plausibility, the submissions did not contain convincing clinical outcome data to support such a hypothesis. Data from the pivotal GHAC study did not conclusively demonstrate an enhanced treatment effect in the severe fracture (SQ3) subgroup compared to the overall Intention-to-Treat (ITT) analysis. Using the overall ITT data (Table 1- page 15), new vertebral fractures occurred in 64 of 544 (14.3%) of women in the placebo group compared to 22 of 541 (5.0%) of women treated with teriparatide 20ug/day (relative fracture risk 0.35, 95% CI 0.22 to 0.55; $p<0.001$). Using the SQ3 subgroup data from the GHAC study, new vertebral fractures occurred in 27 of 95 (28.4%) of women in the placebo group compared to 5 of 86 (5.8%) of women treated with teriparatide 20ug/day (relative fracture risk 0.205, 95% CI 0.08 to 0.51). Because the 95% confidence intervals for each type of population analysis lie within each other, the

treatment effect by population should be interpreted with caution. Moreover, the PBAC has conducted a test for interaction of treatment effect modification across the two baseline fracture subgroups: - the nominated ("severe") subgroup and its comparator ("non-severe", i.e. nil + mild + moderate fracture subgroups at baseline) and shown a non-significant result (p=0.1433) for relative risk reduction. This information is contained within the evaluator's commentary on the July 2005 re-submission (refer to section 7.6-page 13). These results are also shown in this review in Tables 3 and 4 (pages 16-17). Finally, the interpretation of these results should be viewed cautiously. According to Altman and Bland, results of tests for interaction are likely to be convincing only if they were specified at the start of the study. As already discussed in matter 1B, such analysis occurred post-hoc.

Reviewer's summary There is sufficient evidence to support the sponsor claim of biological plausibility that teriparatide may have an improved treatment effect over comparator therapy (antiresorptive drugs) in a population of patients with severe vertebral osteoporosis. This is based on the drug's anabolic mechanism of action and how such an action may be disproportionately effective in severe osteoporosis where there is significant alteration in the pathophysiology of osteoporosis favouring such a therapeutic action.

However, there is insufficient clinical outcome data in the submissions to conclusively support the sponsor claim that teriparatide is more effective in the high risk vertebral fracture subgroup than in the overall treatment population. In addition, It remains unclear whether or not teriparatide has a superior treatment effect compared to alternative anti-osteoporosis agents for patients with more severe forms of osteoporosis as direct comparative (head-to-head) clinical outcome data in this sub-population does not currently exist.

MATTER 3 – VALIDITY OF CONTINUED USE OF UTILITIES

Background to matter arising

The sponsor raised the following in its requesting letter for review.

The sponsor says 'PBAC concern with regard to continuing to use the same utility values in spite of the sponsor's efforts to address these concerns in its responses – paragraph 4 of the PBAC minutes'

The nature of PBAC concern is stated in the minutes of the March 2006 PBAC meeting (item 7.8.25): 'The PBAC remained concerned about the continuing use of the same utilities and disutilities in the model where the sensitivity analyses indicate the model is sensitive to the assumptions used to derive the incremental utility estimates from the trial-based outcome measures.'

The nature of the concerns by the sponsor outlined in the November 2005 sponsor submission relate to the response by PBAC to earlier submissions that stated "the utilities in the base case model might be skewed downward since the experts may not have accounted for some fractures being asymptomatic. The Model should separate out the utilities for a painful and non-painful vertebral fracture rather than relegating this to a sensitivity analysis."

The sponsor argues that the "comparison of their AQoL derived utility weight (0.217) being substantially lower than the utility elicited for a 'good outcome' vertebral fracture using time trade-off (0.257) is not valid. The sponsor proceeds to state that the comparison is not relevant for several reasons:

a) Adjusting the submission figure to account for 30% of fractures being symptomatic the utility is actually 0.266 and close to the Salkeld 0.258 utility.
b) It is inappropriate to compare utilities elicited using different methods such as HUI vs EQ-5D (PBAC overview for the July 2005 meeting, PBAC OVR 7.6.1]
c) Patients in the model have severe previous fractures. Therefore, their utility with a further fracture is expected to be below the Salkeld et al 0.259 utility for an initial or subsequent fracture

On the question of separating out the utilities for a painful and non-painful vertebral fracture rather than relegating this to a sensitivity analysis, the sponsor states that this was excluded from the utility survey "as it was felt that the experts could better evaluate the proportion of fractures that would cause painful symptoms rather than providing a figure (i.e. 70%). However, to account for this unlikely oversight by the experts, a sensitivity analysis assumes that 70% of fractures would be asymptomatic".

"The evaluators acknowledge the possibility that the 70% figure for the asymptomatic portion is, in fact, less in this sub set 97.6 EVAL.22]. In this case the acute and chronic fracture utilities in the model would actually be skewed upwards leading to a conservative analysis."

