Author's response to reviews

Title: Health-related quality of life after vertebral or hip fracture: a seven-year follow-up study

Authors:

Inger Hallberg (inger.hallberg@liu.se)
Margareta Bachrach-Lindström (margareta.bachrach-lindstrom@liu.se)
Staffan Hammerby (staffan.hammerby@lio.se)
Göran Toss (goran.toss@lio.se)
Anna-Christina Ek (anna-christina.ek@liu.se)

Version: 2 Date: 5 June 2009

Author's response to reviews: see over
Authors’ response to reviews

Title: Health-related quality of life after vertebral or hip fracture: a seven-year follow-up study
Manuscript: 1509314422236778

Authors:
Inger Hallberg* - inger.hallberg@liu.se
Margareta Bachrach-Lindström - margareta.bachrach-lindstrom@liu.se
Staffan Hammerby - staffan.hammerby@lio.se
Göran Toss - goran.toss@lio.se
Anna-Christina Ek - anna-christina.ek@liu.se

Version: 2 Date: 5 Jun 2009
Authors’ response to reviews: see following
Referee 1

Major Compulsory Revisions
1. As Figure 1 demonstrates, there were 95 study participants with a vertebral or hip fracture at baseline, 91 study participants at the two-year follow-up, and 67 at the seven-year follow-up, which corresponds to a loss to follow-up of 29.5% of the original study cohort. The authors need to carefully consider the mechanism by which drop-out occurred and provide comparisons of differences between study completers and drop-outs. They should also consider the mechanisms by which missing data may occur, as described by Little and Rubin (2002): missing completely at random, missing at random, and missing not at random. If the missing data pattern in the current study is missing not at random, then the analyses of HRQOL change will be biased.

Thank you for this comment. Missing data in the current study are unfortunately missing not at random, and this is a limitation that may have affected the results. In this study, the analyses included only participants who had completed the seven-year follow-up.

We have now made an analysis regarding SF-36 data at two-year follow-up between those missing (n=24) and those included (n=67) at seven-year follow-up. The missing group had significantly lower values regarding the SF-36: general health and social function as well as lower weight, body mass index and bone mineral density. Age did not differ. New text has been added to Patient Group in Methods; see page 7.

This can be interpreted as the missing group having poorer health and the results at seven-year possibly leading to an overestimation of HRQOL (SF-36) scores for hip and vertebral participants. New text has been added to Statistical Analyses on page 11 and to the discussion of limitations on page 22 in Discussion.

2. The authors used repeated-measures ANOVA to test for change over time, but only for those individuals who participated in all three waves of data collection. There are a number of limitations associated with this approach to the analysis. First, this analysis makes stringent assumptions about the covariance structure of the repeated measurements, which are unlikely to be satisfied in many applications. The model cannot accommodate time-varying covariates, which may be associated with HRQOL and should be controlled for in the analysis of change over time. A random-effects regression models is a flexible method for the analysis of data characterized by missing values, increasing variability across
measurement occasions, and covariates that change with time. Furthermore, the analysis should not be limited just to those individuals who completed all three waves of data collection; a random-effects model can accommodate missingness.

Based on comments from Referee 3, the focus in aim (i) has been changed to concentrate on the change in HRQOL from year two to year seven. We have already published data comparing baseline with year two (ref 29). More information about these results has been added to the introduction on page 5.

The differences between two and seven years were tested using Student’s paired t-test (instead of ANOVA). These results showed that the vertebral group had stable values in all domains except bodily pain, which had decreased significantly (49.7 vs. 41.6 p=0.022 CI (95%) =1.2-14.9), indicating increased pain. The hip group had stable values in all domains (see Statistical Analyses, page 12, and Results, page 14).

From the reference group, women aged 70 and 75 year were compared and showed stable values in all domains, except for vitality, which was significantly lower for the elderly women (59.0 vs. 69.7). Text has been added to Reference Group, page 7.

