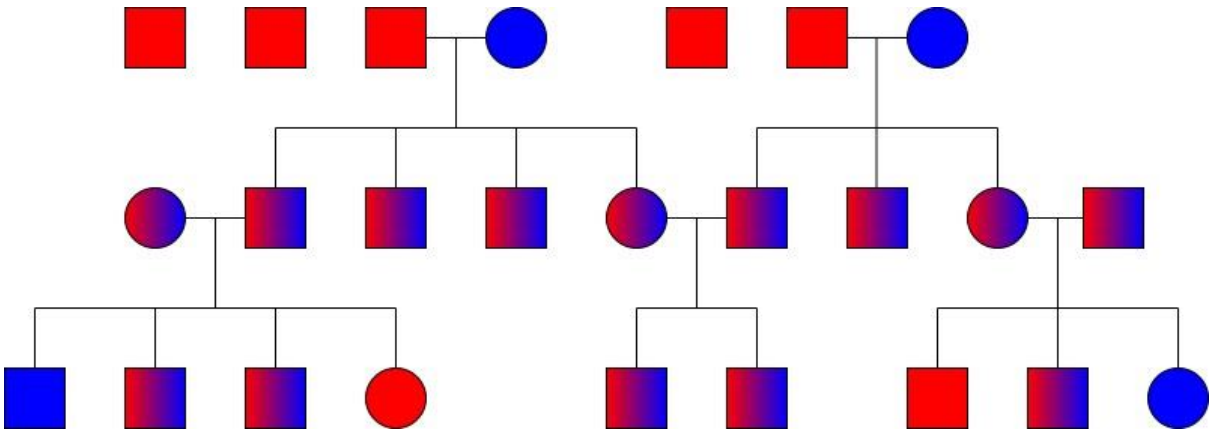
Adding sexes to life history theory

Validating a field-data based estimate of reproductive value in populations with sexes

Pelle Scholten & Sjors van den Hoogen
Advisors: M. Borger & F. Weissing
University of Groningen, the Netherlands

April 7, 2022

ABSTRACT
Reproductive values are a useful measurement of fitness and can lead to insights in evolutionary biology and population dynamics by comparing individuals or categories in a population. A new method, the pedigree-based method, of estimating this reproductive value (RV) is easier to perform but has not yet been properly validated in populations with sexes. We performed computer simulations based on different life history scenarios to assess individual and median estimates of the pedigree-based reproductive value (pRV) and compare these values to the theoretical reproductive values. We found that for the same simple life histories, the individual pRV estimates diverged significantly over time, while the median of the pRV estimates approached the theoretical RV. The median of the estimates was able to clearly show differences in the RV of different categories of individuals based on different ratios between the sexes and different life history parameters. The variation in the pRV estimates means a single pedigree-based estimate of the RV in the field, is not reliable.



INTRODUCTION

An important concept in biology and studying populations is fitness. It tells how well an organism is adapted to its environment. Furthermore, fitness describes how successful an organism should be in passing on its genes (Kimbrough, 1980). To estimate fitness levels of organisms in the wild, is a harder task. Several parameters, such as survival rate and number of offspring are studied and mentioned as possible indicators of fitness (Roff, 1992). Such parameters do help in estimating fitness levels but are not complete, because there are multiple other factors that play a role too in determining fitness levels.

Two proxies that are often used to estimate fitness are expected lifetime reproductive success and the reproductive value (Fisher, 1930). The first is described as the exact number of offspring that an individual produces during its life. Since this method only requires counting the offspring of individuals over their entire life, it is a relatively easy method. The downside of this method is that this value does not include the rate of reproduction and the possibility for different types of offspring. (Brommer et al., 2004) Reproductive value is defined as the expected relative contribution to the future gene pool of an individual of a certain class (Grafen, 2006; Taylor et al. 1974). It includes the specific life history of individuals but thus requires a lot of information about this life history, such as survival rates, fecundities, behavioral choices, and kind of offspring produced, before the reproductive value can be measured. Besides, these life history parameters are estimated and therefore not perfectly accurate and precise. This makes it a lot more complicated to obtain reproductive values compared to the estimated lifetime reproductive success. Also, incorporating all these values in a model to generate reproductive values is much harder than simply calculating the lifetime reproductive success.

