The following debate article was sent to Oikos but rejected for publication following the retraction of the Moller and de Lope (1998) paper.

Fluctuating asymmetry in stone oaks as a response to man-made herbivory, fact or fabrication?

By Jørgen Rabøl and Jette Andersen,

Department of Population Biology,
University of Copenhagen,
15 Universitetsparken,
DK-2100 Copenhagen Ø,
Denmark
e-mail: jrabol@zi.ku.dk; jeta@worldonline.dk

In a recent paper, Møller & de Lope (1998, Oikos 82, 246-252) report that increased levels of fluctuating asymmetry arose in stone oak leaves (*Quercus rotundifolia*) as a response to man-made herbivory (Table 2) and survival (Table 3). In the article it is maintained that the data was measured by J.Andersen, for which she was explicitly and profusely thanked by Møller & de Lope, both in the Acknowledgements: "*J.Andersen did the heroic task of measuring all the leaves*" and again on p.247 "*The leaves were measured - - - by J.Andersen, who was unaware of the origin of the leaves. Measurements were therefore done blindly with respect to the predictions under test*".
When reading the paper, we were surprised by the fact that it contained a number of improbabilities. We repeatedly approached the senior author (A.P. Møller) for the data on which the published results were based, but despite a prolonged e-mail correspondence, we never received any answer to our specific requests. As a consequence, an identical copy of A.P.Møller's original data file (the Andersen file) is used for the calculations in this paper. This file remained unchanged and unmodified in a computer at the Department of Population Biology at the University of Copenhagen from April 1996. At that time, the data file was given to A.P.Møller as a computer copy and on paper, after he resigned from his position as professor of the department. This data was submitted to Møller (along with those leaves that had survived the measurements) exactly as now published on a web site. The file is available on www.zi.ku.dk/terrecol/steneg.htm and contains information on leaves sampled from 343 oak trees.

Improbabilities in the paper

What is designated leaf "area" in Tables 1-3 cannot be leaf area, simply because the mean area of a stone oak leaf cannot be about 20mm^2, when the mean leaf width of the same leaves is approx.

12mm. According to Møller & de Lope ”*Leaf area was determined automatically by a leaf area meter to the nearest 0.01 cm^2*”.

According to the Andersen data file, mm lengths were measured. However, no area measurements are found in the file that also contains, not only the lengths and the largest widths either side of the leaves, but also the numbers and individual lengths of thorns on either side, and the distances between the three middle ones. A length of about 20mm may be reasonable for a leaf of a stone oak, so the rather detailed specification of “area” in the paper by Møller & de Lope is incomprehensible.

2) The mean values in Table 1 of four “old” leaf characters are almost identical to the mean values of the same four characters in another sample of “new” leaves in Table 2 (controls).

This is conspicuous - and unexpected, as the two samples are completely independent. The incredulity increases when we find that, for at least the first three characters out of the four, the standard deviations of the "old" leaves compared with the new ones are higher by almost the same constant (approx. 1.8). Thus we get the nearly impossible combination of four pairs of very similar means, whose corresponding variances all deviate significantly from each other (F-test, $P < 0.01$). In Table A, we summarize and transform (standard errors into standard deviations) parts of Tables 1 and 2 of Møller & de Lope. In order to avoid further confusion, we use Møller’s & de Lope’s designation ”area” for what is most probably the lengths in the Andersen file. Apparently the same data set has been used twice, but probably not with the described combination of sample sizes, 343 or 25. As can easily be confirmed, the standard deviations – in at least these first three characters – would be almost identical for combinations of sample sizes, 105 and 25, or 343 and 81. When compared with the standard deviations in Table B (see below), 105 and 25 is the more likely of the latter two constellations.

3) The third inconsistency is not quite so obvious and by itself without much significance but indirectly it must cast doubt on the credibility of the samples involved. If it can be made probable that some of the table values are constructions, then none of the rest can be trusted.

In Tables 2 and 3, mean width is given with one decimal, mean absolute asymmetry with two (Table 2) or one (Table 1 and 3), and relative asymmetry is given with three. In particular, the single decimal is puzzling as all widths were measured to the nearest 0.01mm with a digital slide gauge, and some of the sample sizes are as large as 343 trees.

