Reviewer’s report

Title: Use of screw locking elements improves radiological and biomechanical results of femoral osteotomies

Version: 2  Date: 8 September 2014

Reviewer: Andrew Judge

Reviewer’s report:

Concern 1 of reviewer 2
My major concern is the statistical pooling of the groups. The authors pool the results of the torsional stiffness of the DCP+6S+SLE and DCP+8S+SLE versus DCP+6S and DCP+8S on the other hand. Plus the 6S groups were sacrificed after 8 weeks and the 8S groups sacrificed after 16 weeks. I do not agree with this procedure and I am convinced that this is methodological not correct. The authors compare apple and oranges and the reason for this procedure is the low power of n=4 per group and p= 0,068 (Wilcoxon Test) for all four groups.

Comments by AJ
In relation to the sample size there are 16 sheep with a femoral osteotomy performed. Half had a screw locking element SLE and dynamic compression plate (DCP), the other half no SLE and a DCP only. 16 sheep appear to have data for analysis at week 8 (6 and 8 screw DCP), and 8 sheep for week 16 (6 screw DCP only).

So there are 4 groups of sample size 4: DCP+6S+SLE, DCP+6S+NOSLE, DCP+8S+SLE, DCP+8S+NOSLE

Pooled comparison can be DCP 6 screws versus 8 screws.
Or SLE versus no SLE.

In table 1 there is no pooling of data, comparisons are done separately in 6 and 8 screw samples of SLE versus no SLE.

Table 2 shows results for torsional stiffness, in operated versus non-operated bone. As the comparison is within the same sheep, there is some matching to the data, so a paired Wilcoxon test should be used. The data is pooled here, and this may reflect use of a non-parametric test with small sample sizes.

Could also provide in table 2 data within the 4 groups, and use a parametric paired t-test. No reason why not to?

Concern 2 of reviewer 2
It is correct to assume that the data are non-parametric since only 4 samples are evaluated by the authors. Nevertheless, the differences between the groups DCP+6S vs. DCP+6S+SLE and DCP+8S vs. DCP+8S+SLE are significant with a
Comments by AJ

See BMJ paper by Altman “Analysis of continuous data from small samples”. http://www.bmj.com/content/338/bmj.a3166. In the case of data with small sample sizes, as in this study, it can be inappropriate to use non-parametric tests. There needs to be some evidence from the existing literature on what the expected distribution of the data is. If larger sample sizes from existing studies suggest the data is normally distributed, then it is perfectly ok, to use parametric tests with small sample sizes.

The t-test used in table 1 seems fine. For table 2, why not just used a paired t-test? As per Altman table with sample size of 4 in each group, a non-parametric test cannot by definition reach statistical significance, which is why data was pooled. I would used a normal paired t-test, which should work with sample size of 4 group comparison.