Response from the Assessment Group to the documents received commenting on the Assessment Group report entitled "Percutaneous vertebroplasty and percutaneous balloon kyphoplasty for the treatment of osteoporotic vertebral fractures"

As of the 28th of September 2012, responses were received from the following organisations:

- Johnson and Johnson (14 Pages)
- Medtronic (10 Pages)
- Synthes (4 pages)
- British Society of Skeletal Radiologists (3 Pages)
- British Pain Society (2 Pages)
- Healthcare Improvement Scotland (1 Page)

These have been taken in turn and responded to by the Assessment Group within this document. Where points are contained in both the executive summary and the main text we have addressed the points in the order of the main text as it is likely that there is fuller detail within the main document. For completeness the majority of comments have been reported within the Table including those that concur with comments within the Assessment Group report or that are statements of fact which have not been disputed by the Assessment Group. In both cases no response has been provided by the Assessment Group. Where the same point is raised multiple times in the main text only the first instance is addressed. Where comments have been omitted, this is indicated, along with the reasons for omissions.

In some instances the comment has been amended for readability reasons. In these cases the Assessment Group have endeavoured to retain the full sentiment of the comment.

Johnson and Johnson	
Comment	Response
The invasive control procedures performed in the Kallmes and Buchbinder studies clearly cannot be considered as a SHAM or placebo, as both involve an invasive procedure. Therefore, it is not possible to conclude from these studies that the treatment effect associated with vertebroplasty (PVP) is due to a 'placebo effect'.	We do not think that infusion of lidocaine 1% into the skin surrounding the affected vertebral body, as per the Buchbinder trial, can be classed as 'invasive'. The important question, as we emphasise at several points in the report, is whether local anaesthetic actually intervenes in the biomechanical processes of VCFs to produce improvements in pain and HRQoL for people with VCFs. Debate is ongoing as to the extent to which this may be the case. For these reasons a large number of scenarios have been presented in the Assessment Report (AR) that examines how the cost-effectiveness of the interventions change given different assumptions regarding the effectiveness of each intervention.
These two publications [Kallmes and Buchbinder] are significantly flawed and highlight the challenges of conducting adequately powered RCTs of vertebroplasty, including barriers to recruitment and the need for appropriate patient selection.	Since their publication, the Buchbinder and Kallmes studies have been the subject of many methodological critiques. The AR covers these comprehensively on pages 144-149. Although we acknowledge that the Buchbinder and Kallmes studies leave a number of questions to be addressed, and agree that the studies highlight the challenges in conducting RCTs of this procedure, the suggestion that these studies were more flawed than the other available evidence does not stand up to critical scrutiny. The Assessment Group (AG) conducted a quality assessment of the relevant RCTs, involving independent assessment by 2 reviewers, and both these studies scored highly. In addition, several observational studies and an individual patient meta-analysis conducted by Staples et al., contradict some of the key points made in criticisms of Buchbinder and Kallmes, including those of patient selection based on fracture acuity and pain severity.
For people with painful osteoporotic VCFs refractory to analgesic treatment, vertebroplasty (PVP) performs significantly better in unblinded trials than NIM/OPM (non-invasive management) in terms of improving quality of life, reducing pain and disability.	We agree. However, it should be added that there is no evidence that PVP performs any better than blinded injection of local anaesthetic, a procedure with fewer risks.
For patients that are refractory to conservative treatment, PVP and BKP are the only routinely performed treatments available.	Generally we are in agreement. However, it is noted that in some centres, facet joint injections are offered prior to PVP or BKP.
There is some evidence that PVP and BKP are associated with reductions in mortality; however, this effect has not been fully investigated in clinical trials with a randomisation procedure, so the causal mechanisms remain unclear.	No specific response required as we believe this comment does not contradict the AR. However, it is commented that the trial size and follow-up required to investigate mortality within a randomised setting is likely

	to make this infeasible.
The results of the Assessment Group's foundation analysis, PVP had an ICER of £7,802 which is cost-effective and well below the standard threshold of £20,000 per QALY.	In isolation this is a factual statement. However, the AG note that many analyses were presented, none of which were designated the base case due to the considerable uncertainty in the decision problem
In the base case submitted by Johnson and Johnson Medical, the ICER for vertebroplasty vs. NIM was £4,392 in the base case, and ranged from £568–£13,595 in the scenarios considered	No response required as this is a factual statement.
PVP with either low or high viscosity cements is cost-effective relative to non-invasive management for patients with osteoporotic vertebral compression fractures. The incremental cost-effectiveness of various cements for either PVP or BKP was not addressed in Johnson and Johnson Medical's submission as such analyses were outside of the scope for this MTA.	No response required on the analyses undertaking within Johnson and Johnson Medical's submission as this is a factual statement. Regarding the scope it is unclear whether different cement types were included, although it is noted that Johnson and Johnson Medical analysed the difference in the safety of different types of cements within their submission, which would appear redundant if cement types were not to be considered. The manufacturer states in their response that PVP with either low or high viscosity cements is cost-effective relative to non-invasive management for patients with osteoporotic vertebral compression fractures, however, no analyses for low-viscosity cements were presented; only PVP using Johnson and Johnson's high-viscosity cement reported. The AG explicitly undertook analyses considering the impact of using high-viscosity cements in all patients.
The AR attempted subgroup analyses on different types of cement viscosity. Given this was not part of the intentions of the assessment as detailed in the scope, and the Assessment Group excluded the only RCT level evidence on the differential safety profile of High Viscosity cement1 (18), we feel this subgroup analysis is not robust. Had the scope detailed that cement type was a consideration of the appraisal we would have investigated these differences more completely. *Anselmetti, (18) a study on the use of HV cement (Confidence Spinal Cement System®) in VBA, reported no significant difference in the rate of intradiscal leakage for HV and LV cement, 6.1% vs. 13.0%, respectively However, the rate of venous leakage in the study, which is more likely to lead to clinically significant complications, was significantly lower with HV cement, at 8.2%, compared to 41.3% with LV cement (p<0.0001)*	It is arguable whether cement type is part of the scope and thus the AG erred on the side of providing the Appraisal Committee with as much information as possible. These analyses were marked as exploratory and a threshold approach was taken to allow the Appraisal Committee to discuss, if desired, whether, if the analyses were deemed robust, such values were likely. Additionally an estimate of the likely QALYs gained were all adverse events avoided has been undertaken which is a more favourable position than the results for PVP reported by Anselmetti.
It is difficult to assess the different cement types in cost-effectiveness analyses, given the current limited availability of comparative level 1 evidence related to HV vs. LV cements. This is to be expected as HV cement has been developed more recently so the evidence is less developed for this technology.	Whilst data for adverse events related to high viscosity cements may be immature, the data for adverse events associated with lower viscosity cements are more robust. These have been used to provide an estimate of the maximum QALY gain that could be achieved were all adverse