Recently, a new type of model came up to estimate the reproductive values of individuals. This model uses genetic pedigree data to construct the reproductive value (Barton and Etheridge, 2011; MacCluer et al, 1986). This method is increasingly used since then (Chen et al., 2019; Hunter et al., 2019). At the start of the measurements the DNA of the initial population, the ancestors, gets sequenced and a unique marker gene that is passed to future generations is assigned to each individual. After a few generations, the DNA of the descendants of these ancestors is sequenced as well. Now, it is possible to determine which unique marker genes are present in the current population and you measure the actual contribution of ancestors to future generations. This is called gene-dropping. The estimate for reproductive value in this model is stated as the average per capita number of descendants of the members of a certain class of individuals. As one can imagine, this method requires deep and complete pedigree data, but the advantage is that it does not need extensive life history parameter values.

Borger et al. (2022) calculated reproductive values using this pedigree method with a simulation study and compared the results with reproductive values calculated from life history models and with the “true” reproductive values (used to setup the simulations). They found that reproductive values estimated from the pedigree method were often highly inaccurate and not precise. In their study, the model assumptions were purposely kept simple, focusing on simple life histories. They also made some more complex life histories including different variables to see if these yielded better results. They included fluctuating environments, sexual or asexual reproduction, extending the number of time steps, two- and

three-stage life histories, they extended the number of time steps to a large time scale, and they looked at the effect of RV estimations based on individuals versus groups.

One parameter that was not included by Borger et al. (2022) yet, but is worth some further research, is the variable sex. Fisher (1930) explained that the sex ratio of most sexually reproducing species is 1:1. If this ratio is skewed, the reproductive value of the sex that is overabundant is lower. Offspring always have one father and one mother, so the reproductive value of all males must be equal to the reproductive value of all females. One can imagine that if there are 200 males and 400 females, the reproductive value of a single male therefore on average must be double the average RV of a single female. Furthermore, there are several examples of different sexes having different life history strategies, fecundities, survival and death rates due to genetical, morphological and behavioral differences (Clobert, 1988). Additionally, there can be differences between sexes in the hierarchy. Especially for males, dominant males might mate with a lot of females while subordinates do not. This makes sex as a variable an interesting addition for the model. This study will test the pedigree-based estimate of the reproductive value in individual-based simulations that include different sexes. The accuracy, precision, and the median of the pRV can be measured over multiple simulations. These simulations will allow us to answer the question: to what extent is the pedigree approach able to quantify the reproductive value for males and females? We will tackle this question by answering the following sub-questions. What is the effect of different sex ratios in a population on the pRV estimates? What is the effect of different life history parameters between the sexes on the pRV? How do differences in life history strategies concerning dominant and subordinate behavior by males affect the pRV?

To answer these questions, we will add sex as a variable in Borger's model. We will extend the model in C++ by adding the sex variable and generate a pedigree which will be used to calculate the pRV. We will track the individuals and their offspring using gene markers to form a pedigree. This pRV will then be compared to the theoretical "true" RV, which can be calculated using a mathematical approach with life history matrices. Life history parameters will be set by us in the model and will yield the "true" RV. These two reproductive values can then be analyzed and conclude if adding sex as a variable improves the accuracy and precision of the estimation of RV via pedigrees.

METHODS

All simulations are individual based and based on specific life history scenarios, as described in figure 1. Scenario 1 only exists of males and females. Scenario 2 has males and females of different age classes. These scenarios will be further explained below.

Time proceeds in discrete time steps, where every time step matches one reproductive season. In our models we usually used 20 time steps, unless stated otherwise. In a reproductive season, an individual can reproduce and after that either survive and stay in their age class, survive and change age class or die.