Considering Tables 2 and 3, the five mean values of relative asymmetry are to the third decimal identical to the mean values of absolute asymmetry (denoted by the two decimals in Table 2 and the one in Table 3) when these are divided by their corresponding mean "widths" (one decimal). For example, 1.51/12.1 = 0.125 (Table 2(2) in Table A). This is a clear indication that Møller & de Lope calculated the means of relative asymmetry directly on the basis of the rounded off table values of mean absolute asymmetry and mean width which is neither correct nor as specified by Møller & de Lope (legend for Table 1). The correct way to calculate a mean relative asymmetry is to do it on the basis of the values of relative asymmetry of each of the N trees in the sample.
When calculated correctly, mean relative asymmetry almost always attains a higher value than that of mean absolute asymmetry divided by the mean width (cf. Table B, and also the Table 1 samples of Møller & de Lope, Table A). Furthermore, as the table means obviously are rounded off, Møller & de Lope must have calculated mean relative asymmetry (3 decimals) directly on the basis of the table values of mean absolute asymmetry (1 or 2 decimals) divided by mean width (1 decimal). If calculated on basis of the original mean values (more decimals) the combined probability of all five means of relative asymmetry being identical to those given in Tables 2 and 3 is approximately 0.003.
How did the standard errors of relative asymmetries arise, if Møller & de Lope calculated mean relative asymmetry on the basis of mean absolute asymmetries and mean widths. The point is that if the mean values and standard errors of relative asymmetry are constructions how then can we trust the other values in Tables 2 and 3. This question is not weakened by the fact that the leaves behind Tables 2 and 3 are not to be found in the original Andersen data file.

Calculations from the Andersen file

The Andersen file contains data on leaves sampled from 343 oak trees, which coincides with the sample size in Table 1 of Møller & de Lope. In the file, the number of trees contributing old and, in particular, new leaves is slightly smaller than 343 (see Table B). The data is subdivided into three treatments (called 1, 2, and 3) and old and new leaves (designated 1 for one year old and 0 for zero years). However, the samples presented in Table 1 in Møller & de Lope are not subdivided into treatments, and the authors even write "*Descriptive statistics are reported for untreated trees in Table 1*". The coincidence in number of trees (343) points to the original Andersen data file as being used at least partly for Table 1 in the paper by Møller & de Lope.

In contrast, Table 2 by Møller & de Lope (see Table A) gives the results of three treatments, but with sample sizes of only 25 trees per treatment (in the paper, Treatment (1) is designated complete herbivory, (2) partial herbivory, and (3) control). These three samples of each 25 trees do not occur in the Andersen file, nor do the two samples of each 22 trees in Table 3.

Møller & de Lope might claim that the original measurements were inaccurate and therefore useless. However, on p.247-48, they write: ”*Measurement errors were estimated by measuring 50 leaves on one day and then re-measuring the same leaves the following week using the procedure of Yeserinac et al. (1992). The percentage measurement errors were 0.5 and 0.7% for leaf area and leaf width. While measurement error was 1.2% for left-minus-right leaf width. All these measurements were sufficiently small to allow reliable determination of the variables of interest*”. As “*Andersen…measured all the leaves”* according to Møller & de Lope, she obviously did it well.

The means and standard deviations based on the measurements in the Andersen file (Table B) and the corresponding values in Table 1 of Møller & de Lope (see Table A) are most often very different. But even very inaccurate original measurements can hardly explain why the old leaves are so much bigger than the new leaves (Table B, both "lengths" and “widths” in all three treatments, $P < 0.05$, t-test). The new leaves appear to have been collected before they were fully developed. This is an interesting point, because in Table 1 of Møller & de Lope (see TableA), there are no significant differences between the "areas" of "old" and "new" leaves.