	events removed. Additionally, as stated, the analyses were marked as exploratory
The clinical advisor to the Assessment Group confirmed there is a role for cements of differing viscosity in treating this patient population but this decision should be based on clinical judgement and based on individual patient characteristics until evidence to investigate this specific point has been generated.	The AG does not dispute this statement. As stated the analyses conducted were to indicate the likely cost-effectiveness of expanding the use of high-viscosity cements to all patients rather than the percentage (estimated at 15%) that are judged to require high-viscosity cement on clinical grounds.
High viscosity (HV) cement was developed to mitigate safety concerns associated with low viscosity (LV) cement by allowing for greater control of speed and location of cement placement within the vertebra. Although complications associated with PVP are rare, they can be serious if cement leakage occurs within the spinal canal or vasculature. In such instances, hardened cement can cause paralysis or lead to a fatal pulmonary embolism. The selection of cement used during PVP and BKP procedures should be left to clinician discretion and be informed by individual patient characteristics.	See previous four responses.
Recognition that the so-called "Sham" treatment arms within the Buchbinder and Kallmes studies are interventions referred to as "Operative Placebo with Local anaesthesia" (OPLA): We were encouraged to see that the Assessment Report appropriately classified the invasive control procedures (so-called "sham" arm) of the Buchbinder and Kallmes studies as interventions, referred to as OPLA. (6, 7) Both studies involved the placement of needles and administration of anaesthesia (either directly to the spine or via the hand/forearm) and were as such, invasive procedures with potential therapeutic effects beyond that obtainable with a true placebo.	There has been extensive debate on whether the improvements observed in the OPLA groups of Buchbinder and Kallmes arise from a placebo mechanism, or whether the anaesthetic provides an actual physiological effect. The use of the term OPLA sought to address both these possibilities.
However, the AR is inconsistent in its classification of the control arms from the Buchbinder and Kallmes studies. Despite having identified the "Sham"/ placebo arms within these studies as an "Operative Placebo with Local anaesthesia" (OPLA), the AR at times refers to these as placebo comparators, and suggests that results of these studies raise questions about the clinical benefit of PVP. This inconsistency in classification makes it difficult to interpret the report	We apologise for this inconsistency and any difficulties this presented in interpreting the report. These should not have been reported as 'placebo comparators', but rather as 'OPLA comparators'. The discrepancy arose because the AG changed terminology part-way through writing the report and did not change the term 'placebo comparator'.
VCF incidence rates in the AR were neither relevant nor recent: The AR model uses data on the incidence of vertebral compression fractures (VCFs) drawn from a large scale prospective Scottish study. (3) Although these are UK-specific data and report VCF rates, this data set is not truly reflective of NICE's population of interest for this appraisal, namely England and Wales and could include patients with VCFs due to malignancy and trauma. Although published in 1998, Singer et al analyse a two year period during 1992 and 1993, a time span over which meaningful changes in screening and diagnosis of VCF may have transpired. (3) In contrast, Johnson and Johnson Medical extracted data from the NHS data source (SUS) by Dr Foster Intelligence, which relates to a much more recent and thus relevant period between	In the AG model future fractures were included to allow future osteoporotic related costs to be considered. As such this was deemed to include all vertebral fractures, not just those which may require vertebral augmentation. It is the same logic that allowed hip fractures to be incorporated despite not being a vertebral fracture that requires vertebral augmentation. The comments on the appropriateness of the Singer data is noted.