3. The use of Pearson correlation coefficients (see Table 4) assumes a normal distribution of responses and a linear association between variables. How did the authors evaluate these assumptions? Isn’t comorbidity measured using a dichotomous scale? If so, then a Pearson correlation, which assumes an interval scale, is not an appropriate choice.

Thank you for the remark. We analysed our data using a partial correlation between two variables instead. We have added a new table (5), which shows the partial correlations in which the relationship between two variables is measured, controlling for the effect that age, fracture group, new co-morbidity and new fracture since two-year follow-up have on both variables. See Statistical Analyses, page 12, and Results, page 16.

4. The authors conducted multiple linear regression analyses to identify predictors of each of eight HRQOL domains. The investigated predictors were balance, hand grip, spinal deformity index, age, bone mineral density, physical activity, new comorbidity, and fall frequency. These analyses are a form of data mining, in which the authors are searching through the data to identify statistically significant predictors. Unless the sample size is large, this approach to modeling may result in overfitting and a failure of the model to validate on a new set of study participants. I recommend that the authors make substantial changes to this section by: (a) focusing on a smaller set of explanatory variables that are justified based on prior theory/research, and (b) validating their analysis results. Given that sample size is small, a cross-validation analysis cannot be conducted. Therefore, the authors should investigate the use of an empirical re-sampling technique, such as the bootstrap, to conduct the validation.
Thank you for this comment. We have deleted the multiple linear regression analyses according to aim (iii), and instead use the partial correlation to interpret the relationships, and it is now more specific. (See Q 3!)

5. The authors used a one-way ANOVA to test for differences in the adjusted means for the vertebral and hip fracture groups. The utility of this analysis for inferring differences in the populations from which these samples were drawn is limited because the authors have not controlled for covariates such as co-morbidity and recent new fracture, which may differ between the two groups and may be confounded with the dependent variable.

According to aim (ii): regarding the differences between hip and vertebral groups, we instead used ANCOVA, controlling for the effect of the covariates age, new co-morbidity and new fracture (see Statistical Analyses, page 12, and Results, page 15).

A new table (4) has been added.

According to aim (ii): regarding the differences between reference and fracture groups, we used an unpaired \( t \)-test because we unfortunately do not have data on co-morbidity or fracture status (see Statistical Analyses, page 12, and Results, page 15).
A new table (3) has been added.
The reference group was handled like a normal background population, recruited from the same general population area during 2006 (added to the discussion of limitations on page 22 in Discussion). New data have been added.

6. The Bonferroni method, which was adopted to control the familywise Type I error rate to \( # \), is not the optimal approach when the tests that comprise the family are correlated. It will have lower power than alternate procedures that account for this correlation. Further information can be found in Blair et al. (1996).
This is correct and we agree, but in this case we had performed a different analysis. Thank you for the reference to Blair et al.!

7. The analysis of longitudinal HRQL is challenged by the potential for study participants to experience response shift over time. Response shift is defined as “a change in the meaning of one’s self-evaluation of a target construct as a result of: (a) change in the respondent’s internal standards of measurement (i.e., scale recalibration), (b) change in the respondent’s valuation of component domains constituting the target construct, or (c) redefinition of the target construct (i.e., reconceptualization)” (Schwarz & Sprangers, 1999). The authors should discuss the implications of response shift for their interpretation of study results.

Thank you for this comment; this is important. It is addressed in Discussion, addressing methodological issues; see page 21–22.
A response shift cannot be excluded in this study. The HRQOL response patterns could have been affected by choice of comparator group.
A study conducted by Fayers et al. found that the majority of questionnaire participants reported using reference frame comparisons when completing an annual
HRQOL questionnaire. The reference to Schwarz & Sprangers, 1999 was added (ref 58).

The following reference was also added (ref 59):

Minor Essential Revisions
8. The first section of the Results, which describes patient characteristics, should be shortened substantially. It repeats much of the information already reported in Table 1. For example, in the sentence that begins “At baseline, 24/42 (57%) in the vertebral group…” the frequencies do not need to be reported in addition to the percentages.