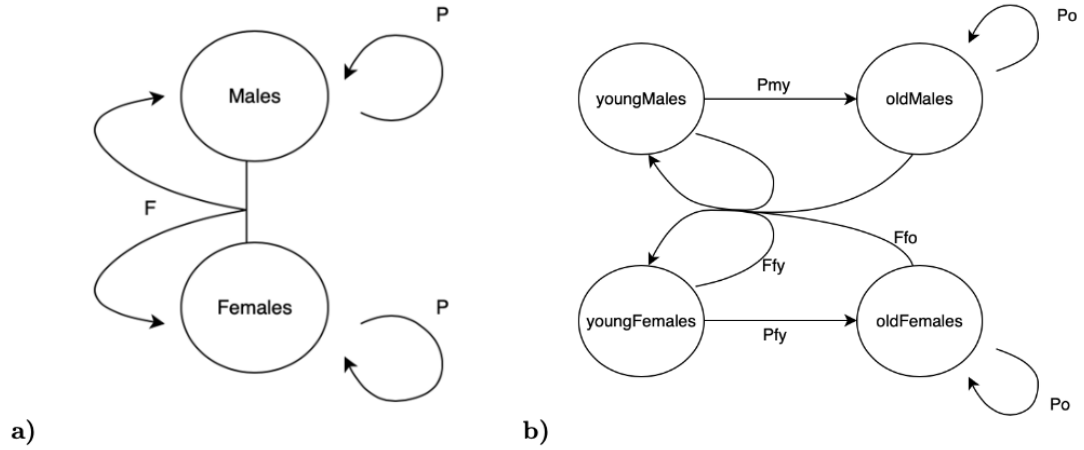


Figure 1: Life history diagrams. Each circle represents a type of individual in the population. The arrows represent the expected fecundity (F) or survival (P) of individuals. **(a)** Scenario 1: P is the density dependent survival probability and F is the average number of offspring produced per individual per reproduction event **(b)** Scenario 2: Pmy and Pfy are the probabilities for young males and young females to survive to adulthood respectively. Ffy is the average fecundity of young females. Ffo is the average fecundity of old females. Po is the density dependent survival probability of old males and females. The fecundity of males depends on their mate.

Reproduction occurs first. We used a reproductive system where all females always mate once. A female is assigned to a male which is picked randomly from the population. This means that some males mate once, some mate multiple times and some do not mate at all. The number of offspring produced per mating is determined by a Poisson distribution around a predetermined fecundity (F). All offspring survive at least until the next reproductive season and enter the population at t+1. Each offspring inherits one gene of its mother and one gene of its father. Which gene is inherited from the parent is determined has equal probability. The sex of the offspring is determined by a discrete distribution, using the probability p of producing a son (and 1-p for the probability of producing a daughter).

After reproduction, survival occurs. The survival probabilities are determined so that the population remains stable in size. Which individuals survive is determined by a Bernoulli distribution with the survival probability as its mean. The survival probability can be density dependent, and it is then calculated by the following equation:

$$\frac{1}{1+\alpha \cdot N} \qquad 1$$

Here, N is population size and alpha is the intensity of the density dependence. We set the initial population size and then adjusted the parameter alpha depending on the model in such a way that the population remained stable around 1000 individuals.

Different types of density dependent population were also investigated. Density dependency in the young population, density dependency in only one sex, and density dependency in the fecundity were investigated.

The effect of the number of generations was also tested to see what happens to the pRV when more time passes. Simulations will be run with 100 generations instead of 20 generations.

Every generation, the pRV is calculated using 'gene-dropping'. The initial population started in a stable stage distribution. Therefore, there was no need to wait with gene-dropping until the populations were stable. At initialization of the populations, every individual was given 2 genes. These genes are unique per life history stage, and individuals in these stages were homozygous for this gene. For example, in the most basic scenario, all males were homozygous for gene '1', and all females were homozygous for gene '2'. These genes spread throughout the population, through inheritance by the offspring. When a male and female mate, the offspring gets one of the genes of its mother and one of the genes of its father. There is an equal probability to get either gene from its parent. The pRV is then calculated by counting how many of each of the genes of the original population are present in the current population. This count is then divided by 2 as every individual has 2 genes and the theoretical reproductive value is measured on the individual level. To further scale this count back to the individual level, this count is divided by the number of individuals of that age class at the start of the simulation, to calculate the number of offspring per original individual in each age life history stage. The reproductive value is always relative, therefore, the pRV of the youngest male population is always set to 1. The pRV of the other stages is then calculated by dividing the number of offspring per original individual of that stage by the number of offspring per original individual of (young) males.

Stochasticity in individual reproductive success is extensive. Estimating RV based on single individuals hence is highly unreliable (Borger, 2022; Chen, 2019). Therefore, we base our estimates of RV on populations rather than individuals. One simulation is comparable to field situations. The pRV will be calculated for 100 simulations per life history and parameter setting. To easily be able to see the variation between simulations we will make graphs that shows the median, the 50% central values and the 90% central values of these 100 simulations.

