

Interactive comment on “Near-future levels of ocean acidification do not affect sperm motility and fertilization kinetics in the oyster *Crassostrea gigas*” by J. N. Havenhand and P. Schlegel

G. Quinn (Referee)

gerry.quinn@deakin.edu.au

Received and published: 7 June 2009

This is an interesting paper on the effects of ocean acidification on sperm motility and fertilization success of oysters. It provides important comparative data for this species of oyster with previous work from Japan. The work revealed an absence of strong effects of reduced pH on sperm and fertilization success in this oyster species. Often, statistically non-significant results are difficult to publish, but this paper incorporated statistical power analysis to help interpret the non-significant results, giving the authors much more confidence that they really were dealing with no effects of pH.

I have a few minor comments on the design of the experiments. While the comparison

C610

of two treatment groups (FSW versus acidified FSW) is straightforward, there were some details lacking:

1. I note that no details are provided as to how the oysters were stored. Were there multiple aquaria or single aquaria for each pH?
2. Were all oysters kept at their experimental pH for the same length of time before being strip-spawned and sperm motility measured?
3. It is not clear what the replicate units were for the fertilization success experiment, i.e. for each combination of sperm and eggs, was there a single “mixture”?

I also have some specific comments to make on the statistical analyses.

1. The P values presented in Table 1 for each male were presumably based on $n=10$ for each pH using the means of the 10 replicate observations. I think including the degrees of freedom for the t-tests in the legend for Table 1 would have been useful.
2. The independent t-tests for each male presented in Table 1 seem appropriate, even though the sperm preparations used for each pH came from the same male. The sperm samples themselves are independent between pHs, justifying an independent test.
3. The two factor mixed model ANOVA that tested for differences between pH and between males is essentially a randomized block analysis using individual male oysters as blocks. This analysis would have used the data from the speed FSW and speed acFSW columns in Table 1 where the number of males is 16 (matches the 15 df for the F-test of pH which uses the pH by male interaction term as the denominator). However, line 8 on page 4578 states there were 14 males? Note that this analysis cannot test for any interaction between pH and individual males (cannot test whether individuals responded consistently to changed pH).
4. The analysis for the fertilization success data should be identical to that for the sperm motility data. I presume the absence of t-tests from Table is because there were no “replicate units” for measuring fertilization success, i.e. no replicate batches? Table

C611

2 includes 17 trials yet the ANOVA tests for pH are based on 12 df (lines 1-2 on page 4580) – obviously trials A-D were not used in the ANOVA but why not?

One of the most important aspects of this paper is the use of power analyses to interpret non-significant results. While power analysis is most useful when used a priori to set sample sizes, this paper illustrates its value when used in a post-hoc fashion. The authors present, on pages 4579 (lines 15-16) and 4580 (lines 2-3), the sizes of differences that could have been detected given the variability between replicates and sample sizes used. These are called “minimum detectable effect sizes (MDES)” and presenting such information is important when trying to interpret statistically non-significant results. In this paper, the analyses could have detected quite small differences with reasonable confidence (80%). Obviously, the interpretation of a MDES still relies on biological judgment as to whether a 4.5% change in FSR_{max} for example is a biologically important difference. It is important to note that with the mixed model analyses used here, the calculation of the “within-group variation” used in the power analysis must be based on the denominator term used for the F-ratio test for pH.

Minor issues:

Page 4574, line 15: “accurate” not right word – “meaningful” probably better

Page 4575, line 12: “life-stages” rather than “life-stage”

Interactive comment on Biogeosciences Discuss., 6, 4573, 2009.