As you have suggested, the first section of the results has been shortened and the information about baseline and two-year follow up has been deleted because the novel aspect of this study is the change in HRQOL from two to seven years (Table 1 has also been changed).

9. The interpretation of the p-values in Tables 2 and 3 is not clear to me. Does the p-value correspond to a test of the omnibus, or overall hypothesis? If so, this should be stated in the table note.

This is no longer an issue because the statistical methods and aim (i) have been amended.

The exact p-value in the table was related to the ANOVA model, and the elevated small letter (for example $p=0.01^a$) refers to significant difference between groups. This table (2) has been changed due to the changed aim (i), as has Table 3.

Discretionary Revisions
10. In the enumeration of study strengths and limitations, the authors should discuss their rationale for not including an osteoporosis-specific measure of HRQOL, such as the Osteoporosis Quality of Life Measure (Badia et al., 1997) in the data collection for the vertebral and hip fracture groups.

As suggested, this has been added to the discussion of limitations on page 22 in *Discussion*.

*Thank you for all the valuable comments and references!*
Referee 2

Major revisions
An interesting long-term study of the effects of vertebral or hip fracture on HRQOL
However one of the problems of such a study are downstream fractures and mortality. How were subsequent fractures handled? One could include patients with subsequent fracture but then one might want to stratify patients with and without subsequent fracture. Significant mortality was observed. How was this handled? Was it ignored. One could model death as HRQOL of zero. Is the loss of QOL with vert fracture due to the single vert fx or a downward spiral as some authors have suggested due to subsequent fractures?
This article is unique in that both seven year data and physical performance measures are shown.

How were subsequent fractures handled?

According to aim (i):
Based on comments from Referee 3, the focus in aim (i) has been changed to concentrate on the change in HRQOL from year two to year seven. We have already published data comparing baseline with year two (ref 29). More information about these results has been added to the introduction on page 5.

The differences between two and seven years were tested using Student’s paired \( t \)-test (instead of ANOVA). These results showed that the vertebral group had stable values in all domains except bodily pain, which had decreased significantly (49.7 vs. 41.6 p=0.022 CI (95%) =1.2-14.9), indicating increased pain. The hip group had stable values in all domains (see Statistical Analyses, page 12, and Results, page 15).

Student’s paired \( t \)-test for the whole group (n=67) showed, when we regrouped the data into new fracture since two-year follow-up (n=29) and no fracture since two-year follow-up (n= 38), the following results: The group with new fracture had significantly lower values at year 7 compared to year 2 regarding Role-Physical, Bodily Pain, General Health and Social Function. The group with no new fracture had stable values (see Statistical Analyses, page 12, and Results, page 14).

According to aim (ii):
Regarding the differences between hip and vertebral groups, we instead used ANCOVA, controlling for the effect of the covariates age, new co-morbidity and new fracture (see Statistical Analyses, page 12, and Results, page 15).
A new table (4) has been added.

Regarding the differences between reference and fracture groups we used unpaired \( t \)-test because we unfortunately do not have data on co-morbidity or fracture status (see Statistical Analyses, page 12, and Results, page 15).
A new table (3) has been added.
The reference group was handled like a normal background population, recruited from the same general population area during 2006 (added to the discussion of limitations on page 22 in Discussion). New data have been added.
According to aim (iii):
We have deleted the multiple linear regression analyses and instead use the partial correlation to interpret the association according to aim (iii).
We analysed our data using a partial correlation between two variables instead. We have added a new table (5) that shows the partial correlations, in which the relationship between two variables is measured controlling for the effect that age, fracture group, new co-morbidity and new fracture since two-year follow-up have on both variables.
See Statistical Analyses, page 12, and Results, page 16, and new Table 5.

Significant mortality was observed. How was this handled?