April 2010 and March 2011. (2) Further, this data source allows for the specific identification of patients with symptomatic osteoporotic VCFS presenting for hospital treatment either non-surgical or an intervention such as PVP or BKP. *(See Appendix A).* [Appendix A not presented here]	
Calculation of Length of Stay is not limited to osteoporotic VCFs: The Assessment Group (AG) relied on the length of stay (LoS) provided by Medtronic and sourced from the standard HES data. These data are likely to overstate LoS by including VCFs of non-osteoporotic etiology (for example, trauma or tumor). In contrast, the Johnson & Johnson submission identified average LoS for the specific population of patients with osteoporotic fractures within the HES / Dr Foster dataset. Both HES and the Dr Foster data use the Secondary Users Service (SUS) data which are routinely collected in the NHS, so the source is the same.	If the analyses from Dr Foster do indeed use the same dataset as HES then this would be a valid point. Within the original submission this was not stated, and it was unclear whether just a subset was used, which led to the decision to use the HES data. It is commented that much more detail is provided in the response to the AR than in the original submission. The costs assumed by the AG were £686 more for PVP, £144 more for BKP and £722 less for OPM than that assumed by Johnson and Johnson. However, there remains the discrepancy between the lengths of stays reported by Dr Foster, and the typical lengths of stay reported by those practicing vertebral augmentation. Additionally lengths of stay are not reported in the trials, potentially implying these were not considerable.
Inadequate clinical input; Reliance on one clinical advisor: The AG appears to have relied extensively on the opinion of one clinician, who reported routine, first-line use of facet joint injections for treatment of osteoporotic compression fractures. This practice is not representative across the wider clinical community in England and Wales.	Facet joint injections were labelled as an exploratory analysis and were seen to improve the cost-effectiveness of vertebral augmentation. Following this comment additional enquiries were undertaken to gather more information on the prevalence of routine facet joint injections in the UK. 35 vertebroplasty practitioners were emailed to ask if they routinely screen patients by performing facet joint injection prior to considering cement augmentation. Four emails bounced, and ten centres did not respond within the required timescale. Of the 21 centres that responded (locations ranging from Exeter to Dundee) ten routinely performed facet joint injections, eleven did not.
The Clinical Advisor estimated the average selling price (ASP) of lower-viscosity cements, which is quoted in the report. Unless validated, this ASP cannot be taken as a robust data point for inclusion in the modeling as it might apply only to a subset of NHS providers eligible for volume-based discounts, rather than mean ASPs across all NHS Trusts. We therefore recommend the ASPs are validated by NICE using NHS procurement data.	We apologise that this statement was possibly ambiguous. The clinical advisor did not estimate the ASP of lower viscosity cements, but estimated the equipment that would be required to perform the operations. This was then multiplied by the list price that was obtained from NICE. It is commented that the AG replicated the manufacturer's analysis (which

	assumed high-viscosity cement for all) as a sensitivity analysis. The base case assumes high-viscosity only in those deemed appropriate by the clinician (15% of PVP patients). It is expected that the additional analysis will be beneficial to the appraisal committee.
Given PVP is being considered as a class, it would be more appropriate to use an ASP calculated across all cement types.	The analyses undertaken for PVP used a weighted average list price of high and low viscosity cements, assuming that high viscosity cement was used in the 15% of patients where it was deemed appropriate by the clinician. The additional analyses evaluated the gains that would be required for high-viscosity cement to be used for all patients and for this strategy to be cost-effective.
Typographical Error. The Assessment Group (AG) comments that there appeared to be a typographical error in the model submitted by Johnson and Johnson Medical; "only 10% of patients receiving BKP were assumed to consume operating room resources; it was assumed that this value was intended to be 100%. As such the overall cost-effectiveness results are likely to be favourable to BKP". This appears to be correct (i.e. there was a typographical error) and this means that the operating phase of the costs for BKP were too low (by £247) but ONLY when the Strom et al costs were considered, this error favours BKP (i.e. creates bias against PVP).	We concur. We apologise if it is not clear that we were discussing the typographical error in the context of the costs from Strom et al.
Acquisition Costs of Medical Devices. Price points are not static with medical devices and are often reflective of a cost to serve or commitment to volume-based agreements. In accordance with the NICE Methods Guide for Technology Appraisals, list prices are the appropriate benchmark for modelling cost-effectiveness. However there must be a recognition that significant discounts are available to the NHS from the quoted list prices based on volume purchased. Unlike with pharmaceuticals, medical devices pricing is not fixed for the life of the patent. Thus, as observed in other competitive markets, such as consumer electronics, prices decrease substantially over time. As a result, acquisition prices and cost per QALY relative to procedures without medical devices will decrease during the appraisal process and thereafter due to market forces. Should NICE rely on using "Average Selling Prices" sourced from Company submissions; it would be prudent to cross-reference these with NHS procurement data to confirm their applicability across the NHS.	The AG make no comment on this apart from to confirm we have used list price within the evaluation
Conclusion There are very few alternative treatment options to PVP or BKP for patients unresponsive to non-invasive management other than months of severe pain, restricted mobility and poor quality of life and depression in up to 40% of cases. It is reported that patients with VCFs are confined to bed nine times more often than those without VCFs, increasing their risk of further VCFs which can further complicate recovery. The impact of VCFs on quality of life (QoL) has been shown to be comparable with chronic obstructive pulmonary	The AG make no comment on the majority of these statements, which appear to be aimed at the appraisal committee. We do comment that our analysis on low cost cement is not intended to result in low viscosity cements to be used in all patients, but was exploring the cost effectiveness (or not) of using high-viscosity cement in all

