Author's response to reviews

Title: Balloon Kyphoplasty in malignant spinal fractures: A Systematic Review and Meta-analysis.

Authors:

Carmen Bouza (cbouza@isciii.es)
Teresa López-Cuadrado (teresalc@isciii.es)
Patricia Cediel (pcediel@isciii.es)
Zuleika Saz-Parkinson (zuleika@isciii.es)
José María Amate (jamate@isciii.es)

Version: 3 **Date:** 3 June 2009

Author's response to reviews: see over

Dear Sir,

We sincerely appreciate your interest in our above-referenced manuscript entitled “Balloon Kyphoplasty in malignant spinal fractures: A Systematic Review and Meta-analysis”. We should also like to thank the reviewers for their comments, which have helped us improve the text. Indeed, the document has been amended in an attempt to incorporate many of the ideas expressed.

As to the specific comments of the reviewers, we provide point-by-point replies below.

We hope you will find these changes to your liking, and we will look forward to hearing from you.

Yours sincerely,

Carmen Bouza, MD, PhD

Referee 1.

1. **English is not the first language of the authors, and I think there is room to both improve clarity and reduce word count.**

We have tried to do this with the support of native-born English translator specialised in scientific biomedical literature.

2. **Methods. Observational studies can be difficult to find by electronic searching [1, 2]. Some people find citation tracking helpful.**

This is true and to increase the literature search sensibility, several electronic databases were used as well as the references of the studies identified. However, the purpose of our study was not to exclusively select observational studies but instead to identify all the available evidence found in the literature regarding balloon kyphoplasty in the management of tumoural spinal fractures. For this reason, the search strategy outlined did not include a limit regarding the study design. There is no doubt that had randomised controlled trials been located, these would have been the object of our analysis. However, and despite the extensive bibliographic search carried out, only observational studies were found.

In any case, and taking into account your suggestion, we have incorporated a summarised version of our search strategy.

3. **As in many palliative care treatments, there is a dearth of high quality evidence from large double blind randomised controlled trials. There are many well-documented reasons why it is difficult or impossible to conduct such trials in this population. That should not deter us from examining other evidence, and indeed observational studies can provide equally robust evidence provided criteria of quality, validity and size are met. The authors have clearly tried to assess potential bias from various sources, and comment on the methodological quality of the studies. However, apart from mentioning the “limited number of patients” at the beginning of the discussion, they make no mention of the susceptibility of the included study results to the random play of chance due to small numbers [3].**

We agree with this comment and have included the reference provided. We have also added a sentence at the end of our discussion recognizing that our results are susceptible to the random play of chance due to the small sample size of the original studies found in literature.

4. **It is unfortunate that the majority of the data for pain relief are presented as population means, which can mask significant differences between individuals. In the two studies for which adequate data are available, most patients do seem to have severe pain at baseline (>6), which is reduced to none or mild (<3) postoperatively. This kind of clinical interpretation gives more meaning than simply saying there was a statistically significant reduction. What matters is how many patients return to a tolerable level of pain, not whether the difference was statistically significant. Similar comments relating to use of mean data and statistical versus clinical significance apply to Functional Capacity, Quality of Life, Vertebral height and Kyphotic angle. The data presented do not tell me what difference the procedure made to the patient. I appreciate, however, that the original studies probably reported only mean data.**

We completely agree with this comment regarding the importance of clinical results and their individual application. In fact, one of our organisation's major commitments is to identify clinical benefits and ensure patient's safety when using new technologies. This is particularly important in this case, given on the one hand the frailty and comorbidities of the target population and, on the other, the serious consequences of non-treatment of a vertebral fracture.

At the same time, we also recognize that we must not lose sight of the inevitable methodological limitation of any study based on sample or population data, namely, that group outcomes may differ from those for the individual, thus constituting the so-called "ecological fallacy" phenomenon [Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. Ann Intern Med 1997;127:820-826.].

However, as mentioned in the reviewer's comment, the original studies provide only group data and do not provide individual data.

Compulsory revisions

1. **Abstract This is a little wordy, but lacks important information. For example: "Methods: We searched several electronic databases to September 2008, and reference lists of relevant studies for studies of any design that reported on balloon kyphoplasty in patients with spinal fractures secondary to osteolytic metastasis and multiple myeloma. Outcomes sought included pain relief, functional capacity, quality of life, vertebral height, kyphotic angle and adverse events. Studies were assessed for methodological bias, and estimates of effect calculated using random-effects model. Potential reasons for heterogeneity were explored."**

This has a slightly higher word count than the original, but gives a clearer picture of what the review is about, and word count can be reduced in Results and Conclusions.