Reviewer's understanding of key issues
Based on the matters raised by the sponsor regarding utilities, the specific matters reviewed are:

- whether the continued use of the same utilities and disutilities in the model in the cost utility analysis in the submission is sufficiently justified.
- whether the AQoL derived utility weights incorporated expert judgement on the proportion of fractures that would cause painful symptoms. In the event that the expert's valuation does not include a judgement about the proportion of symptomatic fractures, are the sensitivity analyses adjustment for the disutility associated with asymptomatic fractures plausible?
- whether the AQoL derived utility weights for vertebral fractures are comparable with utilities derived for other health states resultant to using TTO or other multi-attribute instruments.

Materials considered
All materials considered by the PBAC relating to the derivation of utility weights were reviewed. In particular, key reference documents (page 50) used in making judgements about concepts and methods relating to the derivation of the utility weights in the submissions included the PBAC Guidelines for cost effectiveness analysis submissions and the textbook authored by Drummond on Methods for the Economic Evaluation of Health Care Programmes.

Reviewer's opinion

1. Is the continued use of the utility weights used in the submission sufficiently justified?

(a) Choice of QoL measurement instrument
In the modeled economic evaluation of the July 2005 and March 2006 submissions, quality of life was accepted as an appropriate final outcome of therapy and hence the need for a utility-based measure of quality of life to generate a QALY measure.

For economic evaluation, the sponsor uses a QALY framework, which requires the derivation of QALY weights (referred to as 'utility' weights by the sponsor).

According to the PBAC Guidelines "Where a quality of life instrument is used, details should be provided on the instrument. Because currently there is controversy over which quality of life instruments are most acceptable, special attention should be paid to the following parameters:
a. the validity of the instrument;
b. the reliability of the instrument;
c. the responsiveness of the instrument to differences in health states between individuals and to changes in health states over time experienced by any one individual; and
d. the clinical importance of any differences detected by the instrument.

The generally accepted requirements to generate a QALY measure are summarized by Drummond et al which state that to "satisfy the QALY concept……the quality weights must be (a) based on preferences, (b) anchored on perfect health and death, and (c) measured on an interval scale.

The sponsor uses the Assessment of Quality of Life (AQoL) Instrument to derive utility weights for five health states related to the comparator and outcomes of drug therapy.

Conclusion
The AQoL instrument would be regarded as meeting guideline requirements for a valid and reliable instrument that is sensitive to clinically important changes in health status. The choice of the AQoL instrument is therefore justified in the context of this submission.

(b) Protocol for the derivation of utilities

Background
The current PBAC guidelines do not require the sponsor to outline the protocol for generating utilities. However, a protocol is the best means by which a reviewer can justify the final utility weights used in their QALY model. An example of a framework for generating a protocol for utility measurement and valuation by Furlong et al is contained in appendix 1 (page 51). The absence of a protocol for the derivation of utilities in the submission means that elements of the utility survey are not adequately justified in the text of the submission.

Instead, the sponsor describes the process by which a group of clinical experts were asked to complete a postal survey using the AQoL instrument. The utility weights used in the QALY calculation were based on the responses of 8 experts, although there are 9 declaration forms from experts contained in Appendix J of the July 2005 submission.

To assess whether the sponsor has adequately justified the utilities used in the QALY model, each element of utility measurement and valuation is addressed in the review. The elements of utility measurement and valuation include:

- Choosing health states and descriptors
- Selecting a measurement instrument
- Whose preferences (respondents and sample size)

(i) Choice of health states and descriptors

The March 2005 and November 2005 sponsor submissions state that utility weights were derived for five health states, they are:

- the comparator (base case) – "An individual with experience of 3 vertebral fractures one of which was a severe grade vertebral fracture (SQ3 grade)", and
- four additional health states are described, based on the QoL for an individual at 2 weeks, 6 months, 12 months and up to 10 years post an additional moderate vertebral fracture.

The comparator health state descriptor is only partially consistent with the PBS listing-authority required in which the sponsor states that the recipients of treatment must have evidence of one severe **painful** osteoporotic vertebral fracture. The word painful does not appear in any of the health state descriptions. It is not clear whether the 'expert' respondents would have imputed from the health state that a severe vertebral fracture was painful or not. This could be clarified if the sponsor provided the mean score on the AQoL for item 9 'pain or discomfort experienced'.

The health state descriptor and duration of the health states are appropriate, with one exception. The use of the word 'moderate' in the health state descriptions is crucial to the valuation of the potential outcomes. The economic model makes no distinction as to the type of vertebral fracture avoided beyond accounting for clinically evident and asymptomatic vertebral fractures in the sensitivity analysis. The use of the descriptor 'moderate' in the health states may represent a simplification of the range of possible vertebral fracture outcomes (from asymptomatic through to a severe vertebral fracture) but may also induce some bias in the QoL measurement by failing to make any distinction between symptomatic and asymptomatic vertebral fractures.

