



## Comment

**Cite this article:** Henshaw JM, Holman L.

2015 Bet-hedging via polyandry: a comment on 'Mating portfolios: bet-hedging, sexual selection and female multiple mating'.

*Proc. R. Soc. B* **282**: 20150346.

<http://dx.doi.org/10.1098/rspb.2015.0346>

Received: 13 February 2015

Accepted: 10 April 2015

**Author for correspondence:**

Jonathan M. Henshaw

e-mail: [jonathan.henshaw@anu.edu.au](mailto:jonathan.henshaw@anu.edu.au)

The accompanying reply can be viewed at <http://dx.doi.org/doi:10.1098/rspb.2015.0866>.

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2015.0346> or via <http://rspb.royalsocietypublishing.org>.

# Bet-hedging via polyandry: a comment on 'Mating portfolios: bet-hedging, sexual selection and female multiple mating'

Jonathan M. Henshaw and Luke Holman

Division of Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra, Australia

Garcia-Gonzalez *et al.* [1] conducted an original and elegant experiment examining whether fertilization of a female's eggs by multiple males (polyandry) can provide fitness benefits via 'bet-hedging' (i.e. due to decreased variance in offspring fitness). The authors measured these benefits in both stable and variable environments, and also quantified the joint fitness consequences of bet-hedging and sperm competition. We believe that the study's experimental design is sound, but that its statistical analysis was incorrect. Here, we reanalyse the raw data and find that all but one of the study's results is consistent with the null hypothesis that polyandry does not provide benefits via bet-hedging, contrary to the original conclusions.

Garcia-Gonzalez *et al.* [1] compared the fitness of females under experimentally imposed monandry and polyandry. The eggs of 12 females were divided into three batches, representing three 'generations'. For each female and generation, half of the eggs were fertilized using the sperm of a single male (monandry) and the other half with sperm from three males (polyandry). The authors measured fertilization rates and offspring viability (the latter in two different environments, termed A and B) for each of the 12 females and three generations.

From the viability and fertilization data, the authors calculated the between-generation geometric means of each fitness measure ( $W_{BG}$ ) under polyandry and monandry, and calculated the difference in  $W_{BG}$  between treatments as

$$\Delta_{Geo} = \text{Poly } W_{BG} - \text{Mon } W_{BG}. \quad (1.1)$$

A positive value of  $\Delta_{Geo}$  implies that polyandry improves geometric mean fitness. To account for variation in  $\Delta_{Geo}$  due to the arbitrary ordering of generations, the authors repeated this calculation  $10^4$  times while randomly varying the order of generations for each female, and generated 95% central ranges for  $\Delta_{Geo}$ , which they took as approximate confidence intervals. Similar randomization procedures were used to simulate various regimes of stable and fluctuating environmental conditions.

This method of generating confidence intervals is problematic. It accounts for one source of variation in the sample (i.e. the ordering of generations) but neglects the variance that results from sampling a finite number of males, females and offspring from the population. This causes the analysis to substantially underestimate the size of the confidence intervals around  $\Delta_{Geo}$ , and thereby to produce false positives.

Moreover, we believe that the geometric mean is not the most appropriate fitness measure for Garcia-Gonzalez *et al.*'s analysis. In cases where (i) the absolute fitness of each strategy comes from the same probability distribution in all generations and (ii) fitness is uncorrelated among individuals of the same genotype (here, monandrous and polyandrous females), it is more appropriate to apply Gillespie's measure [2–4]

$$\Delta_{Gill} = \mu_{\text{Poly}} - \mu_{\text{Mon}} - \frac{1}{N}(\sigma_{\text{Poly}}^2 - \sigma_{\text{Mon}}^2), \quad (1.2)$$

here  $\mu_i$  and  $\sigma_i^2$  are the population mean and variance in reproductive success of each strategy within a generation, and  $N$  is the population size. The assumptions behind Gillespie's measure are met for those treatments of Garcia-Gonzalez *et al.*

**Table 1.** Recalculated effect sizes and  $p$ -values for the experiments of Garcia-Gonzalez *et al.* [1]. The first column shows the difference in fitness between polyandry and monandry treatments, while the second and third columns show the 95% central ranges for this figure. The last column shows the approximate two-tailed  $p$ -value, interpretable as the probability that the experiment produces a difference at least this great by chance alone given the null hypothesis that polyandry has no effect on fitness. Results are shown for Gillespie's measure  $\Delta_{\text{Gill}}$  (first row), the difference in geometric mean fitness  $\Delta_{\text{Geo}}$  following our analysis (second row) and Garcia-Gonzalez *et al.*'s original analysis based on  $\Delta_{\text{Geo}}$  (third row, in parentheses) for each expression and trait. Positive values of  $\Delta_{\text{Gill}}$  and  $\Delta_{\text{Geo}}$  indicate that polyandry had higher fitness than monandry. We multiplied all values by 100 for ease of comparison with the original study, in which fertilization rates and offspring viability were measured as the percentage of eggs fertilized and the percentage of offspring surviving until 8 days, respectively.