We appreciate this comment and have adjusted the abstract to the given proposal.

Results: Claims of efficacy should perhaps be tempered with "in these studies", since the data is so limited.

We have added the suggested wording to our original sentence.

2. **Background**

In the first paragraph the authors state that "spinal fractures significantly increase the risk of new fractures". Is this true? Or is it the case that those who have had one fracture are already predisposed to fractures, and hence more likely to have another one?

It is true that patients who suffer a spinal compression fracture have a real predisposition to develop a new spinal fracture which could perhaps occur because they suffer from osteoporosis, a tumour or are receiving corticosteroids. However, different authors have shown that the presence of spinal fractures, even those classified as mild or moderate, are accompanied by a greater risk of presenting new fractures [see reference 8 of our text, as well as Delmas PD, Genant HK, Crans GG et al. Severity of prevalent vertebral fractures and the risk of subsequent vertebral and nonvertebral fractures: results from the MORE trial. Bone 2003; 33:522-32].

3. **Methods**

Clinical outcomes sought are poorly described. Did you look for any information on any of these outcomes? If so, did you have prespecified criteria for which were your desired outcomes and/or what a clinically useful outcome would be? For example, did you look for particular validated scales? Did you accept undefined "improvement" in pain, and was pain patient-reported or investigator reported (reduced validity)? The recent IMMPACT consensus statement suggests measures of change in pain and other outcomes that have clinical relevance in chronic pain conditions [4].

The selection of the outcomes of interest was based both on previous experience of our research group, as well as the revision carried out in specialised literature on application of minimally invasive techniques in treating spinal fractures. In this sense, standard clinical results regarding procedure efficacy include assessment of changes in pain, functional capacity, vertebral height and kyphotic angle. The scales applied to evaluate pain and functional capacity are objective, consistent and validated, as well as being the ones most frequently used.

As regards pain, except in the study by Vrionis, all other studies included in our report assess patient-reported pain in an objective manner using the VAS.

4. **Results**

a**. Why were references without abstracts excluded?**

Though systematic reviews should not routinely exclude references without abstracts, this subject still remains a controversial issue. However, abstracts are often used to eliminate clearly irrelevant reports, obviating the need to obtain the full text of those reports and thus reducing the time and costs of the searching process.

Please identify the included studies (e.g. "These studies, three retrospective [refs] and four prospective [refs] single-centre….").

Although the type of design of each study is outlined in Table 1, we have incorporated, as suggested, the corresponding references within the text.

One of the studies excluded because of data duplication in subsequent or more complete publications [your ref 24] is cited when reporting details of the patients' condition (2nd paragraph). Is this a mistake, or did only that publication provide that data?

This is a mistake which has been corrected.

b. On what basis are the studies described as "representative"?

This wording refers to one of the items evaluated to analyse the selection bias risk in non-experimental studies. In this specific case, it refers to the fact that selected studies include patients with

characteristics similar to those of the general population of patients with painful tumoural spinal fractures and, therefore, can be considered representative of this target population.
Since another of the reviewers asked us to incorporate the inclusion and exclusion criteria, these have been added to Table 1 so this aspect of representativeness of the population can be observed.

c. Pain Relief.

• Was pain relief patient-reported? Investigator reported pain estimates frequently do not correlate well with patient reported estimates.

• What is meant by "improvement"? Is it clinically relevant improvement?

Except in the study by Vrionis, all other studies included in our report assess patient-reported pain in an objective manner using the Visual Analogue Scale (VAS). It is, in fact, a clinically relevant improvement. Vrionis, however, does not indicate who measures the observed improvement nor is the measurement instrument employed described.

d. Safety. Cement leaks. I question the value and validity of combining these data to give a statistical output, given the small number of events.