If treatments 1, 2 and 3 designate complete, partial and no herbivory, respectively, then according to the Andersen data file, no significant effect exists of herbivory on ”relative asymmetry” (Table B, Treatment 1 against Treatment 3 (t-test, $0.20<P<0.30$) or Treatment (1+2) against Treatment 3 ($0.10<P<0.20$)). The following circumstantial evidence supports this conclusion:

a) If the old leaves are important for the growth of new leaves, one might expect smaller new leaves, when many old leaves are removed before sprouting. According to Table B, the leaves are smallest in Treatment 1 (complete man-made herbivory).
b) As is also seen in Table B, approx. 97 (Treatment 3, no herbivory) trees contribute with both old and new leaves; 10 further trees contribute with only old ones. This may be an indication that 10 control trees did not grow any new leaves. If Treatments 1 and 2 designate complete and partial herbivory, then we might expect that, compared with the controls, fewer trees

would grow new leaves (because they died or were otherwise retarded as a result of the treatment). Approx. 234 Treatment 1 and 2 trees contribute with old leaves, but only 178 with new ones, too. The difference between Treatment 3 and Treatment 1+2 is highly significant (P about 0.003, Chi-square test).

c) As already mentioned, the sample size of 343 trees results in too large standard deviations (Table 1, Møller & de Lope, see Table A), whereas a sample size of approx. 100 trees is much more reasonable and found in treatment 3 in Table B.

In conclusion, there are many inconsistencies that must be clarified by Møller & de Lope. The file on which the tables are calculated, Møller & de Lope did not use the Andersen data file; the area measurements; the almost identical mean values with significantly different variances; the calculation of mean and standard error of relative asymmetries in Tables 2 and 3; the designations: Treatments 1, 2 and 3 in the Andersen file. Therefore, it would be appreciated if Møller & de Lope make their data file available to other researchers.

New leaves	area (sd)	width (sd)	abs.as. (sd)	rel.as. (sd)
Tab. 1, N=343	20.43(12.41)	10.3 (7.4)	1.0 (1.9)	0.102 (0.13)
Tab.2 (1) N=25	20.68 (5.60)	13.0 (4.5)	2.08 (1.05)	0.160 (0.055)
Tab 2(2) N=25	20.77 (5.50)	12.1 (3.5)	1.51 (1.10)	0.125 (0.050)
Tab 2(3) N=25	20.58 (5.75)	12.7 (4.0)	1.01 (1.05)	0.080 (0.070)
Old leaves				
Tab 1 N=343	20.59 (10.37)	12.7 (7.4)	1.0 (1.9)	0.081 (0.11)

Table A. Based on Tables 1 and 2 of Møller & de Lope (standard deviations calculated on the basis of N and standard errors given). Treatments (1), (2) and (3) designate complete herbivory, partial herbivory and control, respectively. The almost identical means and the very different standard deviations of "new" leaves in Table 2 (3) and "old" leaves in Table 1 are conspicuous.

New leaves	length (sd)	width (sd)	abs.as. (sd)	rel.as. (sd)
1 N = 93(93)	14.94 (6.21)	11.55 (4.91)	0.58 (0.29)	0.055 (0.025)
2 N = 85(83)	17.31 (5.61)	13.06 (4.44)	0.66 (0.27)	0.057 (0.030)
3 N = 97(97)	16.76 (4.64)	12.57 (4.24)	0.61 (0.31)	0.051 (0.026)
Total N = 275(273)	16.32 (5.58)	12.37 (4.56)	0.61 (0.29)	0.054 (0.027)
Old leaves				
1 N = 115(115)	19.54 (4.86)	15.26 (4.11)	0.64 (0.26)	0.044 (0.016)
2 N = 119(119)	19.06 (4.71)	14.39 (3.72)	0.66 (0.29)	0.048 (0.019)
3 N = 107(106)	20.45 (4.59)	15.64 (3.89)	0.66 (0.29)	0.042 (0.016)
Total N = 341(340)	19.66 (4.75)	15.06 (3.93)	0.65 (0.28)	0.045 (0.017)

