Theor Appl Genet (2007) 115:933–944
DOI 10.1007/s00122-007-0619-9

ORIGINAL PAPER

Reliable confidence intervals in quantitative genetics: narrow-sense heritability

Thomas Fabbro · Anthony C. Davison · Thomas Steinger

Received: 21 December 2006 / Accepted: 20 July 2007 / Published online: 15 September 2007

Abstract Many quantitative genetic statistics are functions of variance components, for which a large number of replicates is needed for precise estimates and reliable measures of uncertainty, on which sound interpretation depends. Moreover, in large experiments the deaths of some individuals can occur, so methods for analysing such data need to be robust to missing values. We show how confidence intervals for narrow-sense heritability can be calculated in a nested full-sib/half-sib breeding design (males crossed with several females) in the presence of missing values. Simulations indicate that the method provides accurate results, and that estimator uncertainty is lowest for sampling designs with many males relative to the number of females per male, and with more females per male than progenies per female. Missing data generally had little influence on estimator accuracy, thus suggesting that the overall number of observations should be increased even if this results in unbalanced data. We also suggest the use of parametrically simulated data for prior investigation of the accuracy of planned experiments. Together with the proposed confidence intervals an informed decision on the optimal sampling design is possible, which allows efficient allocation of resources.

Communicated by C.-C. Schön.

T. Fabbro · T. Steinger
Department of Biology, University of Fribourg,
1700 Fribourg, Switzerland

A. C. Davison
Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

T. Fabbro (✉)
Unit of Evolutionary Biology, University of Basel, Vesalgasse 1,
4051 Basel, Switzerland
e-mail: thomas.fabbro@unibas.ch

Introduction

Quantitative genetics, one of the most promising frameworks for the unification of the fields of macroevolution and microevolution (Steppan et al. 2002), allows for the study of inheritance at the phenotypic level. The quantitative genetic approach is especially useful if genetic details are of intermediate importance and if random genetic drift, fluctuating adaptive landscapes, and genetic mechanisms do not have large impact. Then the patterns of genetic and phenotypic variation among individuals can be used to study the evolutionary origin or possible future trajectories of quantitative traits. If quantitative genetic parameters fluctuate over short periods of evolutionary time, meaningful predictions are difficult, though "no one expects genetic variances and covariances to remain unchanged for millennia" (Ayers and Arnold 1983). The rate of quantitative genetic parameter change has therefore to be compared to the rate of speciation, population differentiation, and changes in the adaptive landscape. Furthermore, quantitative genetic parameters might not only change over time but might also vary randomly among populations. Evolutionary arguments from one particular population would then not be valid for the species as a whole (Stearns 1982).

To investigate empirically changes in quantitative genetic parameters over time and the variation among populations we need to compare different parameter estimates. Therefore we need not only precise estimators but also reliable methods to assess their accuracy.

Assessment of estimator uncertainty is usually made using confidence intervals. Most quantitative genetic textbooks report only methods to estimate the variance of estimators (Roff 1997; Lynch and Walsh 1998), but these can only be transformed into statements about uncertainty, such as confidence intervals, if the distribution of the estimator is known, at least approximately (Davison 2003). Quantitative genetic parameters are often functions of variance components, for which the exact distributions of estimators are difficult to derive, especially in the presence of missing values. To achieve the same accuracy, variance estimation requires many more replicates than does the estimation of means, and our intuition concerning the accuracy of variances and related quantities can be strikingly wrong. This has consequences for the sampling design for experiments and especially for the sample size. Using parametrically simulated data sets we can determine a priori the expected accuracy of the estimator.

Standard analysis of variance methods (ANOVA) rely on the assumption that the available data are balanced, that is, the numbers of classes and subclasses do not vary. In empirical studies, however, death of some individuals during the experiment often occurs and so balance is more the exception than the rule. For variance estimation with unbalanced data, restricted maximum likelihood estimation (REML) or analysis of variance with unweighted sums of squares (ANOVAuw) are recommended (Searle et al. 1992; Burdick and Graybill 1992). However, analytical calculations of the influence of missing values on variance estimators and their uncertainty can easily become very tedious. Studying simulated data sets with randomly missing individuals provides important information on how missing values influence the uncertainty of REML and ANOVAuw estimators.

Here we investigate the accuracy of one of the most frequently estimated quantitative genetic statistics, heritability. The classic breeder's equation,

$$\mathsf{Response} = \mathsf{h}^2 \cdot \mathsf{Selection\ differential}, \tag{1}$$

relates the response, the across-generation change in trait mean, to heritability, h^2, times the selection differential, the within-generation change in trait mean. Heritability measures how efficiently a trait can respond to selection, and this is important for natural and artificial selection. It can be calculated as the ratio of the heritable variation divided by the phenotypic variation of a trait (Lynch and Walsh 1998).

We show how to assess the uncertainty of heritability estimators by calculating confidence intervals following a method developed by Sen et al. (1992), and we use simulation to evaluate the reliability of confidence intervals based on balanced and unbalanced data sets. Further we created an R package called qgen to facilitate parametric resampling and analysis of quantitative genetic data sets (Appendix).

Methods

Heritability can be estimated as the proportion of the total phenotypic variance attributable to heritable effects. In order to partition the observed phenotypic variance into heritable and non-heritable components, the relatedness of the individuals must be known, and this typically entails use of a mating design.