Conclusion The health state descriptors do not distinguish between asymptomatic and clinically evident vertebral fractures. This is discussed further in the next section.

(ii) Measurement instrument

The AQoL is a multiattribute utility instrument that has used both the time trade-off (TTO) and person trade-off (PTO) methods to develop a utility-based scoring algorithm. The sponsor omitted three QoL questions from the survey, those concerned with vision, hearing and speech, because these would be unaffected by vertebral fracture. This was stated on the survey form received by the respondents. It is inappropriate to omit questions from a standardized survey instrument even if their exclusion is highly unlikely to affect the final valuation.

Conclusion The AQoL is an appropriate utility-based QoL measurement instrument

(iii) Whose preferences?

The AQoL scoring algorithm is based on the responses from a community sample so the values used in the sponsor's QALY estimates are appropriate. The key question is whether the measurement process, where 9 clinicians were asked to map the health states into the AQoL was adequately justified.

The PBAC Guidelines provide no information on whose preferences should be used to map health states in a multiattribute utility instrument. In addition, there is not any guidance on the sample size required for this exercise. The sponsor does not provide an adequate justification for the selection of the 9 experts nor for the sample size.

Reviewer's summary
The use of the AQoL multiattribute utility instrument is appropriate for the purpose of deriving utility weights for the vertebral fracture health states. The instrument is valid and reliable and uses community values to ascertain a utility score. The role of the nine clinical experts in mapping the 5 health states onto the AQoL index was appropriate but there was inadequate justification of the selection of the experts and the sample size. The health state descriptors do not distinguish between asymptomatic and clinically evident vertebral fractures.

Conclusion Both the PBAC guidelines for deriving health state utility weights and the sponsor's justification of their approach to deriving the utility weights are inadequate.

2. Whether the AQoL derived utility weights incorporated the expert's judgement on the proportion of fractures that would cause painful symptoms and whether the sensitivity analyses adjustment for the disutility associated with asymptomatic fractures are plausible?

Background

The sponsor comments in the re-submission for consideration at the March 2006 meeting of the PBAC say: ' The re-submission included a modelled economic evaluation using alendronate as the comparator. This model was constructed to be deliberately conservative in the use of the clinical and epidemiological evidence over the ten-year period of the model. Including less conservative assumptions in the model would reduce the ICER substantially. However all sensitivity analyses conducted were done so using this base case model. The evaluator has presented the table of sensitivity analyses presented in the re-submission (Table 83 of the submission) but has not commented on the two-pages of discussion of the plausibility of these analyses. We have re-presented this Table, with this response, but have re-ordered the analyses based on the ICER. (see Table -"reproduced without alteration in this review as Table 7 on page 46")

Several scenarios were included in the sensitivity analysis of this conservative model to determine the effect on the modelled evaluation. Essentially these analyses suggest that the cost-effectiveness ratios are sensitive to changes in assumptions, however, the incremental cost-effectiveness ratio for teriparatide remains below \$45,000-\$75,000[1] per QALY gained for all sensitivity analyses, with the exceptions being the most extreme and least plausible analyses. Even under these scenarios the ratio remains in the range \$45,000-\$105,000[2]. The cost-effectiveness of teriparatide is insensitive to the assumptions around residential care, the mortality risk increase following vertebral fracture, and the discount rate.

A two-way sensitivity analysis was carried out, following that performed by the evaluators in the PES commentary on the March 2005 submission. In that commentary, an analysis was presented where the disutilities from new fracture were low and used in conjunction with the upper 95% confidence limit for fracture relative risk (0.51) for the entire GHAC population.

When these scenarios are used together with the mean relative risk of fracture for alendronate (0.533), the ICER increases to >\$200,000[3]. However, to accept this sensitivity analysis as plausible would assume that a new fracture in an SQ3 patient would not be associated with any disutility. As we have presented in the submission, these patients are already suffering from multiple fractures, experience clinical levels of pain and have a low HR QoL. New fractures in this patient will be associated with further pain and disability and negative impacts on the activities of daily living, thus it is reasonable to expect that new fractures will

[1] ICER replaced with ICER range, consistent with PBAC procedures for PSD
[2] As above
[3] As above

further compound the patient's condition and will be associated with further disutility.

(a) Accounting for asymptomatic vertebral fractures in the QALY model

Key issues
The main issue here is how the QALY model accounts for an estimated 70% of vertebral fractures being asymptomatic and 30% being symptomatic. There is an additional issue about the exact proportion of vertebral fractures that are asymptomatic in the sub set of patients modeled in the economic evaluation.