Scenario 1:

This scenario is the simplest scenario. There is a male population and a female population. The density dependent survival rate is determined solely on the size of the adult population, not on the number of offspring that was just born. Each simulation was done for twenty generations and was repeated a hundred times. The RV of the males was set as one, and the female RV was calculated using Fisher's Principle. For simplicity, we assumed that all offspring from 1 female came from the same male individual, and that the initial population was homozygous.

We used Fisher's Principle to estimate the true reproductive values. This principle states that because offspring have one father and one mother, in a population the total RV of all males and females is the same. This means that when there are more males than females, the RV of a single female individual is on average higher than the RV of a single male individual. We used set ratios and proportional start population sizes so that the true RV was equal to the sex ratio. For example, when there are 3 times as many males as females, the RV of a female is 3 compared to the RV of a male which is set to 1.

Scenario 2:
This scenario has four life history states. Males and females can be young or old. Both young and old individuals reproduce, but can have different fecundities. Old females can mate with young and old males, as so can young females.. The fecundity is solely determined by the age of the female. The survival probability of young individuals is fixed and can differ between the sexes, Pmy for males and Pfy for females. The survival probability of old individuals is density dependent and is the same for both males and females.

In scenario 2 it is not possible to use Fisher's principle to calculate the true RV. Therefore, recurrence relations were used **(**see appendix A**)**. Equations including the life history parameters were made to calculate the stable stage distributions and the reproductive values and were solved in Excel. The life history parameters that were fixed, such as fecundity and death rate of young individuals were put in directly. The survival probability of adult individuals was calculated by solving the density dependent formula (1) for the alpha used in the model and N = 1000.

Technical note
Simulations were written in C++. Figures were made in R3.4.1, with the packages ggplot2 and cowplot.

RESULTS
Scenario 1:
When looking at multiple simulations of a population with the same life history and the same life history parameters, there is a significant variation in the pRV (see figure 2) . The variation arises due to the stochastic effects in the population, which can outweigh the effects of the true reproductive value of an individual. This variation is partly dependent on the values of the life history parameters. For example, when the fecundity and the death rate are both higher, there is more variation in the pRV compared to a population with a low fecundity and death rate. Individual simulations usually have a pRV that is either significantly lower or higher than the true RV. They also do not stabilize over time. The median of the pRV's does approach the theoretical value of the RV. The accuracy of the pRV estimates consequently seems fine. The average sometimes also reflects the theoretical RV, but in other cases is a multiple of the true RV because there are very large estimates of the pRV in the simulations. This implies that the precision of the pRV seems systematically incorrect.

The results of the one generation model clearly reflect Fisher's principle. When the population ratio is skewed, one sex gets a higher reproductive value (see figure 2). At first glance it looks like there is more variation in the scenario with a 3 to 1 male to female ratio. However, as the RV is higher for this ratio compared to the other ratios, the variation is also higher. The relative variation is equal for the different ratios.

The variation becomes larger with a smaller population size. The graphs in figure 3 show population sizes that are more realistic for field-work. Here the variation is significantly larger than in the larger population sizes, even for this relatively simple life history.