Assessment of cement leaks is essential in the analysis of safety in the management of spinal fractures using minimally invasive techniques, and therefore in assessing BKP. It is known that cement leaks may potentially cause serious clinical events such as pulmonary embolism and other cardiopulmonary complications. In addition, any cement extravasation into the spinal canal or the venous system is a cause for concern because it is not known if there are other possible systemic effects or clinical consequences, such as development of new spinal fractures [Lin EP, Ekholm S, Hiwatashi A, Westesson PL. Vertebroplasty: cement leakage into the disc increases the risk of new fracture of adjacent vertebral body. AJNR Am J Neuroradiol 2004; 25(2):175-80.]
The small number of cement leaks is, in fact, one of the theoretical advantages of BKP, but no doubt this potential advantage needs to be checked in an analytical way, which is why it was included in this study.

5. Discussion

a. **The limited nature of the results, in terms of numbers of patients/procedures and outcomes reported should be acknowledged. The fact that they are in broad agreement with those in osteoporotic fractures may give added weight and increase our confidence in them**.

b. **Paragraph 1. I think "assumptions" is the wrong word. The authors probably mean "conclusions", although I would not like to draw firm conclusions from this limited data. Perhaps "Nonetheless, combined analysis gives results that are in broad agreement with earlier reports that focussed on osteoporotic fractures".**

Thank you for your suggestion, we have included the sentence in our manuscript.

c. It would be helpful to state what pain relief value is considered clinically significant in Jensen (ref 38), and also to refer to the recently published IMMPACT recommendations [4].

Jensen points out that a pain reduction of 33% is the threshold to determine whether a given therapy is clinically useful for the patient.

The IMMPACT study points out that a change of 2.0 to 2.7 points in the Visual Analogue Scale, which is equivalent to a reduction between 30% and 41%, is the threshold for important changes in pain, and therefore clinically relevant for the patient.
Although we had initially thought of including both references, after careful consideration of the data in the IMMPACT study and those from the Jensen study, we have decided that it is more appropriate to exchange the Jensen study by the IMMPACT study. Although, in essence, the determination of clinically significant pain relief is similar, the IMMPACT study is better adjusted to the units we have used in this work, it is a more recent reference and it is a consensus conference.
d. Safety. These studies were not designed to evaluate adverse events, so although the information they provide is useful, and may give insight into potential common events, they cannot evaluate less common, and potentially serious events. This should be acknowledged.
We agree that the studies may not analyse adverse events which are rare or with few signs or symptoms and this may lead to a lack of information regarding the technique's safety. However, we believe that all authors would have been able to recognize serious adverse effects, given their clinical repercussion and the existing data in the literature regarding possible complications of this technique even if they refer to a different pathology such as osteoporosis. Whether or not these have been described in their publications is another matter, but we understand that the adverse effects described by the authors are the ones that their patients truly presented.
e. **Conclusions. Future studies should have consistent reporting of clinically useful outcomes.**
We have included the suggested sentence.

Minor revisions
Table 1. **Add BKP and VP in footnotes.** We have added these.
Table 2.
Functional capacity and Quality of Life: it would help to state the range of the scale used, as is done for VAS. We have done this, although with this modification, Table 2 is now Table 3
Be consistent with abbreviations for weeks, months, years. We have corrected these inconsistencies.
State (in footnote?) that continuous data is presented as means ± SD. We have done so.
Table 3. Give range for VAS and ODI. We have done so.
Figure 1. Exclusions do not clearly tie in with the text (paragraph 1 Results).
Although our idea was to try not to lengthen the manuscript, we have reworded the text so it ties in properly.

Referee 2.
1) The relative effectiveness of non-surgical management is discussed fairly dismissively in the introduction. I think that the authors have overstated toxicity and undertated potential efficacy of combined modality approaches. I have read each of the 4 references cited in support of this assertion. I come to the conclusions that the authors claims that approaches

involving radiotherapy, analgesia and biphosphonates are so poor or so toxic as to be dismissed from further evaluation exaggerated are not adquately supported by the cited evidence. Some studies of radiotherapy in this setting have demonstrated substantial efficacy for example: [1]. Unless substantially stronger evidence to support this can be presented, I would suggest a much more circumspect approach such as saying that outcomes are variable and that single modality approaches are rarely effective. I=
This concern has implications for both the introduction and the discussion. In the discussion the authors address the future research agenda to define relative efficacy. It would appear that further studies of relative efficacy should probably address this issue relative to non-invasive and as well as other invasive approaches.
We appreciate this comment. Our intention was not to suggest that toxicity was such that these procedures should be dismissed or that their efficacy was lacking completely. Simply we wanted to point out that there is a problem with toxicity, and efficacy is, in many cases, incomplete. In fact, all the included studies make some reference to these aspects of toxicity and refractoriness to pain in many patients, to analgesics, to physical measures, biphosphonates, and radiotherapy.
However, we understand the reviewer's point of view and have modified our sentence, adding a more cautious tone, to the following: "outcomes are variable and single modality approaches are rarely effective".
At the same time, we have added a sentence in the discussion regarding the need to evaluate BKP efficacy as compared to invasive and non-invasive techniques.