Background to matter arising

In its November 2005 submission, the sponsor states that "the major concern that the PES and ESC had with the utilities was that the descriptions of the health states in the utility study might not capture the fact that around 70% of vertebral fractures are asymptomatic. The sponsor goes on to say that the PES itself pointed out that "*one would expect experts to be aware that most morphometric fractures are not clinically evident {July 2005 Commentary on the Resubmission 7.6.22}*

In the PBAC July 2005 minutes, the quote in italics above was accompanied by a clear concern by PBAC about the impact of not explicitly stating this fact in the utility survey. The July 2005 PBAC minutes, state that "although one would expect experts to be aware that most morphometric vertebral fractures are not clinically evident, it may be that the survey by its structure and/or implementation skewed the utilities downward by not mentioning this fact prominently in the instructions….The fact that the scenario to be assessed involves a patient with a severe vertebral fracture and that it is one of three vertebral fractures in total is prominently emphasized, and reinforced by noting that the next vertebral fracture encountered is "moderate." On the other hand, it is possible that the 70% figure for the asymptomatic portion is, in fact, less in this sub-set. No good data is available regarding any aspect of this topic, which makes the derivation of utilities highly uncertain and a particularly important element to capture in the assessment of uncertainty. In all cases it seems evident that the primary model should use utilities that account for a significant portion of vertebral cases being asymptomatic, rather than relegating this scenario to a sensitivity analysis."

Reviewer's opinion-survey design and QALY model
The problem here is in the QoL survey design and the QALY model. The probability of entering a health state should be considered separate to the task of valuing that health state. Hence there should have been separate health states for a symptomatic and asymptomatic fracture used in the AQoL health utility survey. The proportion of asymptomatic vertebral fractures (be it 70% or otherwise) would be a transition probability in the economic model (as would the probability of having a symptomatic fracture). The probability would be multiplied by the relevant utility weight for a symptomatic or asymptomatic vertebral fracture.

Conclusion
Based on the data presented in the submission, there is no way of assessing whether the expert respondents did or did not consider the proportion of asymptomatic fractures in mapping the health states into the AQoL.

There is a flaw in the design of the utility survey and cost utility model. The probability of entering a health state should have been separated from the task of valuing the health state.

(b) Reviewer's opinion-sensitivity analysis

Where there is uncertainty in one of more parameters in the economic model, sensitivity analysis should be used to assess how varying the parameter(s) impacts on the study results. Sensitivity analysis is also used to quantify the level of uncertainty relating to the methodological assumptions of the study. In the two submissions under review here, there is uncertainty relating to the mean utility weight and to the method used to derive the utility weight.

The sponsor has allowed for uncertainty in the method (for deriving the utility weights) and in the utility score itself by modeling two scenarios: a) one where 70% of asymptomatic fractures are assumed to have one third of the disutility associated with clinical fractures and another b) where 70% of asymptomatic fractures are assumed to have no disutility relative to the baseline health state.

Conclusion This is an appropriate way to deal with the uncertainty surrounding the derivation of the final utility weights. It is plausible to assume that there is disutility associated with a symptomatic fracture in a SQ3 patient.

(c) Reviewer's opinion-Inclusion of asymptomatic vertebral cases
On the question of whether the base case in the QALY model should include a significant proportion of vertebral cases being asymptomatic, there is a difference of opinion between the sponsor and PBAC on whether the experts' valuations, using AQoL-based utility survey, do allow for a proportion of vertebral fractures being asymptomatic.

There is no way of to make an objective assessment on this without going back and asking the experts whether they did or did not consider the proportion of asymptomatic versus symptomatic fractures. This points to a flaw in the survey design. It was not clear whether the vertebral fracture health states represented a clinically evident or asymptomatic fracture. It is inappropriate to expect the valuer (expert) to impute proportions when mapping the health state into the AQoL matrix. It is the health state that is being valued, not the probability of entering that state. The probability, is a measure of chance of entering the health state and rightly belongs as a separate parameter in the economic model.

Conclusion The base case ICER (cost per QALY) should be based on the utility valuation of separate health states (symptomatic versus non-symptomatic) and an appropriate transition probability of a person in the model entering that state. It is impossible to assess whether the sponsor's base case cost per QALY does allow for the proportion of symptomatic versus asymptomatic vertebral fractures because of a flaw in the design of the utility survey.

The two scenarios in the sensitivity analysis that allow for 70% of asymptomatic fractures are assumed to have one third of the disutility associated with clinical fractures and another where 70% of asymptomatic fractures are assumed to have no disutility relative to the baseline health state. The two scenarios (a and b described above) are both clinically plausible and better reflect the likely impact of uncertainty surrounding both the method used to derive the utility weights and the final utility weights themselves.