disease (COPD), a point which should not be underestimated. **Vertebroplasty (PVP) has been shown to be a clinically effective intervention** for people with painful osteoporotic VCFs refractory to analgesic treatment. VBA performs significantly better in un-blinded trials than NIM (non-invasive management) in terms of improving quality of life and reducing pain and disability and may be one way to mitigate some of the problems associated with the NIM **Vertebroplasty (PVP) has been shown to be a cost-effective intervention** across the range of list prices for cements with high- and low-viscosity, having an ICER well below £20,000 per QALY in the Assessment Group's foundation analysis and in the base case in the model submitted by Johnson and Johnson Medical. High viscosity (HV) cement was developed to mitigate safety concerns associated with low viscosity (LV) cement. There is RCT evidence which substantiates the difference in venous leakage rates (which although rare, can be clinically significant) between High and Low viscosity cement. However it is recognised that the evidence base concerning cement viscosity is emerging and therefore it is recommended **that the decision on which viscosity cement is suitable to different patients, is left to the judgment of the clinician.**	patients rather than those where it was thought clinically appropriate (15%). Within Johnson and Johnson Medical's submission it was implicit that all patients would receive high-viscosity cement, indeed the word judgment (or judgement) did not occur within the main text of the initial submission indicating that the sentiments conveyed in the words in bold are new, even were the intention to include these originally. It is noted that Johnson and Johnson submission also implicitly assumed that the control procedure ("sham") would also use high-viscosity cement, which would appear to provide no benefit over low-viscosity cement in this use.
Medtronic	
Comment	Response
The heterogeneous nature of the OVCF population is not appropriately described in the report, nor are the implications recognised. Vertebral Augmentation (VA) procedures – PVP and BKP - are only appropriate for a sub-set of the scoped population. These are acute (≤ 6 weeks) symptomatic (at least VAS pain ≥ 5 correlating with fracture) vertebral compression fractures. BKP, in particular, is considered appropriate for patients who have proof of on-going fracture process and spinal deformity (Anselmetti 2012). These criteria are similar to the inclusion criteria used in the more recent larger RCTs on VA procedures, (FREE1, VERTOS II2) and [three] comparative non-RCTs. These criteria represent current NHS clinical practice; whereby hospitalised patients are referred to spine surgeons for further examination and diagnostic work-up due to their level of disability. In general, VAS pain ≥ 5 correlating with fracture, i.e. positive MRI STIR image revealing oedema and x-ray showing vertebral collapse, deem a patient suitable for vertebral augmentation.	The claim that PVP and BKP are only suitable for acute (≤ 6 weeks) and severe (≥5 VAS) cases, has been made repeatedly in commentaries on the Buchbinder and Kallmes trials. This assertion is, however, contradicted by an Individual Patient Data meta-analysis from Staples et al. When these investigators combined the data from Buchbinder and Kallmes, the combined data provided > 80% power to assess whether vertebroplasty had a 2.5 unit advantage over control for patients with acute fractures or severe pain. Patients with acute fractures or more severe pain still failed to demonstrate a benefit of PVP over local anaesthetic. Furthermore, as highlighted on p. 145 of the AR, several large case series seem to suggest there is no association between fracture age and clinical outcome. The reviewers suggest that VERTOS II applied more stringent inclusion criteria. However, while VERTIOS II claimed to include only patients with ≤ 6 weeks' pain, the delay between recruitment and performance of PVP

	(9.4 days ±8.1) meant that many patients would have pain of more than 6 weeks' duration by the time PVP was performed. Indeed, if these suggested entry criteria were to be strictly applied, most of the evidence would have to be discarded. Several open-label studies did not specify a VAS cut-off point of ≥5 (Liu, Rousing, VERTOS, Farrokhi); only three reported attempting to localise pain (Farrokhi, VERTOS, VERTOS II), and four included subacute (>6 weeks) fractures (Rousing, VERTOS, FREE, Farrokhi). Finally, the suggestion that this criteria more closely mirrors NHS current practice contradicts the suggestion given by the AG's clinical advisor that vertebral augmentation is typically performed around 3 months after the VCF which is a greater duration from fracture than deemed appropriate by the manufacturer.
Failure to recognize the majority of current NHS clinical practice has led to selection of inappropriate population and therefore studies The AG's failure to recognise the appropriate patient population for vertebral augmentation has led to the inclusion of inappropriate studies in its review and meta-analysis. In particular, the INVEST and Buchbinder studies are not consistent with the indication for the procedures under consideration as they include a significant proportion of OVCF patients who would not be considered clinically appropriate for a vertebral augmentation procedure in the NHS. This flaw in the design of these trials has been extensively pointed out. (cf. section 5.3 Medtronic submissions) but is not reflected in the report. The individual-patient level meta-analysis conducted on the aforementioned studies8 reported that 24 participants were required in each treatment group to show a 2.5 unit reduction in pain scored; however, only 25 of 106 PVP patients in the meta-analysis had onset of pain before 6 weeks – i.e. Acute fractures. Furthermore, it is unknown if all these patients had severe pain at baseline, or if they were a mix of patients with mild, moderate or severe pain. Equally important, fracture severity was not reported in the INVEST study and only 23% of fractures in Buchbinder's study were reported as "severe", albeit the staging of fracture severity is not provided so not comparable to other studies. Hence the clinical review in the report is potentially misleading by including these studies. This is particularly the case given that the AG considered these studies to be the best quality trials available to evaluate vertebral augmentation. Furthermore, all parameters in the AG's model that are estimated using the results of these trials are unreliable. This is of	Given that the significance of these subgroups (fracture age ≤6 weeks, pain score ≥5) is questionable, and that INVEST and Buchbinder represent the first attempts to perform double-blind, 'OPLA'-controlled RCTs of VA, the AG takes the view that excluding these data from the analyses would not be justifiable. The AG does not accept the claim that the debate around these subgroups was not reflected in the report; it is in fact discussed on pp. 144-146. Further, while it is true that the Staples et al article is unclear as to whether the patients with acute fractures also had severe pain, it was not only the Staples et al meta-analysis which was included in that discussion. A number of observational studies contradict the association between fracture age and outcome, and there are significant issues in using pain ratings as an entry criterion. The debate regarding which trials are most appropriate to use as the evidence base was the reason why the AG presented a large number of analyses and why none were labelled as the base case. It appears that Medtronic are not in favour of certain scenarios (where data from INVEST and Buchbinder are included) and favour other scenarios (where the data from FREE are used directly) but this is an Appraisal Committee decision