trait	mean fitness difference ( $\Delta_{\text{Gill}}$ or $\Delta_{\text{Geo}}$ )	lower 95% central range	upper 95% central range	approx. $p$ -value
experiment 1				
offspring viability	5.2	−0.4	10.9	0.08
environment A	5.3	−0.4	11.1	0.08
	(5.4)	(5.0)	(6.0)	
offspring viability	−4.7	−10.8	1.1	0.14
environment B	−4.7	−10.9	1.2	0.14
	(−4.8)	(−5.2)	(−4.4)	
offspring viability	0.3	−4.4	5.1	0.91
across environments	0.3	−4.4	5.1	0.91
(A + B)	(0.3)	(0.1)	(0.7)	
fertilization rates	−7.5	−16.5	1.1	0.11
	−7.7	−16.9	1.2	0.10
	(−7.8)	(−8.7)	(−6.1)	
experiment 2				
offspring viability	−0.2	−4.3	3.5	0.91
environment A	−0.2	−4.4	3.5	0.92
	(−0.3)	(−0.5)	(0.0)	
offspring viability	2.2	−2.9	7.0	0.40
environment B	2.3	−3.0	7.1	0.39
	(2.2)	(1.8)	(2.7)	
offspring viability	1.0	−2.4	4.3	0.59
across environments	1.0	−2.4	4.4	0.59
(A + B)	(1.0)	(0.8)	(1.3)	
fertilization rates	12.8	5.4	20.6	0.002
	12.7	5.4	20.7	0.002
	(12.8)	(12.5)	(13.5)	

in which the environment is assumed constant between generations. By contrast, the geometric mean fitness is a more appropriate measure when the expected absolute fitness of a strategy fluctuates between generations, but there is minimal within-generation variation in fitness among individuals of the same genotype [3–6]. The geometric mean can also be applied to relative fitness, regardless of the underlying assumptions, but this approach is of limited practical use because calculating relative fitness requires knowing the frequencies of each genotype in the population [3].

Garcia-Gonzalez *et al.* also simulated fitness in deterministically alternating environments of the form ABA and BAB. In these environments, neither of the above fitness measures strictly applies. This is because there is both between-generation variance in a strategy's expected success (making Gillespie's measure inappropriate) and also substantial within-generation variance among individuals playing the same

strategy (which violates the conditions for the geometric mean fitness). These environmental regimes necessitate a more complex analysis, which we omit for brevity (cf. [7]).

Here, we estimate confidence intervals using a bootstrapping method that accounts for the missing sources of sampling variance. We also estimate the probability of obtaining the observed values under the null hypothesis that there is no difference in fitness between monandrous and polyandrous treatments (a statistic not provided in the original paper). We present results both for the difference in geometric mean fitness  $\Delta_{\text{Geo}}$  and for Gillespie's measure  $\Delta_{\text{Gill}}$ .

## 2. Generating effect size confidence intervals

The original experiment generated 36 data points (12 females measured over three generations) for each fitness component

(fertilization rates and offspring viability in environments A and B) under both monandry and polyandry. We resampled from these data  $10^4$  times to perform the bootstrap analysis. For each run, we sampled 36 monandrous data points from the original 36 with replacement. Matching polyandrous data points were selected so as to maintain the pairings of female and generation from the original experiment.

For each run, we calculated the difference in geometric mean fitness ( $\Delta_{\text{Geo}}$ ) and Gillespie's measure ( $\Delta_{\text{Gill}}$ ) using equations (1.1) and (1.2), respectively. For the geometric mean fitness, we split the 36 data points into three generations of 12 females at random, maintaining the pairing of treatments as above. For Gillespie's measure, we approximated the population means and variances by the sample means and unbiased sample variances and, for consistency, assumed a population size of  $N = 12$ . Because individual females and males appear multiple times in the original experiment, this procedure will tend to underestimate the true population variances, and hence the strength of bet-hedging effects. We nevertheless believe this pseudoreplication is unlikely to affect the results strongly.