Reviewer's summary

The correct approach to modeling the utility/disutility associated with a symptomatic and asymptomatic vertebral fracture would have been to value the health states separately in the AQoL expert survey. This was not done in the AQoL survey contained in the sponsor's submission and there is no way of assessing whether the expert respondents did allow for the proportion of asymptomatic fractures in their responses. Given this limitation, it is reasonable and plausible to include the two scenarios in the sensitivity analysis allow (a) for 70% of asymptomatic fractures assumed to have one third of the disutility associated with clinical fractures and (b) another where 70% of asymptomatic fractures are assumed to have no disutility relative to the baseline health state.

3. The comparability of the AQoL derived utility weights with other health states, such as the utility associated with a hip facture

Background
The comparability of the acute vertebral fracture utility elicited using the AQoL (0.217) to other health states, such as hip fracture, may provide evidence on the validity of the utility weights used in the two submissions.

The sponsor states that a comparison of the acute vertebral fracture utility elicited using the AQoL (0.217) to the 'good outcome' hip fracture utility weight using TTO (Salkeld et al) is not relevant. Three reasons are given by the sponsor:

1) the adjusted (for 30% of fractures being asymptomatic) utility is 0.266 not 0.217 and 0.266 is close to the hip fracture utility of 0.31;
2) that the PES state that it is inappropriate to compare utilities elicited using different methods;
3) patients in the model have severe SQ3 fractures and their utility with a further fracture would be expected to be lower than the hip fracture utility weight.

Reviewer's opinion
It is inappropriate to compare mean utility weights using different utility measurement instruments without some attempt to transform them into a common index to ensure comparability. Whilst most multiattribute utility instruments, such as the AQoL, HUI3, EQ-5D, and direct health state description valuation methods such the SG, TTO PTO all satisfy the conditions for the derivation of a QALY weight, the underlying theory and methodological assumptions behind each in terms of deriving 'utility' weights are slightly different. The usefulness of utility theory is to ensure that valuations across people, time and place are comparable. As the underlying theory for the different instruments listed above are different, a direct comparison is not possible. Therefore, the comparison to the Salkeld et al hip fracture utility weight of 0.31 is not relevant to the submission(s).

To highlight the impact of different approaches to utility valuation, the Harvard University-based Cost Effectiveness Registry has hundreds of published preference weights (they do not use the word 'utility'), classified by disease group and health state. A summary of preference weights for various vertebral fracture health states are presented in Table 6 (pages 44).

Table 6: Other published preference weights for vertebral fracture for the period 1998 - 2001

Health state	Preference weight	Method	Valuer	Author
1st year following vertebral fracture in post menopausal women	0.64	Standard Gamble	42 women with osteoporosis	Coyle D et al 7 2001
Year with vertebral fracture	0.9	Clinician judgment	Clinician	Willis M et al 8 2001
Vertebral fracture 1st year	0.704	Unknown	Clinician	Armstrong K et al 9 2001
Vertebral fracture subsequent years	0.858	Unknown	Clinician	Armstrong K et al 9 2001

Source: http://www.tufts-nemc.org/cearegistry/data/default.asp

In the absence of a standardized approach to QALY weight measurement, it is known that different utility measurement techniques will produce different weights for identical health states. Where comparability of QoL weights is important to resource allocation decisions, as it can be for PBAC, there is considerable merit in recommending a standardized approach to utility-based weights for QALY ratios.

Reviewers Summary for Matter 3

1) The continued use of the utility weights used in the submission is not sufficiently justified. The base case ICER (cost per QALY) should be based on the utility valuation of separate health states (symptomatic versus non-symptomatic) and an appropriate transition probability of entering that state.

2) It is impossible to assess whether the sponsor's base case cost per QALY does allow for the proportion of symptomatic versus asymptomatic vertebral fractures because of a flaw in the design of the utility survey. The one-way sensitivity analysis adjustment for the disutility associated with asymptomatic fractures ($45,000-$75,000[4] for the one-third disutility and $75,000-$105,000 [5]for zero disutility) are plausible.

3) The AQoL derived utility weights for vertebral fracture health states are not directly comparable to the utility weights associated with a hip facture and which are derived using a different utility measurement technique.

Recommendations
That the PBAC make explicit in the guidelines, the steps required to identify, measure and value QoL outcomes for inclusion in a QALY-based cost utility analysis. To a large extent this has been achieved in the PBAC Guidelines Draft for Consultation July 2006. The draft Guidelines 2006 requires a more detailed justification for the choice of utility measurement approach – a MAUI or Scenario – Based approach and the criteria for justifying the selection of a particular instrument.