significant importance given the predictive analysis of QALY improvement suggested that worse health states at baseline provide larger gains in QALY (Borgström 2012, Medtronic submission Supplementary document 8). In contrast, the model submitted by Medtronic focuses on the relevant patient population by using FREE (>50% of patients had more than 25% of deformity) and VERTOS II (>60% of patients had more than >40% deformity) and estimates parameters based on appropriate clinical evidence. The concern of placebo effect should be weighted not only against the bias of introducing these studies as source for utility gain but also against emerging evidence suggestive of reduced morbidity for VA patients and BKP in particular using more objective measures (Edidin 2012 Morbidity); as well as the 1 year results of the early terminated RCT comparing BKP to PVP (KAVIAR, NCT00323609f) that observes a trend of a lower rate of subsequent fractures in favour of BKP. http://www.clinicaltrials.gov/ct2/show/results/NCT00323609?term=Kaviar&rank=1	and the AG makes no comment. It is commented that QALY improvement is correlated with initial health states (with worse states getting a higher improvement). Importantly the manufacturer's supplementary document (p17) reports that [redacted] Additionally, this hypothesis does not appear supported by the analyses presented in the AR, Fig 23 (p188) which show a correlation between higher initial VAS scores and lower stable VAS scores. However it is acknowledged that this discrepancy could be due to a number of factors such as the use of aggregate data and the EQ-5D encompassing more than stable VAS. It is noted that the KAVIAR results appear to have been published after both the manufacturer's submission and the AR. There is a trend for better outcomes in BKP, however, this was not statistically significant. Our calculations indicate p>0.15 at both 12 and 24 months.
Not scoped and inappropriate inclusion of comparator – Operative Placebo Local Anaesthesia (OPLA) The failure to characterise the correct patient population for vertebral augmentation has also led to the inclusion of an inappropriate comparator–(OPLA) - into the report and model with the suggestion that this sham procedure has the potential to be a second line treatment alternative for OVCFs. OPLA would not be considered appropriate for the vertebral augmentation population clarified in point 1 above. Patients who are most likely to benefit from OPLA are those who develop facet joint pain after the natural healing of their fracture and may experience short term alleviation of their pain due to the injection; albeit the long term impact is less clear and will not address their post-fracture segmental kyphosis	It is arguable if OPLA should be a comparator and thus the AG erred on the side of providing the Appraisal Committee with as much information as possible. Multiple scenarios have been provided that are intended to provide the Appraisal Committee with sufficient evidence regardless of their decision on the appropriateness of OPLA. It is commented that Johnson and Johnson also believed that a control procedure was a valid comparator.
Underestimation of the QALY gain associated with BKP A further limitation of the AG report is that the relative benefit of BKP compared to both OPM and PVP is likely to be underestimated and imprecise in the majority of scenarios modelled. The problem is manifested as follows: Firstly, utilities used for economic modelling were either derived from regression analysis of the VAS pain scale against the EQ-5D ('mapping') or	The initial two problems highlighted by the manufacturer appear to be regarding the choice of the scenario that they feel is most appropriate. This is an Appraisal Committee decision and the AG makes no comment.

by using pooled EQ-5D scores at 4 weeks from INVEST, Buchbinder and FREE studies directly . Mapping utilities from VAS pain may ignore between 40 to 55% of the balloon kyphoplasty effect which relate to economically relevant dimensions of HRQoL - mobility and self-care. This suggestion derives firstly, from dimensional analysis undertaken on patient-reported outcomes instruments from FREE9 and BKP data from SwissSpine Registry revealing the relative contribution of each dimension to the overall EQ-5D value (Borgström 2012, Medtronic submission Supplementary document 8; Borgström SSR analysis 2012e). Secondly, from available exploratory factorial analysis conducted on EQ-5D and ICECAP-O10 as well as EQ-5D and OHS11 suggestive that the scales are more complements than substitutes. As indicated by the AG, mapping has the advantage of incorporating data from all studies and thus will not discard data, although will not be as precise as using EQ-5D directly from the trials. This imprecision is probably relevant, as by removing the INVEST study from the mapping the fit of VAS pain to EQ-5D increases from an r2 of 0.62 to 0.86. The second problem is that the use of a network meta-analysis included studies with meaningful differences between randomised groups in VAS pain scores at baseline. By conducting the meta-analysis in terms of absolute VAS rather than difference from baseline, the results may well be biased. Furthermore, as the FREE study showed the smallest difference between groups at baseline (likely due to a larger sample size, n = 300), the discrepancies in the PVP vs. OPM baseline values may have biased against the BKP vs. PVP comparison. This would impact on the results of the network meta-analysis and the scenarios in which these are used in the cost effectiveness analysis. We would, therefore, suggest that the scenarios using the results of the meta analysis and VAS mapping (scenarios 1, 3, 5) should not be considered in the Appraisal Committee's deliberations.	
Underestimation of the QALY gain associated with BKP The third problem relates to the choice of EQ-5D data for scenarios 2, 4 and 6.The AG have selected individual trials for each sensitivity analysis, forcing them to assume equivalence of BKP and PVP and thus resulting in BKP being dominated in the cost-effectiveness analysis scenarios in which the mortality benefit of BKP over PVP is removed (scenarios 4 and 6). To adequately capture the full utility impact of the differences observed on segmental spinal deformity correction between BKP and PVP, a more sensitive instrument on this dimension is likely needed. For example, the recent analysis on radiographic measurements and relationship with other outcomes from the FREE study observed a significant association of improved physical functioning (SF-36 PCS) with increased correction of segmental kyphosis (Van Meirhaeghe 2012). Furthermore, the correlation analysis of QALY/AUCscore	The third highlighted problem is with the estimation of EQ-5D for PVP when only BKP and OPM were available in the FREE trial. The manufacturer is correct in that the AG have assumed equivalence for PVP and BKP. This decision was taken based on the results of the network meta-analysis (of VAS) which included a small head to head trial of BKP and PVP (Liu). The approach taken by the manufacturer was a form of indirect comparison (using absolute differences rather than relative ratios). This was limited by the availability of the evidence as VERTOS II published QALY differences at baseline 1 month and 12 months, whereas FREE published utility values at baseline, 3, 6, 12 and 24 months. In their