2) Another issue for consideration is that compression fractures >50% of vertebral height are commonly associated with evidence of epidural compression [2] [3]. . The paper does not address this issue and it begs many important questions: If there is evidence of epidural compression does this impact on decision making regarding balloon kyphoplasty? Was this an exclusion criterion in any of the studies? Does the approach differ if there is bone fragment or tumor in the epidural space? What is the appropriate imaging modality to evaluate this? Do all patients need MRI evaluation before the procedure?

The exclusion criteria of the involved studies should be clearly articulated and consideration of spinal cord compression addressed.

The presence of medullar compression clearly affects the decision to carry out a BKP, and generally is a contraindication to do so. This aspect was not discussed in the manuscript in an attempt to not lengthen it. However, we consider that it may be important for the possible readers to clearly identify inclusion and exclusion criteria in each of the studies for which we have added this information in Table 1.
Practically all the literature concerning BKP recommends carrying out an MRI before the procedure. In fact, in all of the studies selected, symptomatic levels were identified by correlating the clinical data

with MRI findings of marrow signal changes consistent with compression fractures (marrow edema and fat suppression sequences). We have added this sentence to our manuscript.

Referee 3.

1. Abstract: Delete "square" after I2. Done

2. Background: The phrase "Due to the fragility…" is too long. I recommend dividing it in two statements. (2nd paragraph, 2nd phrase) Done

3. Methods: Clarify what "IME" is for non Spanish readers. For example, write "IME (Índice Médico Español)" instead of "IME" alone. (1st paragraph, 2nd line).

As no studies were found in the IME, we have opted to exclude this from our text.

4. Results:

4.1. Commas instead of semi-colons in the second phrase of Results (1st paragraph) and in the last phrase of the second paragraph of Results (just before Efficacy outcomes). Done

4.2. Functional Capacity: Write "(Table 3)" at the end of the phrase, not in the middle. Done

4.3. Quality of Life: Cite "Table 2" instead of "Table 3". Thank you for this correction which has been made.

4.4. Cement Leaks: Colon after I2, that is, "I2: 0%". Done

4.5. Clinical Complications: The information about complications found by Kose are different in the text and in Table 2, that is, the text shows "two patients…
wound infection and temporary respiratory difficulties…" and Table 2 shows "1
balloon rupture…".

The problem here, as stated in the text, is that it was not possible to know whether the patients with these minor complications, as described by Kose, had received BKP or VP treatment which is why we initially did not include them in the Table but have done so now, as it is truly the most correct way.

5. Discussion: The phrase "To our knowledge, there are no…" is too long. I recommend dividing it in two statements. (last paragraph) This has been corrected.

6. Table 1:

6.1. Show reference citations next to the Study, "Lieberman 2003 (25)", for
example. It helps to the reading. Done

6.2. Kose's row: Colons after BKP and VP, that is "BKP: 18/22" and "VP: 16/26". Done

6.3. Vrionis' row, Estimated age of VF: ¿"10" " means "10 months" perhaps? Yes, thank you, it is indeed 10 months.

6.4. Abbreviations: Add VP, VF and BKP, and their meanings, to the list of abbreviations. Done

7. Table 2:

7.1. Show reference citations next to the Study, "Lieberman 2003 (25)", for example. Done

7.2. There are some data centered in some cells. This has been modified.

7.3. I think you should delete "ODI" in the Pflugmacher 2008's row and write it in
the head of the column, that is, "Functional capacity, ODI", for example. Done

7.4. Fourney's row: Delete º (grades symbol) or show it for the rest of studies. The symbol has been eliminated.

7.5. Take care of use of commas and semi-colons; I recommend substitute all commas with semi-colons in order to show the data more homogenous. Done

7.6. Write "p=0.001" or "p=.001" but do not mix both ways because it's confuse, in my opinion. Sorry for the confusion, we have corrected this.