[4] ICER replaced with ICER range, consistent with PBAC procedures for PSD
[5] As above

Table 7 Sensitivity analyses in the modelled evaluation

Change from baseline	Incremental cost	Incremental QALYs	Incremental cost per QALY gained
Vertebral fracture relative risk d. Teriparatide upper confidence interval, alendronate lower confidence interval	<$15,000 (ICER replaced with range consistent with PBAC procedures for PSD)	-0.075	Dominated
11. Comparator is placebo		0.533	<$15,000 (ICER replaced with range consistent with PBAC procedures for PSD)
1. Vertebral fracture relative risk c. Teriparatide lower confidence interval, alendronate upper confidence interval		0.485	
6.Treatment benefit duration b. Remains constant to 10 years		0.314	$15,000-$45,000 (ICER replaced with range consistent with PBAC procedures for PSD)
7. Vertebral fracture utilities b. Decreased by 10%		0.306	
8. Costs b. Vertebral fracture acute cost increased by 50%		0.264	
12. Non-vertebral fractures included		0.276	
1. Vertebral fracture relative risk a. Teriparatide lower confidence interval, alendronate lower confidence interval		0.295	
1. Vertebral fracture relative risk f. Antiresorptive relative risk from MORE SQ3 subgroup analysis (0.74)		0.277	
9. Discounting a. Costs and outcomes undiscounted		0.287	
8. Costs d. Residential care cost increased by 50%		0.264	
Baseline		0.264	
7. Vertebral fracture utilities e. Residential care disutility excluded		0.263	
8. Costs e. Residential care cost decreased by 50%		0.264	
3. Relative risk of mortality a. Relative risk of mortality after vertebral fracture = 1.66		0.260	
10. Two vertebral fractures at baseline		0.259	
4. No increased risk of residential care after vertebral fracture		0.263	

Change from baseline	Incremental cost	Incremental QALYs	Incremental cost per QALY gained
3. Relative risk of mortality b. No increased mortality after vertebral fracture		0.247	
9. Discounting b. Costs and outcomes discounted at 10% per annum		0.244	
5. Starting age in model b. Starting age is 75 years		0.240	
8. Costs c. Vertebral fracture acute cost decreased by 50%		0.264	
7. Vertebral fracture utilities a. Increased by 10%		0.222	
5. Starting age in model a. Starting age is 65 years		0.215	
8. Costs a. All non-drug costs excluded		0.264	
6. Treatment benefit duration a. Expires at 5 years		0.201	
7. Vertebral fracture utilities c. Adjusted so 70% have 1/3 disutility		0.154	$45,000-$75,000 (ICER replaced with range consistent with PBAC procedures for PSD)
2. Lower increase in vertebral fracture risk after fracture (2.2)		0.160	
1. Vertebral fracture relative risk e. Relative risk for teriparatide uses entire OP patient sample (0.35)		0.140	
1. Vertebral fracture relative risk b. Teriparatide upper confidence interval, alendronate upper confidence interval		0.116	
7. Vertebral fracture utilities d. Adjusted so 70% have no disutility (zero disutillity if asymptomatic)		0.097	$75,000-$105,000 (ICER replaced with range consistent with PBAC procedures for PSD)
Two-way sensitivity analysis Teriparatide upper confidence interval for vertebral fracture relative risk, and vertebral fracture utilities adjusted so 70% of fractures have no disutility		0.006	>$200,000 (ICER replaced with range consistent with PBAC procedures for PSD)

Source: Response to the Pre-PBAC consultation – Re-submission for consideration at the March 2006 Meeting of PBAC.

KEY REFERENCES

Matter 1A – Indirect data comparison

Black DM, Cummings SR, Karpf DB, et al. Randomised trial of effect of aledronate on risk of fracture in women with existing vertebral fractures. Fracture Intervention Trial Research Group. Lancet 1996; 348: 1535-41.

Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997; 50: 683-91.

Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005; 331: 897-900.

Delmas PD, Genant HK, Crans GG, Stock JL, Wong M, Siris E. Severity of prevalent vertebral fractures and the risk of subsequent vertebral and nonvertebral fractures: results from the MORE trial. Bone 2003; 33: 522-32.

Ensrud KE, Black DM, Palermo LMS, et al. Treatment with aledronate prevents fractures in women at highest risk: results from the Fracture Intervention Trial. Arch Int Med 1997; 157: 2617-24.

Lunt M, O'Neill TW, Felsenberg D et al. Characteristics of a prevalent vertebral deformity predict subsequent vertebral fracture: results from the European Prospective Osteoporosis Study (EPOS). Bone 2003; 33: 505-13.

Neer RM, Arnaud CD, Zanchetta JR et al. Effect of parathyroid hormone (1-34) on fractures and bone mineral density in postmenopausal women with osteoporosis. N Engl J Med 2001; 344: 1434-41.

Nevitt MC, Chen P, Dore RK et al. Reduced risk of back pain following teriparatide treatment: a meta-analysis. Osteoporosis International; published online: 2 September 2005 (accessed July 10 2006).