improvement in FREE has shown VAS-pain and RMDQ give modest explanations for the variance in EQ-5D (12% and 15%) and SF-36 utility (18% and 27%). This suggests that these measurements are not appropriate predictors for overall quality of life, at least in comparison with multi-dimensional instruments such as EQ-5D and SF-36 (Borgström 2012, Medtronic submission Supplementary document 8).	analyses the manufacturers used linear interpolation to ascertain the QALY difference at 6 months within Vertos II. The manufacturer's assumptions are arguably more contentious than that of the Assessment Group, particularly given the relatively small differences (without a measure of uncertainty) assumed by the manufacturer. These data are reproduced in Table 42 p169 of the Assessment Report. It is seen that the undiscounted utility gain associated with BKP rather than PVP is less than 0.01 QALY within the initial 2 year period. We believe our assumptions to be reasonable, but acknowledge that they are less favourable to BKP than those of the manufacturer. If an additional 0.01 QALYs were added to BKP within our analyses then BKP would no longer be dominated by PVP in scenarios 4 and 6, although the ICER for BKP compared with PVP would be greater than £150,000 (due to the cost difference of approximately £1,600). We do not have a copy of Van Meirhaeghe 2012 and a full reference has not been provided, therefore we cannot comment on it.
Underestimation of the QALY gain associated with BKP The fourth problem relates to how the relative benefit of BKP is modelled with respect to the mortality effect. Although the mortality effect of vertebral augmentation interventions is considered plausible by the AG, mainly due to the strength of effect, no consideration is given to its plausibility and consistency. The most plausible assumption is that BKP shows a difference in size of effect on mortality, relative to PVP, as Medicare data analysis (Edidin 2012, Medtronic submission Supplementary document 3) adopted thorough propensity score analysis to reduce selection bias, used a large sample size (858,978 patients) and its findings were partially replicated in a smaller European healthcare setting (AOK Niedersachsen German sickness fund, 2.4million insurants in 2011, 3'607 included in survival analysis) (Lange 2012,Medtronic submission Supplementary document 4). Furthermore, given the well-known cascade from a primary vertebral fracture to hyperkyphosis to increased morbidity and mortality, the mortality benefit is most likely to be linked to a meaningful impact on physical functioning subsequent to spinal deformity correction, particularly for this co-morbid patient population. More specifically, the differences in morbidity risks from Medicare (Edidin 2012 Morbidity) that has emerged since Medtronic submission reports that BKP vs. PVP propensity-matched OVCF patients that survived first year had - 16% risk of being admitted to hospital with pneumonia; -22% risk of death with pneumonia; -4% risk of subsequent	For the fourth problem we reiterate that we believe it is possible that there is a causal difference in mortality between patients treated using OPM and those patients that received vertebral augmentation given the size of the effect. However, we state that it is also possible that there is no causal difference. In summary, we do not have sufficient evidence to make a judgement of whether there is an actual difference in mortality between patients undergoing BKP and PVP. This is because: i) matching is sensitive to the matching method used, the size of the comparison group as well as the amount of overlap between the treated and comparison group and the robustness of the results to these issues has not been demonstrated. ii) Matching, even if performed correctly, reduces the selection bias due to observables but does not deal with any bias arising from selection on unobservables. Lange 2012, (Medtronic submission Supplementary document 4) does not address either of these issues and in addition suffers from a small sample size and a very limited number of covariates on which to perform the matching. We do not have a copy of Edidin 2012 (Morbidity) and a full reference has

hospitalisation and -6% risk of Urinary Tract Infections (UTI). Additionally, it is observed that same matched cohorts of BKP vs. OPM (but not PVP) patients had -12% risk of myocardial infarctions/ cardiac complications and -12% risk of being admitted with Deep Vein Thrombosis (Edidin et al 2012).	not been provided and therefore we cannot comment on it. Note: the sample size of 858,978 refers to Edidin 2011 not Edidin 2012 Medtronic submission Supplementary document 3 and; the correct sample size for the matched sample for comparisons between BKP and PVP in the propensity score analysis is reported as 151,277 for the overall sample (traumatic and OVCF fractures).
Inaccurate acquisition cost of balloon kyphoplasty The acquisition cost modelled by the AG is the list price cited for BKP (£2663) in the Medtronic submission which is significantly higher than the average selling price (ASP £1900) as sourced from NHS tender offerings. This tender process is transparent and consistent, with the price offering agreed for a given timeframe in line with tender specifications. Further to an unsolicited request from NICE c/o Stuart Wood (Technology Appraisal Team), we revised our submission to formally release the ASP for BKP from commercial in confidence (CIC) and, under sections 1.11 ,6.5.8, 8.5.4 (table), 8.5.5 (table) and 8.5.9 (table), publically disclosed an ASP of £1900. We also amended our check list to align with this revised submission (26/07/12). Therefore, Medtronic suggest that AG either use our ASP in the foundation scenario or run sensitivity analysis on ICER estimates.	The AG have checked with NICE and it has been confirmed that the list price is the price that should be used in the AR. In order to provide as much information as possible to the Appraisal Committee a quick analysis was undertaken, further to the AR, to see if the conclusions would be changed were BKP priced at the ASP. In Scenarios 1 and 2, BKP was the most cost-effective intervention and a price reduction would not alter this conclusion. In scenarios 3 to 6, PVP dominated BKP (due to a better or identical efficacy) and the cost-difference was greater than the reduction in BKP price were the ASP considered. Thus the conclusions would remain unchanged. If a potential utility benefit of 0.01 (see above) was considered alongside the reduction in BKP price the ICER compared with PVP would be above £50,000 per QALY. This is likely to be favourable to BKP and this comparison uses the list price of PVP rather than ASP.
Synthes	
Comment	Response
We have been surprised by one sentence mentioned at page 29: “Anecdotally, stenting is associated with a greater risk of procedure-generated adjacent fractures, and some operators cement the adjacent vertebrae as a preventive measure.” We wonder where this statement is coming from. As manufacturer of VBS we are absolutely unaware of this. On the contrary, the currently available evidence that we summarized in our submission suggests that the rate of adjacent vertebral fracture lies somewhere around 9%, which seems similar or even somewhat lower than the rates reported in the literature for	On reflection, we agree with the manufacturer and retract this statement.