For both fitness definitions, we calculated the mean and 95% central range from the bootstrap distributions (table 1); the latter provides an estimate of the 95% effect size confidence intervals (the true CIs are probably wider, due to additional variance in the population that is not captured by the sample). This procedure was performed using data from environments A, B, and an average of these two environments, as described in the original study and in our Mathematica code (see electronic supplementary material).

### 3. Null hypothesis significance testing

We also used bootstrapping to simulate the expected distribution of the mean values of  $\Delta_{\text{Geo}}$  and  $\Delta_{\text{Gill}}$  under the null hypothesis that there is no difference between the polyandrous and monandrous treatments. In each of  $10^4$  runs, we randomly swapped the monandrous and polyandrous data points within each of the 36 data pairs with probability 0.5. This gave us two new 'treatment groups', each consisting of random mixtures of the original monandry and polyandry treatments. We calculated mean  $\Delta_{\text{Geo}}$  and  $\Delta_{\text{Gill}}$  for each run using the same method as above in order to obtain their approximate distributions under the null hypothesis.

## References

- García-González F, Yasui Y, Evans JP. 2015 Mating portfolios: bet-hedging, sexual selection and female multiple mating. *Proc. R. Soc. B* **282**, 20141525. (doi:10.1098/rspb.2014.1525)
- Gillespie JH. 1974 Natural selection for within-generation variance in offspring number. *Genetics* **76**, 601–606.
- Frank SA, Slatkin M. 1990 Evolution in a variable environment. *Am. Nat.* **136**, 244–260. (doi:10.1086/285094)
- Starrfelt J, Kokko H. 2012 Bet-hedging—a triple trade-off between means, variances and correlations. *Biol. Rev.* **87**, 742–755. (doi:10.1111/j.1469-185X.2012.00225.x)
- Dempster ER. 1955 Maintenance of genetic heterogeneity. *Cold Spring Harbor Symp. Quant. Biol.* **20**, 25–32. (doi:10.1101/SQB.1955.020.01.005)
- Gillespie JH. 1973 Natural selection with varying selection coefficients—a haploid model. *Genet. Res.* **21**, 115–120. (doi:10.1017/S001667230001329X)
- Seeger J, Brockmann HJ. 1987 What is bet-hedging? In *Oxford surveys in evolutionary biology*, vol. 4 (eds PH Harvey, L Partridge), pp. 182–211. Oxford, UK: Oxford University Press.

We next calculated the proportion of mean  $\Delta_{\text{Geo}}$  and  $\Delta_{\text{Gill}}$  generated under the null model that were at least as large (in absolute value) as the means calculated in the previous section, giving an approximate two-tailed  $p$ -value (table 1). One can interpret this  $p$ -value as the probability of seeing a result at least as extreme as the observed one, under the assumption that mating treatment has no effect on the distributions of  $\Delta_{\text{Geo}}$  and  $\Delta_{\text{Gill}}$ . Exact  $p$ -values would probably be larger, due to additional variance in the population that is not captured by the sample.

## 4. Results and discussion

Table 1 shows that most of the comparisons made by García-González *et al.* [1] yielded results consistent with the null hypothesis ( $\alpha = 0.05$ ) after reanalysis, suggesting that fitness did not differ significantly between polyandry and monandry treatments. Nevertheless, the estimated 95% CIs often included large differences, suggesting that this dataset does not rule out the existence of a substantial benefit from bet-hedging via polyandry. The two fitness measures ( $\Delta_{\text{Geo}}$  and  $\Delta_{\text{Gill}}$ ) gave quantitatively similar results, largely because variance in both measures was dominated by statistical fluctuations in the sample means. These fluctuations would be reduced with a larger sample size. Because the revised confidence intervals are large, our analysis highlights that a greater degree of replication (especially of females) is required to measure the benefits of bet-hedging with sufficient precision. We applaud García-González *et al.*'s efforts to establish a 'proof-of-principle' approach for studying bet-hedging in isolation from other factors, and we hope that our modified statistical approach proves useful to future experiments.

**Data accessibility.** Mathematica code for the statistical model is provided in the electronic supplementary material.

**Authors' contributions.** J.M.H. and L.H. conceived of the study and wrote the manuscript. J.M.H. performed the statistical analysis.

**Competing interests.** We declare we have no competing interests.

**Funding.** Funding was provided by an Australian Postgraduate Award to J.M.H. and an Australian Research Council DECRA to L.H.

**Acknowledgments.** We thank Francisco García-González for insightful and generous comments throughout the writing and review of this manuscript. We also thank Anne Lizé and two anonymous reviewers for their helpful feedback.