Song F, Altman DG, Glenny A-M, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. BMJ 2003; 326: 472-6.

Matter 1B – Subgroup analysis

Assman SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet 2000; 355: 1064-69.

Gallagher JC, Genant HK, Crans GG, Vargus SJ, Krege JH. Teriparatide Reduces the Fracture Risk Associated with Increasing Number and Severity of Osteoporotic Fractures. J Clin Endocrinol Metab 2005; 90(3): 1583-7.

Lagakos SW. The challenge of Subgroup Analyses-Reporting without Distorting. N Engl J Med 2006; 354: 1667-9.

Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. Ann Intern Med 1992; 116: 78-84.

Parker AB, Naylor CD. Subgroups, treatment effects, and baseline risks: some lessons from major cardiovascular trials. Am Heart J 2000; 139: 952-61.

Pfeffer MA, Jarcho JA. The Charisma of Subgroups and the Subgroups of CHARISMA. N Engl J Med 2006; 354: 1744-6.

Rothwell PM. Treating Individuals 2. Subgroup analysis in randomized controlled trials: importance, indications, and interpretation. Lancet 2005; 365: 176-86.

Scott IA, Greenberg PB. Cautionary tales in the clinical interpretation of therapeutic trial reports. Int Med J 2005; 35: 611-21.

Yusif S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991; 266: 93-8.

Matter 2 – Biological plausibility

Altman DG, Bland JM. Statistics Notes Interactions revisited: the difference between two estimates. BMJ 2003; 326: 219.

Horwitz M, Stewart A, Greenspan SL. Sequential parathyroid hormone/aledronate therapy for osteoporosis-robbing Peter to pay Paul? J Clin Endocrinol Metab 2000; 85: 2127-8.

Jiang Y, Zhao JJ, Mitlak BH, Wang O, Genant HK, Eriksen EF. Recombinant Human Parathyroid Hormone (1-34) [Teriparatide] Improves both Cortical and Cancellous Bone Structure. J Bone Miner Res 2003; 18: 1932-41.

Klotzbeucher CM, Ross PD, Landsman PB, Abbott 3rd TA, Berger M. Patients with prior fractures have an increased risk of future fractures: a summary of the literature and statistical synthesis. J Bone Min Res 2000; 15: 721-39.

Parkinson IH, Fazzalari NL. Interrelationships Between Structural Parameters of Cancellous Bone Reveal Accelerated Structural Change at Low Bone Volume. J Bone Min Res 2003; 18: 2200-05.

Zanchetta JR, Bogado CE, Ferretti JL, Wang O, Wilson MG, Sato M, Gaich GA, Dalsky GP, Myers SL. Effects of Teriparatide [Recombinant Human Parathyroid Hormone (1-34)] on Cortical Bone in Postmenopausal Women with Osteoporosis. J Bone Min Res 2003; 18: 539-43.

Matter 3 – Validity of continued use of Utilities

Armstrong K, Chen TM, Albert D, Randall TC, Schwartz JS. Cost-effectiveness of ralozafine and hormone replacement therapy in postmenopausal women: impact on breast cancer risk. Obstetrics and Gynaecology. 2001;98,6:996-1003.

Commonwealth Department of Health and Ageing. Draft for Consultation July 2006. Guidelines for the Pharmaceutical Industry on preparation of submissions to the pharmaceutical Benefits Advisory Committee. Including major submissions involving economic analyses.
URL:http://www.health.gov.au/internet/wcms/publishing.nsf/Content/pbac_draftguidelines (accessed 11th August 2006)

Commonwealth Department of Health and Ageing. Guidelines for the Pharmaceutical Industry on preparation of submissions to the pharmaceutical Benefits Advisory Committee. Including major submissions involving economic analyses. September 2002. (Section 3.5 of submission)

Commonwealth Department of Health and Ageing. Guidelines for the Pharmaceutical Industry on preparation of submissions to the pharmaceutical Benefits Advisory Committee. Including major submissions involving economic analyses. September 2002. (Appendix H of submission)

Coyle D, Cranney A, Lee KM, Welch V, Tugwell P. Cost-effectiveness of nasal calcitonin in postmenopausal women: use of Cochrane Collaboration methods for meta-analysis with economic evaluation. Pharmacoeconomics. 2001;195:565-575

Drummond MF, O'Brien B, Stoddart GL and Torrance GW. 1997 (2nd Ed) Methods for the Economic Evaluation of Health Care Programs (2nd Ed). Oxford Medical Publications, Oxford. Pg 170

Furlong W, Feeney D, Torrance GW, Barr R, and Horsman J. (1990). Guide to the design and development of health-state utility instrumentation. McMaster University: CHEPA Working Paper $90-9.