vertebroplasty and balloon kyphoplasty. For example, Kasperk et al. reported 9.7% of adjacent level fractures after 3 years when using kyphoplasty and Mudano et al. reported 18.8% after 1 year for vertebroplasty and kyphoplasty. We would therefore strongly suggest that this sentence is removed from the report, particularly considering that there is no source given. We consider that statements that are "anecdotal" and without any supporting evidence should not be put forward in a Multiple Technology Assessment by NICE. [Further evidence presented to support the case have been omitted]	
Regarding the two double-blinded RCTs13 14 we were very pleased to see that Assessment Report acknowledges that the comparator (local injection of anesthesia) was not a real "sham" procedure but rather a pseudo treatment form which is rightly described by ScHARR as "operative placebo with local anaesthesia" (OPLA). It remains nevertheless doubtful whether such a treatment form can produce more than short term pain relief.	No comment required.
British Society of Skeletal Radiologists	
Comment	Response
I am a musculo-skeletal radiology consultant of 16 years with an interest in spinal intervention. I have been performing vertebroplasty and kyphoplasty for over 10 years. Below are my thoughts from observations during that time and based on reading the current literature on the subject: Osteoporosis is a complex demineralising condition with a wide spectrum of clinical severity which affects many, predominantly elderly patients with 750,000 new vertebral fractures occurring in the United States per year. The majority of osteoporotic vertebral fractures will heal without long term sequelae and a significant number will have occurred sub-clinically, but a small proportion remain painful and in some cases result in severe debilitating pain and progressive deformity. It is these cases which need to be recognised at an early stage and treated more aggressively with vertebral cement augmentation to prevent progressive deformity and to relieve the severe pain. Epidemiological studies support the fact that osteoporotic vertebral compression fractures are not as benign a clinical entity as perhaps originally thought. Mortality rates are observed to increase significantly when the number of vertebral fractures increase, (Kado DM et al) and	A considerable amount of this response is background and has not been commented on. We do comment that the AGp model did take mortality associated with future fractures into consideration. It may be true that a proportion of fractures leading to chronic, severe pain and progressive deformity could benefit from early intervention. However, to our knowledge, there are as yet no well-validated methods to identify such patients *a priori*. In the absence of such techniques, the concern would be that some patients were treated unnecessarily with invasive procedures bearing a small risk of serious complications. This risk was highlighted by the recruitment pattern in VERTOS II, in which 53% of potentially eligible patients became ineligible during the course of the screening process due to spontaneous pain relief. The low participation rate in the Buchbinder and Kallmes trials is a limitation, and it may be true that the Buchbinder and Kallmes studies were somewhat underpowered. However, this is to some extent open to interpretation. The Kallmes study had >80% power to detect a difference

the conclusion from the study by Suzuki et al was that, instead of the generally believed good prognosis for the greater majority of those with vertebral fractures, the acute vertebral body fracture was the beginning of a long lasting severe deterioration of their health. It is this heterogeneity of the clinical spectrum which has made evaluating the efficacy of treatments such as vertebroplasty and kyphoplasty for osteoporotic vertebral fractures difficult. The studies by Buchbinder and Kallmes are both randomised, controlled, double blinded trials but are significantly under-powered studies and certainly in the case of the Kallmes study, the inclusion criteria are not as stringent as might appear at first glance. The original power calculation suggested that 294 patients should be included in the study and when the study was terminated prematurely only 131 patients had been enrolled, and it should be noted that 1682 patients were excluded from the study for a variety of different reasons. In the Kallmes study, at one month, clinical improvement in patients with painful osteoporotic vertebral fractures was similar among those treated with vertebroplasty and those treated with local anaesthetic injections in and around the posterior elements of the painful segment. Similarly, the Buchbinder study concluded that there was no significant early benefit from vertebroplasty over a sham procedure (local anaesthetic injection into the posterior para-spinal tissues among patients with recent osteoporotic fractures. It should be noted that potential trial participants for both studies, when presented with a choice between an apparently established treatment (vertebroplasty) and a clinical trial which might mean no effective treatment, are likely not to enter either trial. It is therefore probable that this will result in exclusion of the most symptomatic patients who are perhaps most likely to benefit from vertebroplasty. My concern, therefore, is that these studies do not address satisfactorily the sub-population of patients with vertebral compression fractures at the severe end of the spectrum who may progress to rapid deformity and multiple vertebral compression fractures over a short period of time. The Vertos and FREE studies are also not without their limitations. These are not placebo-controlled trials and there is therefore still doubt about the mechanism of the effect of the intervention being evaluated. On the other hand these are adequately powered studies looking at patients with more clearly defined early vertebral fractures with higher pain scores, compared to the randomised controlled trials of Kalmes and Buchbinder.	of 3.0 in the RDQ and a 1.5 point difference on the VAS at 1 month. The Buchbinder trial reported a sample size of 24 participants per group would be required to detect a 2.5 point advantage of PVP over pain with an SD of 3.0. The sample size of 78 therefore provided a large enough sample size to address the primary aim of the study. With respect to the potential sub-group of patients with acute fractures ($\leq$ 6 weeks old) or more severe pain ($\geq$8 VAS score), this was addressed in the Individual Patient Data meta-analysis performed by Staples et al. When these investigators combined the data from Buchbinder and Kallmes, the combined data provided > 80% power to assess whether vertebroplasty had a 2.5 unit advantage over control for patients with acute fractures or severe pain. Patients with acute fractures or more severe pain still failed to demonstrate a benefit of PVP over local anaesthetic. Furthermore, as highlighted on p. 145 of the AR, several large case series seem to suggest there is no association between fracture age and clinical outcome. We concur that pain in VCF is complex multifactorial phenomenon, and the adjunctive biomechanical processes mentioned by the reviewers may help explain why local anaesthetic caused such a high response in the Buchbinder and Kallmes trials. However, we are unaware of any validated procedures which could be used to identify such patient subgroups. At the very least, these biomechanical processes highlight the need for caution in seeking to provide clinical benefit through cement injection when comparable improvements could be gained for many patients from less risky procedures. In addition we have presented multiple scenarios to be considered by the Appraisal Committee in their deliberations which explicitly use different evidence sources.