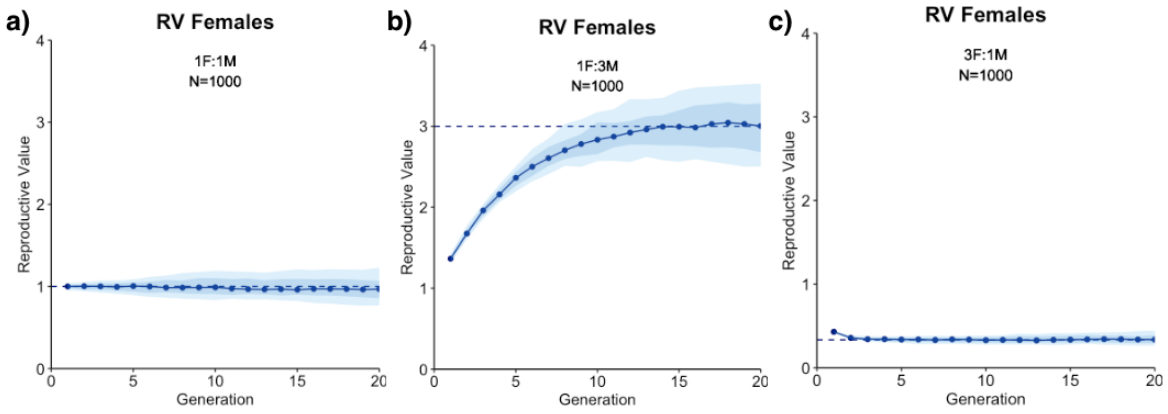


Figure 2. Example simulations of scenario 1, for three different male:female ratios, our one-stage model. Dashed horizontal lines represent the true RV. Dotted lines indicate the medians of the 100 simulations per time step. The dark blue (light blue) bands represent the 50% (90%) confidence intervals of the simulations. Parameter values: F=1 for all graphs, alpha is 0.001 for a, 0.00033 for b, 0.003 for c. **(a).** Reproductive value of females in a population with a 1F:1M ratio. **(b)** Reproductive value of females in a population with a 1F:3M ratio. **(c)** Reproductive value of females in a population with a 3F:1M ratio.

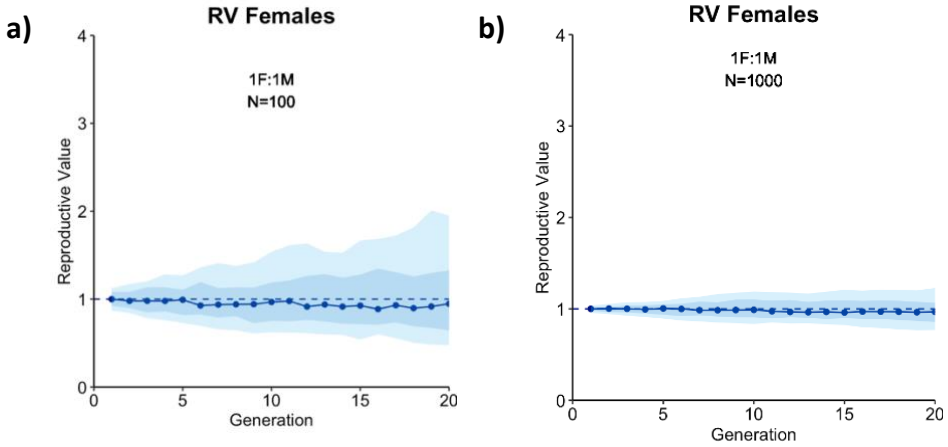
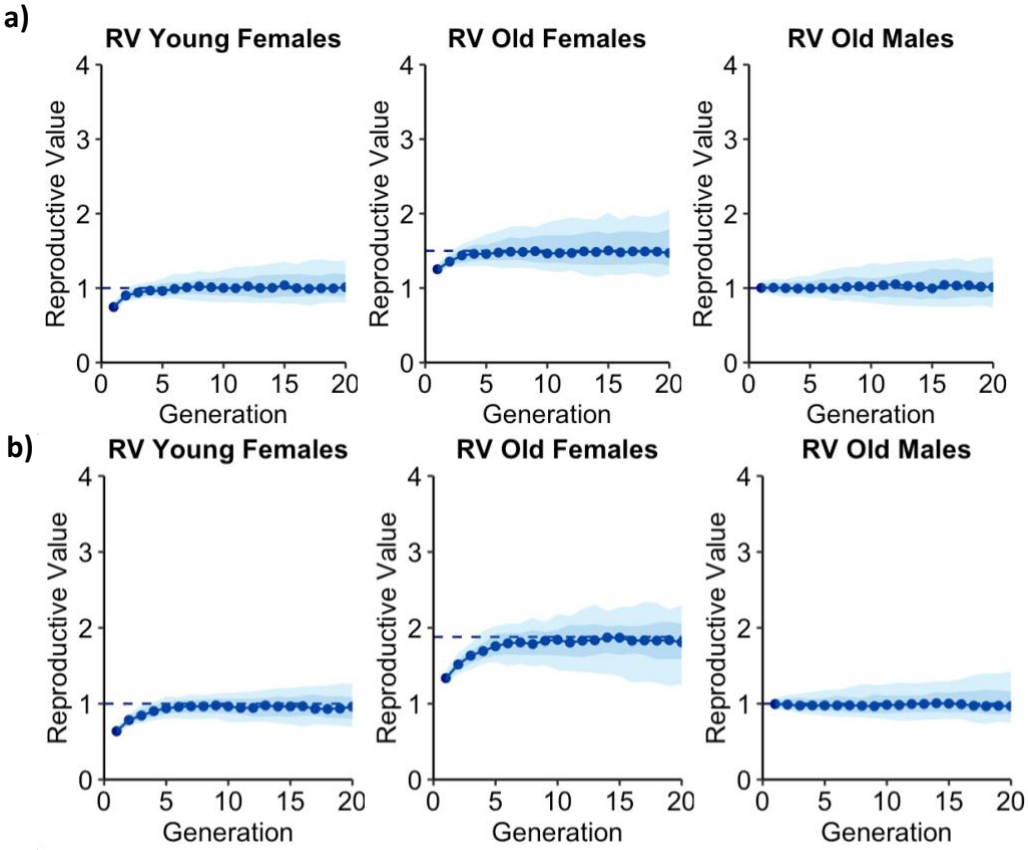