Thank you for this comment. Missing data in the current study are unfortunately missing not at random, and this is a limitation that may have affected the results. In this study, the analyses included only participants who had completed the seven-year follow-up.
We have now made an analysis regarding SF-36 data at two-year follow-up between those missing (n=24) and those included (n=67) at seven-year follow-up. The missing group had significantly lower values regarding the SF-36: general health and social function as well as lower weight, body mass index and bone mineral density. Age did not differ. New text has been added to Patient Group in Methods; see page 6.
This can be interpreted as the missing group having poorer health and the results at seven-year possibly leading to an overestimation of HRQOL (SF-36) scores for hip and vertebral participants. New text has been added to Statistical Analyses on page 11 and to the discussion of limitations on page 22 in Discussion.

Is the loss of QOL with vert fracture due to the single vert fx or a downward spiral as some authors have suggested due to subsequent fractures?

The small sample in the study does not allow for this question to be evaluated, but Student’s paired t-test for the whole group (n=67) showed, when we regrouped the data into new fracture since two-year follow-up (n=29) (of whom 22 belonged to vertebral group) and no fracture since two-year follow-up (n= 38) (of whom 20 belonged to the vertebral group), the following results: The group with new fracture had significantly lower values at year 7 compared to year 2 regarding Role-Physical, Bodily Pain, General Health and Social Function. The group with no new fracture had stable values. New text has been added to Statistical Analyses, page 12, and Results, page 14.
This can be interpreted as the vertebral and hip fracture group with subsequent fractures (all fractures, not only new vertebral ones) having poorer health (measured with the SF-36) at seven-year follow-up (added to Discussion, page 17).

We need more information about the reference group. Was it matched by
comorbidities, age, etc. It should be. Please show the demographics of this group in table 1.

Regarding reference group: We unfortunately do not have data on co-morbidity or fracture status. The reference group was handled like a normal background population, recruited from the same general population area during 2006 (added to the discussion of limitations on page 22 in Discussion). New data has been added to Reference Group, page 7.

Minor revisions
reference tables and figures in results

Some revisions have been made in Tables 1-5.

Thank you for all the valuable comments!
Referee 3

Major Compulsory Revisions:

1. When was the reference group selected? There may be baseline differences between the groups if the selection of the reference group occurred 7 years after the hip and vertebral fracture group. The management of the osteoporosis, medications given, and lifestyle factors may be quite different for these two time periods. In addition, we do not know the prevalence of fractures in the reference group. As a consequence, the comparison between the reference group and fractures groups may be misleading and the conclusions drawn may be false. Do the authors have any evidence that the groups are similar for important baseline characteristics?

   Regarding the reference group: We unfortunately do not have data on co-morbidity or fracture status. The reference group was handled like a normal background population, recruited from the same general population area during 2006 (added to the discussion of limitations on page 22 in Discussion). New data has been added to Reference Group, page 7.

   According to aim (ii): regarding the differences between reference and fracture groups, we used an unpaired t-test because we unfortunately do not have data on co-morbidity or fracture status (see Statistical Analyses, page 12, and Results, page 15).
   A new table (3) has been added.

2. The novel aspect of this study is the change in HRQL from year 2 to year 7. However, the results (from table 2) may be difficult to explain. It appears that HRQL decrease from year 2 to year 7 which is counter-intuitive. One would expect that HRQL would gradually increase over this time period. Is there any plausible explanation for the observation? Is this problem due to drop-outs or new fractures? Are these comparisons reliable given these findings?

   Thank you for this suggestion. The focus in aim (i) has been changed to concentrate on the change in HRQOL from year two to year seven. We have already published data comparing baseline with year two (ref 29). More information about these results has been added to the introduction on page 5.