Hawthorne G, Richardson J, Day N. A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. Annals of Medicine 2001,33;5: 358-370

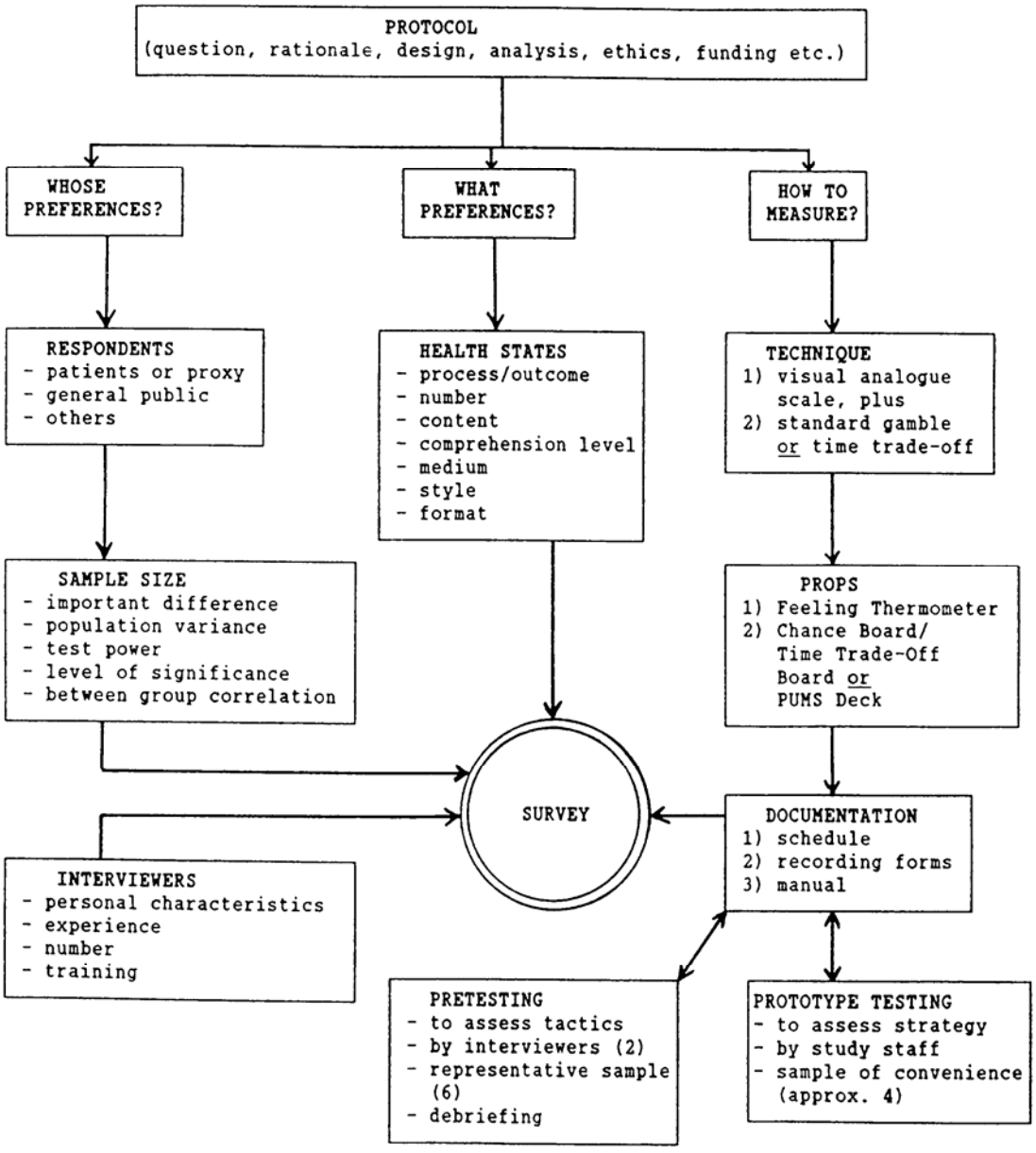
Salkeld G. Cameron ID, Cumming RG, Easter S, Seymour J, Kurrle SE, Quine S. Fear of falling and hip fracture related quality of life - measuring the preferences of older women: a time trade-off study. BMJ 2000;320:341-346.

Willis, Odegaard K, Persson U, Hedbrnat J, Mellstorm D, Hammar A. A cost-effectiveness model of tibolone as treatment for the prevention of osteoporotic fractures in postmenopausal women in Sweden. Clinicial drug Investigation. 2001;21,2:115-127.

Appendix 1

Figure 1.0

FLOW DIAGRAM OF BASIC STEPS IN DEVELOPMENT OF HEALTH-STATE UTILITY MEASUREMENTS



Source: Furlong et al (1990)