Figure 3. Simulations of scenario 1, one-stage model. Dashed horizontal line represent the true RV. Dotted line indicate the medians of the 100 simulations per time step. The dark blue (light blue) band represents the 50% (90%) confidence intervals of the simulations. Parameter values: F = 1.0 for both graphs **(a).** Reproductive value of females in a population with a 1F:1M ratio and a population size of 100 individuals and an alpha of 0.01. **(b)** Reproductive value of females in a population with a 1F:1M

In the beginning of the simulations the variation is the smallest but in most cases it is not yet near the theoretical reproductive value. Over time, variation in the pRV increases and the median comes closer to the theoretical value. How fast this happens differs between life histories and parameter settings. Usually, the estimate is best after 5 to 10 generations. However, sometimes the best estimate is earlier, but can even be as late as 12 generations.

Increasing the time horizon from 20 to 100 generations did not change the results.



ratio and a population size of 1000 individuals and alpha of 0.001.

Figure 4. Simulations of scenario 2, the four-stage model. The population has 1000 individuals and the offspring sex ratio is 1F:1M. Dashed horizontal lines represent the true RV. Dotted lines indicate the medians of the 100 simulations per time step. The dark blue (light blue) bands represent the 50% (90%) confidence intervals of the simulations. **(a)** Equal survival probability for young males and young females. Parameter values: Fy=0.5, Fo=1.5, Pmy = 0.5, Pfy =0.5, alpha = 0.001 **(b)** Different survival probability for young males and young females. Parameter values: Fy=0.5, Fo=1.5, Pmy = 0.6, Pfy =0.4, alpha = 0.000667

Scenario 2:

The results from the life history with 2 age classes are similar to the results of 1 age class. Again, the median of the simulations corresponds well with the theoretical RV. Variation arises over time and appears similar to scenario 1 (see figure 4.a).

None of the different types of density dependent population size control gave different results than the scenario with density dependency in the adult survival probability.

When the survival probability of males is higher than females, the median of the pRV's still coincides with the theoretical value. However, this came with an increase in the variation (see figure 4.b).

DISCUSSION

Reproductive value is a proxy to measure fitness. This can represent individuals or populations (Grafen, 2006). The reproductive value is the expected contribution to future generations of individuals. The pedigree based reproductive value is increasingly used and

researched by us. We find that estimates of the reproductive value are unprecise, also when adding sex in the life history. Our estimates of the reproductive value are somewhat more accurate than models that did not included sex as a variable. Still, there is too much variation to consider the pedigree-based method reliable.

Previous research of Borger et al (2022) showed that the method has some severe drawbacks. Now, our research shows approximately the same results as Borger's. Pedigree based reproductive value estimates are unprecise. She argued as well that the traditional way of model-based pedigree estimates is better. The median of 100 simulations does reflect the true RV in our study, so it seems a bit more accurate than models without sexes, but still the variation is huge, especially for realistic population sizes (N = 50 – 100). Thus, using this method for a single population in the wild is not reliable. When the population size is increased a lot (N=1000), the variation seems to drop, but this is an unrealistic population size because the pedigree has about 100000 individuals each year. Additionally, individual simulations differ a lot from each other, and this divergence is even stronger when the time span increases. Individual estimates of the reproductive value can be a multiple of the true reproductive value. Since one simulation reflects one empirical study population, the pedigree estimate is normally not a good indication of the true RV, if it is based on one study. The individual estimates of the RV did not improve in populations with different sex ratios. In populations with different age classes and different survival parameters between the sexes, the median was again able to closely reflect the true reproductive values but did not cause a decrease in the variation of the individual estimates.