   According to aim (i):

   The differences between two and seven years were tested using Student´s paired t-test.
   These results showed that the vertebral group had stable values in all domains except bodily pain, which had decreased significantly (49.7 vs. 41.6 p=0.022 CI (95%) =1.2-14.9), indicating increased pain. The hip group had stable values in all domains (see Statistical Analyses, page 12, and Results, page 14).
   A new table (2) has been added.
Student’s paired $t$-test for the whole group ($n=67$) showed, when we regrouped the data into new fracture since two-year follow-up ($n=29$) and no fracture since two-year follow-up ($n=38$), the following results: The group with new fracture had significantly lower values in Role-Physical, Bodily Pain, General Health and Social Function. The group with no new fracture had stable values. The groups with new co-morbidity ($n=46$) had stable values compared with the group without new co-morbidity ($n=21$).

This can be interpreted as the vertebral and hip fracture group with subsequent fractures (all fractures, not only new vertebral ones) having poorer health (measured with the SF-36) at seven-year follow-up. No value had significantly improved at group level in any domain.

**Regarding dropouts:**
Missing data in the current study are unfortunately missing not at random, and this is a limitation that may have affected the results. In this study, the analyses included only participants who had completed the seven-year follow-up. We have now made an analysis regarding SF-36 data at two-year follow-up between those missing ($n=24$) and those included ($n=67$) at seven-year follow-up. The missing group had significantly lower values regarding the SF-36: general health and social function as well as lower weight, body mass index and bone mineral density. Age did not differ. New text has been added to Patient Group in Methods; see page 7.

This can be interpreted as the missing group having poorer health and the results at seven-year possibly leading to an overestimation of HRQOL (SF-36) scores for hip and vertebral participants. New text has been added to Statistical Analyses on page 11 and to the discussion of limitations on page 22 in Discussion.

3. Comparisons between vertebral and hip fracture groups are interesting. However, these results need to be adjusted for other variables (i.e. other disease conditions). This would allow for better interpretation of the results.

**According to aim (ii):**
Regarding the differences between hip and vertebral groups, we instead used ANCOVA, controlling for the effect of the covariates age, new co-morbidity and new fracture (see Statistical Analyses, page 12, and Results, page 15). A new table (4) has been added.

4. While there may be some statistically significant differences between groups the authors should state in the manuscript whether these differences are clinically meaningful. Significant difference between groups or over the duration of the study may be a function of the sample size. It is extremely important to state that the authors believe that these differences or clinically meaningful.

Thank you for valuable comments; this is important. Our results show, for the most part, clinically meaningful differences. A half standard deviation ($½$ SD) is a conservative estimate of clinical significance, but the minimally important difference may be below $½$ SD in specific cases. New text has added to Discussion, page 23.
The following reference has been added (ref 57):
*Applying Quality-of-Life Data Formally and Systematically Into Clinical Practice.*
Mayo Clinic Proceedings Mayo Clinic Proceedings J1 - Mayo Clinic Proceedings,

5. How were new fractures adjusted for in the analyses? These new fracture may explain differences between groups and over time in HRQL. The authors may want to run the analyses for individuals that did not have new fractures.

Please see Q 2, concerning aim (i), and Q 3, concerning aim (ii).

6. Table 4. The results are based on bivariate analysis so they may be misleading. The authors need to adjust for other important variables. In general, I would suggest that the authors focus on the novel aspects of the findings and attempt to adjust for important covariates in all analyses. It is very important that the authors explain way HRQL decreases from year 2 to year 7. There may be some factor that causes this change other than the fracture groups.

**According to aim (iii):**
We have deleted the multiple linear regression analyses and instead use the partial correlation to interpret the association according to aim (iii). We analysed our data using a partial correlation between two variables instead. We have added a new table (5) that shows the partial correlations, in which the relationship between two variables is measured controlling for the effect that age, fracture group, new co-morbidity and new fracture since two-year follow-up have on both variables. See *Statistical Analyses*, page 12, and *Results*, page 16, and new Table 5.

See Q 2, according to aim (i). This can be interpreted as the vertebral and hip fracture group with subsequent fractures (all fractures, not only new vertebral ones) having poorer health (measured with the SF-36) at seven-year follow-up.

*Thank you for all the valuable comments!*