7.7. Abbreviations: Add NS, ODI and VCF, and their meanings, to the list of abbreviations. Done

8. Table 3:

8.1. Show reference citations in a new column or in the "Studies providing data" column: "(25,33,34)", for example. It helps to the reading. Done

8.2. Take care of use of commas and semi-colons. Done

8.3. Write "0.20" or ".20" but do not mix both ways because it's confuse. Sorry for the confusion, we have corrected this.

8.4. Write "Kyphotic deformity" (as in the text) next to "Cobb angle" in order to help the readers to relate easily the text and the data in the table. Done

8.5. Abbreviations: Add CI and its meaning to the list of abbreviations. Done

Discretionary Revisions:

9. Background: Clarify the meaning of PMMA. (last phrase of 2nd paragraph). We have included the name of the compound.

10. Methods: The reference 23 does not seem "recent". Perhaps the authors could cite other reference. (last phrase of Methods). This reference is a publication from the year 2000, and although it may not be described as most recent, we do believe it is current and is also included in the last version of the Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1 [updated September 2008). The Cochrane Collaboration. [http://www.cochrane-handbook.org]

11. Results, Functional Capacity: Show that the Index decreases when the disability decreases. Perhaps it is obvious but the readers thank usually this kind of clarification, in my opinion. We have included this in the text.

12. Table 2: I would move Kyphotic angle column before Vertebral height column and New VCF column before Clinical complications column, with the aim of keep the rank they appear in the text. We have changed the order of the columns as suggested

Referee 4.

1. Did the authors of this study contact other authors about the work done regarding the research question and if not, why?

Following Royle et al. recommendations (reference number 17 of our manuscript) we contacted experts on the subject in order to adequately focus the review and identify published studies. However, we have at no time tried to contact the company which currently manufactures the technology.

We have added a sentence in the methods section of our manuscript regarding this issue.

2. It is neccesary to make a clear statement about the methods of dealing with unpublished studies

We believe this issue has been clearly mentioned in our manuscript. As noted in the discussion page, this study is limited to peer-reviewed literature and thus, does not include unpublished data. In the same paragraph, we state that, “in line with prior reports, we decided not to include unpublished data from industry given both the difficulties encountered in obtaining this information and the recognition that the use of these data may not necessarily reduce the bias in meta-analysis [39,40]”.

3. **Event here is said that "the possibility of bias in the studies was evaluated using publisher guidelines for systematic reviews", it will be neccesary the report of what type of bias were considered and how were evaluated.**

Regarding this comment we should point out that Table 2 (previously Table 1) clearly shows the biases considered as well as the elements which were considered for their evaluation. As stated in the methods section, such biases (selection, procedure, detection and attrition bias) were analysed following the guidelines established in The Cochrane Handbook for Systematic Reviews of Interventions (bias in non-experimental studies) [reference number 20 of our manuscript]. Such guidelines are publicly accessible and are widely known for which we believe it is unnecessary and redundant to explain them in detail in this manuscript.

4. It will be very useful to know what was the reason to use the random effects model for meta-analysis.

According to meta-analysis methodology literature, there are several reasons to use a random effects model. Firstly, the observational design of the studies increases the risk of statistical heterogeneity and thus, according to the recommendations of the reviewer’s manual of the Cochrane Collaboration as well as other well-known reports [Khan KS et al. Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for Carrying Out or Commissioning Reviews. Report 4 - 2nd ed. York, UK: NHS Centre for Reviews & Dissemination. University of York. 2001], this model is appropriate to estimate the global effect of the procedure. In second place, as we are only dealing with published studies, we cannot categorically assert that our study includes all existing literature and, in general, it is assumed that this model provides a more conservative global estimate. Thirdly, in this setting, using this model allows a greater generalisation of results.

In addition, and although this model was chosen during the methodology design phase of the study, we have done a simulation using the fixed-effects model and the results were not found to differ notably.

5. It will be also useful if there is a clarification about what sensitivity analysis to explore statistical heterogeneity was choosed and not use other method.

Due to the small number of articles found in the literature, a meta-regression was not appropriate in this case. Therefore, the only possible option to explore statistical heterogeneity, was a sensibility analysis. This analysis has been performed following the guidelines established in the reviewer’s manual of the Cochrane Collaboration.