Other studies done on pedigree reproductive values (pRVs) also indicate these results of too much variation in pedigree-based estimates of the RV. In Chen's (2019) research, different genetic contribution to future generations were made, and different estimates for pRVs were measured for different cohorts in the same population. Chen suggested this was caused by stochastic effects and high mortality in young individuals. In Hunter's (2019) research she shows a high amount of variation in the pRV of sparrow populations, also caused by stochastic effects. Interestingly, our study did not show the zigzagging pattern at the start of the simulations that Borger et al (2022) found. In our results, this zigzag is a curve that moves towards the median. The time it took for the median to converge to the true RV remained approximately the same as in the study of Borger et al (2022).

The large variation of RV estimates arises due to the stochastic effects that occur in a population. These effects are likely even larger in real populations, as in real life, life histories are more complex than the ones we used. For example, emigration or immigration is not included in our models, but are happening in real populations. Another limitation for the pedigree-based method is that it is expensive to map a whole population, let alone the fact to catch and sample a whole population. Furthermore, emigration might take place, and could cause problems because it is unknown if an organism has emigrated or died. Real populations are also more dynamic than our model populations, so numbers per population will change more over time, and it is unlikely that they will stay in a stable stage distribution over many generations.

Our model was kept simple, focusing on a few life histories. In more complex life histories, estimates might vary even more. On the other hand, different life histories could also improve the accuracy and precision of the RV estimates. For example, a life history with

low stochasticity might dampen the variation. However, all the different life histories and parameter combinations we tried resulted in imprecise estimates. RV estimates probably would get worse in more complex life histories. For example, if survival probabilities for males are lower than that of females, and species have a high lifespan. This is the case in many species but was not yet included in our models. This might be a relevant subject for future research. The research question we did not answer is how the pRV is affected by a population with dominant and subordinate males, and a consequently different number of mates per male. We suspect that this will only increase variation more, since fewer individuals contribute to the future generations in that case, and hence the pRV will be more susceptible to stochasticity.

To conclude, our results that the pedigree-based method for calculating the RV is not accurate and precise, may disappoint some biologists. The traditional way of calculating the RV with life history parameters still is more accurate and precise than the pedigree estimates. However, these history parameters can be hard to estimate, which influences the reliability of the model-based RV estimates.

REFERENCES

Barton NH, Etheridge AM, (2011). "The relation between reproductive value and genetic contribution". *Genetics* 188:953-973. doi: 10.1534/genetics.111.127555.

Borger, M.J., Komdeur, J., Richardson, D.S., Weissing, F.J. (2022) "Putting life history theory to the test - the estimation of reproductive values from field data", *The American Naturalist*, under review.

Brommer, J. E., (2007). "The evolution of fitness in life-history theory". *Biological Reviews*, 75(3), 377–404. https://doi.org/10.1111/j.1469-185X.2000.tb00049.x

Chen N, Juric I, Cosgrove EJ, Bowman R, Fitzpatrick JW, Schoech SJ, Clark AG, Coop G, (2019). "Allele frequency dynamics in a pedigreed natural population". *Proceedings of the National Academy of Sciences of the United States of America* 116:2158-2164. doi: 10.1073/pnas.1813852116.

Clobert, J., Perrins, C. M., McCleery, R. H., & Gosler, A. G. (1988). "Survival rate in the great tit parus major in relation to sex, age, and immigration status". *Journal of Animal Ecology*, 57(1), 287–306.

Fisher RA, (1930). "The Genetical Theory of Natural Selection". Oxford: *Oxford University Press.*

Grafen A, (2006). "A theory of Fisher's reproductive value". *Journal of Mathematical Biology* 53:15-60. doi: 10.1007/s00285-006-0376-4.

Hunter D, Slate J, (2019). "Understanding genetic changes between generations". *Proceedings of the National Academy of Sciences of the United States of America* 116:1834-1836. doi:10.1073/pnas.1820938116.

Kimbrough, S.A. (1980). “The concepts of fitness and selection in evolutionary biology”. In: *Journal of Social and Biological Structures* 3: pp. 149–170.