Perhaps the unifying conclusion from all of these studies is that at the very least, patient selection is critical if vertebroplasty is to be found to be effective and that if at all possible, more (ideally better powered) randomized controlled trials are required. Spinal pain in the setting of vertebral compression fractures is a complex, multifactorial phenomenon. As well as pain from the micro-movement at the fracture plane, it is likely that sagittal imbalance from the kyphotic deformity results in operation of the biomechanical stresses on the posterior elements and paraspinal muscles, and it is therefore logical to understand that local anaesthetic injections into these areas may have a short term effect on pain and potentially a more intermediate term effect by breaking complex pain cycles. Further well designed studies will be difficult to orchestrate and complete, but are essential to understand more comprehensively the complexities of vertebral compression fractures. What is most important is to identify the subset of patients in whom early intervention with percutaneous cement augmentation is likely to be beneficial both in terms of pain relief and prevention of progressive deformity.	
The British Pain Society	
Comment	Response
This is a well considered document, which seems to cover all facets of using Vertebroplasty (PVP) & Kyphoplasty (BKP) (although this is a technique which I do not currently perform). It is a very large document and is sometimes difficult to keep track of. This is a very comprehensive review running to 417 pages. The assessment report seems to have reviewed most important studies in the field & explained why they have not included others. The report provides a good overview of osteoporosis and vertebral compression fractures and the short-term and long-term problems associated with them. The report provides a summary of techniques of vertebroplasty (PVP) & balloon kyphoplasty (BKP).	No Comment required
It appears that NICE will come down in favour of PVP & BKP with the caveat that 2 RCTs where patients were blinded did not show an improvement compared to sham procedures (see below). The 2 RCTs mentioned above have both been criticised in this document and elsewhere, mainly for the fact that the sham procedure was potentially not sham at all.	The AG cannot speculate on the decision of the Appraisal Committee
There is detailed discussion of the evidence limited to RCTs 9 studies were considered in total: 6 considered PVP against optimum pain management, 1 considered PVP against BKP and	No Comment required

2 considered PVP against a sham (placebo) procedure (Buchbinder et al, NEJM 2009 & INVEST, NEJM 2009). Both of these latter 2 studies were double blinded and appeared in the same edition of the New England Journal of Medicine. First 7 studies favour PVP. The last two showed no significant benefit when compared to sham. However, criticism of these 2 studies have been made. Specifically that the PVP groups had generally lower volumes of cement than is usually the case and that sham was local anaesthetic technique, which may not have been a true placebo.	
One of the references cited in the document (Wilson et al, European Radiology) suggests the use of facet joint injections as a potential treatment prior to PVP & BKP. Although in principle I would have no objection to this, I feel that that the implication is made in the report that these are minor procedures with little in the way of complications. As we know this is not the case. Although I would be happy to see Facet joint injections put forward as a potential treatment, I think that the final assessment report should emphasise that these are procedures with potentially serious consequences albeit not on the scale of PVP & BKP.	The use of facet joint injections has been labelled as an exploratory analysis. Our clinical advisor has reported no complications within his career (over 25 years) but we have noted the comment made, and that we have a small sample size in terms of practitioners.
Healthcare Improvement Scotland	
Comment	Response
The authors have conducted a very comprehensive review of the literature. The conclusions are that Vertebroplasty and Kyphoplasty clearly provide better short term pain relief that standard care in patients with recent vertebral fracture. However the Buchbinder and Kallmes studies (and associated network meta-analysis, which attempts to extend this to KP) demonstrate that it is unclear to what extent these benefits are due to the augmentation procedure (VP or KP) or other components of the technique (such as going to theatre, having sedation, having local anaesthetic injected into the spine and so on). In other words, this could be a placebo effect. There is nothing wrong with a placebo (it can be helpful in many areas of medicine) but when the intervention carries the risk of serious adverse effects such as embolism, spinal cord compression and so on (as these interventions do), then the risk benefit is doubtful. Therefore my opinion is that these interventions cannot be supported as routine treatments of patients in NHS Scotland. I should say that we have been reviewing all this evidence lately as part of the SIGN osteoporosis guideline and the group collectively came to the same conclusion as I have	No Comment required

outlined above.	
I am unable to find any fault with the content or methodology of the document (though I am not qualified to assess the accuracy of the statistical analyses). The conclusions and recommendations also appear logical and reasonable. NICE IPG's 12 and 166 both endorse the creation of multidisciplinary teams to assess these patients and refer for treatment appropriately, and I would have liked to see more emphasis on this in the document. One of the main reasons for the apparently equivocal results of treatment is likely to be the difficulty in confirming the cause of pain and ensuring that suitable patients receive treatment promptly, given the haphazard and fragmented nature of current service provision.	The AG believes that the components of multidisciplinary teams were outside of the scope of this assessment, which focussed on the procedures themselves. Thus, the advice on this provided in IPG 12 and 166 should not be considered to be affected in any way by the AR.