Table B. Means and standard deviations (in mm except rel. as.) based on the Andersen data file. The sample size N refers to number of trees in which length of the leaves were measured. The numbers in parentheses refer to the sample sizes of the last three of the four characters. The total number of trees measured was 343 (or 344, including one tree of which only one leaf length was measured). In approx. 70 trees only "old" leaves were measured, and in a few trees only "new" leaves. Concerning the four characters, “length” is directly available in the Andersen file, whereas “width” is left plus right width. “Abs. as” is the unsigned left minus right width, and “rel. as.” is “abs. as.” divided by “width”. Each tree contributes with a single mean value of an “old” and/or a “new” leaf. The mean denoted is thus the (second order) mean of these means (with corresponding standard deviations). Six obvious typing errors in the data file are corrected: 1) Treatment 1.34 age 1 ("old"). The left width of a single leaf is denoted as 811, not 8.11. If not corrected - or omitted - this error has serious influence on, in particular, the mean of absolute asymmetry and the standard deviations of width and absolute asymmetry. Including the error leads to 15.68 (11.63) mm and 1.24 (10.92) mm for mean (sd) of total N = 340 ("old" leaves) width and absolute asymmetry, respectively. 2) Treatment 1.10 age 1. Right width of a single leaf is denoted 10:28 - and corrected to 10.28. 3) Treatment 2.37 age 0. The age of a single leaf is not denoted, but obviously it is 0 ("new"). 4) Treatment 3.93 age 0. The left widths of two leaves are denoted as 0. These leaves are omitted before further calculations. 5) Treatment 2.111 age 1. The left width is denoted as 3..75 – and corrected to 3.75. 6) Treatment 3.17 age 1. The right width is denoted as 8..36 and corrected to 8.36.Including the errors of 2) through 6) has no significant influences on the means (sds) in the table above.

Malmer

Copy of part of a letter sent from the Editor-in-Chief, Nils Malmer OIKOS to APM. The space between the first two sections in the letter signals a section on about seven lines taken out before the present part of the letter was sent to Jørgen Rabøl from OIKOS. Perhaps also other sections were taken out.

Lund. 24 November, 2000.

Professor Anders Pape Møller
Laboratoire d Ècologie
Universite Pierre et Marie Curie
Paris

I have received your email dated 2000-11-09 dealing with the very serious problems concerning your paper together with de Lope published in Oikos 82: 246ff (1998). I am anxious to finalise the case as soon as possible but, unfortunately, the content in your mail does not give much help. I have also a submitted paper pending for decision whether it should be accepted or rejected depending on the outcome of our discussions.

In the Editorial Office we have now the two sets of data which you have admitted refer to measurements carried out on the material of leafs dealt with in your paper. One, here called data set A, has been submitted by J. Rabøl and J. Andersen; the other one, here called data set B, has been submitted by yourself. The measurements included in data set B were done after you had seen the measurements in data set A. There are several differences between the two data sets.

In your printed paper you explicitly say that you have only used data set A for the calculations and results presented there. Later you have in letters said that you discarded data set A and instead only used data set B for the paper. In data set A there are no measurements of leaf area but instead of leaf length. In the published version no values are given for leaf lengths. For leaf area are indicated values that are unrealistic and you have explained that the decimal points have been misplaced.

For every article published in a scientific journal it is always expected that it should be possible to use the original measurements to confirm the results presented. Calculations based on the measurements in data set A do not give the results presented in the published paper or support the conclusions there. The reasons you have given to discard these measurements are not particularly convincing. As was indicated in an earlier letter to you, our calculations on data set B showed several striking and important deviations from the results presented in the published version of the paper. I cannot see that you in your email of 00-11-09 have been able to rebut the results of our calculations. To be sure that we have done our calculations correctly I last week asked another person, also well familiar with statistical calculations in ecology, to independently repeat these calculations. The results obtained in these new calculations were exactly the same as in the earlier ones. Therefore, it has not been possible for us to confirm the results presented in the published paper.

Various kinds of errors may now and then occur in scientific articles and can be corrected. However, to me it is now quite obvious that Oikos with your paper has published an article that does not reach the most elementary scientific requirements and therefore has to be retracted. From what you say in the last paragraph of your email of 00-11-09 it seems to me that you also are inclined to do so. Only a short note from you and your co-author is necessary. In order to make it easier for you I can also propose the following text:

Retraction

Herbivory affects developmental instability of stone oak, *Quercus rotundifolia*. – Oikos 82:246-252, 1998.

In this article we reported data on size and (absolute and relative) asymmetry of stone oak (*Quercus rotundifolia*) leaves from Spain. It now appears that the measurements and analyses behind the data in the article were flawed and misinterpreted, implicating that the conclusions drawn are invalid. We therefore retract the article.

A. P. Møller F. de Lope

As I have indicated in my earlier letters, another possibility is that the Editor is made responsible for the retraction of the article. I have then to rely on the paper submitted by Rabøl and Andersen and write a text summarising most of what I have written above.

(Nils Malmer)
Editor-in-Chief