MacCluer, J. W., VandeBerg, J. L., Read, B., & Ryder, O. A. (1986). “Pedigree analysis by computer simulation”. *Zoo Biology*, 5(2), 147–160. https://doi.org/10.1002/zoo.1430050209

Roff, D. A. (1992). “The evolution of life histories: theory and analysis”. *Springer Science & Business Media.*

Taylor, H. M., Gourley, R. S., Lawrence, C. E., & Kaplan, R. S. (1974). “Natural selection of life history attributes: an analytical approach”. *Theoretical Population Biology*, 5(1), 104–122. https://doi.org/10.1016/0040-5809(74)90053-7

APPENDIX

A: Equation for calculating the theoretical RV in scenario 2

The following formulas were used to calculate the stable stage distributions and the theoretical reproductive value in scenario 2. The input values were taken from the simulation parameters. When there was density dependency, equation 1 was used to calculate the survival probability where the α was taken from the simulation and N was assumed to be 1000.

2. $n_{f,o} = P_{f,y} \cdot n_{f,y} + P_{f,o} \cdot n_{f,o}$
3. $n_{m,o} = P_{m,y} \cdot n_{m,y} + P_{m,o} \cdot n_{m,o}$
4. $n_{f,y} = F_{f,y} \cdot n_{f,y} \cdot (1-s) + F_{f,o} \cdot n_{f,o} \cdot (1-s)$
5. $n_{m,y} = F_{f,y} \cdot n_{f,y} \cdot s + F_{f,o} \cdot n_{f,o} \cdot s$
6. $V_{f,y} = P_{f,y} \cdot V_{f,o} + \frac{1}{2} \cdot F_{f,y} \cdot s \cdot V_{m,y} + \frac{1}{2} \cdot F_{f,y} \cdot (1-s) \cdot V_{f,y}$
7. $V_{f,o} = P_{f,o} \cdot V_{f,o} + \frac{1}{2} \cdot F_{f,o} \cdot s \cdot V_{m,y} + \frac{1}{2} \cdot F_{f,o} \cdot (1-s) \cdot V_{f,y}$
8. $V_{m,y} = P_{m,y} \cdot V_{m,o} + \frac{1}{2} \cdot F_{f,y} \cdot s \cdot V_{m,y} \cdot \frac{n_{f,y}}{n_{m,y}+n_{m,o}} + \frac{1}{2} \cdot F_{f,y} \cdot (1-s) \cdot V_{f,y} \cdot \frac{n_{f,y}}{n_{m,y}+n_{m,o}} + \frac{1}{2} \cdot F_{f,o} \cdot s \cdot V_{m,y} \cdot \frac{n_{f,o}}{n_{m,y}+n_{m,o}} + \frac{1}{2} \cdot F_{f,o} \cdot (1-s) \cdot V_{f,y} \cdot \frac{n_{f,o}}{n_{m,y}+n_{m,o}}$
9. $V_{m,o} = P_{m,o} \cdot V_{m,o} + \frac{1}{2} \cdot F_{f,y} \cdot s \cdot V_{m,y} \cdot \frac{n_{f,y}}{n_{m,y}+n_{m,o}} + \frac{1}{2} \cdot F_{f,y} \cdot (1-s) \cdot V_{f,y} \cdot \frac{n_{f,y}}{n_{m,y}+n_{m,o}} + \frac{1}{2} \cdot F_{f,o} \cdot s \cdot V_{m,y} \cdot \frac{n_{f,o}}{n_{m,y}+n_{m,o}} + \frac{1}{2} \cdot F_{f,o} \cdot (1-s) \cdot V_{f,y} \cdot \frac{n_{f,o}}{n_{m,y}+n_{m,o}}$

Where $n_{f,y}$ is the number of young females, $n_{m,y}$ is the number of young males, $n_{f,o}$ is the number of old females, $n_{m,o}$is the number of old males, s is the sex ratio (proportion of males), $P_{f,y}$ is the probability the a young females survives and becomes old, $P_{f,o}$ is the probability that an old female survives, $P_{m,y}$ is the probability the a young males survives and becomes old, $P_{m,o}$ is the probability that an old male survives $F_{f,y}$ is the average fecundity of a young female, and $F_{f,o}$ is the average fecundity of an old female.