A common mating design is the North Carolina I, where males are crossed with several females to obtain full-sib families nested within paternal half-sib families. This allows the estimation of the component of variance due to additive genetic effects, separate from the components due to dominance and maternal effects, so estimates of heritability are not inflated by dominance or maternal effects. In this design all individuals are paternal half-sibs, full-sibs, or are unrelated. Therefore the variance of a trait can be partitioned into three components, referred to as male σ^2_{M}, female σ^2_{F}, and residual σ^2_{R} (in the animal breeding literature commonly called sire, dam, and residual). Statistical analysis of such twofold nested data must take all three variance components into account; methods for computation of confidence intervals involving just two variance components such as that of Harville and Fenech (1985) are not applicable here.

The three observable variance components are assumed to represent four independent underlying causal sources of variance (e.g. Falconer and Mackay 1996; Roff 1997; Lynch and Walsh 1998):

$$\sigma^2_{\mathsf{M}} \simeq \frac{\sigma^2_{\mathsf{Add}}}{4}, \tag{2}$$

$$\sigma^2_{\mathsf{F}} \simeq \frac{\sigma^2_{\mathsf{Add}}}{4} + \frac{\sigma^2_{\mathsf{Dom}}}{4} + \sigma^2_{\mathsf{Mat}}, \tag{3}$$

$$\sigma^2_{\mathsf{R}} \simeq \frac{\sigma^2_{\mathsf{Add}}}{2} + \frac{3\sigma^2_{\mathsf{Dom}}}{4} + \sigma^2_{\mathsf{Env}}, \tag{4}$$

with the subscript Add for additive genetic variance, Dom for dominance genetic variance, Mat for maternal variance, and Env for (micro-)environmental variance. All other possible sources of variance are assumed to be negligible (Lynch and Walsh 1998, p. 87). The male and female effects are assumed to arise from the joint action of a large number of genes, each with an individually small contribution to the phenotype, and therefore to be normally distributed.

Under the above assumptions, trait measurements taken on individuals from this particular mating design can be described by the model

$$E[y \mid \mathrm{m}, \mathrm{f}] = \mu + \mathrm{m} + \mathrm{f}, \tag{5}$$

M male (m = 1,2, ... , M), $\quad \mathrm{m} \sim \mathcal{N}(0, \sigma^2_{\mathsf{M}})$,
F female within male (f = 1,2, ... , F_m), $\quad \mathrm{f} \sim \mathcal{N}(0, \sigma^2_{\mathsf{F}})$,
P progeny within female (p = 1,2, ... , P_{mf}),

$$\mathsf{var}(y \mid \mathrm{m}, \mathrm{f}) = \sigma^2_{\mathsf{R}}.$$

Different methods can be used to partition the variance into the components for male, σ^2_{M}, female, σ^2_{F}, and residual, σ^2_{R}.

Heritability, h^2, is the proportion of the phenotypic variation attributed to heritable effects and can be estimated as

$$\widehat{\mathsf{h}^2} = \frac{4\widehat{\sigma}^2_{\mathsf{M}}}{\widehat{\sigma}^2_{\mathsf{M}} + \widehat{\sigma}^2_{\mathsf{F}} + \widehat{\sigma}^2_{\mathsf{R}}}. \tag{6}$$

Sen et al. (1992) developed a method to calculate two-sided (1–α) confidence limits for ratios of variance components from unbalanced twofold nested models. The lower confidence bound for heritability is

$$\widehat{\mathsf{L}}_{(1-\frac{\alpha}{2})} = \frac{4\left\{w_3\mathrm{MS}_{\mathsf{M}} - w_1\mathcal{F}_{\frac{\alpha}{2};\mathrm{df}_{\mathsf{M}},\mathrm{df}_{\mathsf{F}}}\mathrm{MS}_{\mathsf{F}} - (w_3 - w_1)\mathcal{F}_{\frac{\alpha}{2};\mathrm{df}_{\mathsf{M}},\mathrm{df}_{\mathsf{R}}}\mathrm{MS}_{\mathsf{R}}\right\}}{w_3\mathrm{MS}_{\mathsf{M}} - (w_1 - w_2)\mathcal{F}_{\frac{\alpha}{2};\mathrm{df}_{\mathsf{M}},\mathrm{df}_{\mathsf{F}}}\mathrm{MS}_{\mathsf{F}} - (w_3 - w_1 + w_2 - w_2w_3)\mathcal{F}_{\frac{\alpha}{2};\mathrm{df}_{\mathsf{M}},\mathrm{df}_{\mathsf{R}}}\mathrm{MS}_{\mathsf{R}}}, \tag{7}$$

where

$$w_1 = \frac{\sum_m (1/F_m)}{\sum_M (1/F_m \bar{\mathrm{P}}_{Hm})}, \quad w_2 = \frac{M}{\sum_m (1/F_m \bar{P}_{Hm})},$$
$$w_3 = \frac{\sum_m F_m - 1}{\sum_m \frac{F_m - 1}{\bar{\mathrm{P}}_{Hm}}},$$
$$\bar{\mathrm{P}}_{Hm} = \frac{F_m}{\sum_f (1/P_{mf})},$$
$$\mathrm{df}_{\mathsf{M}} = M - 1, \quad \mathrm{df}_{\mathsf{F}} = \sum_m F_m - M,$$
$$\mathrm{df}_{\mathsf{R}} = \sum_m \sum_f P_{mf} - \sum_m F_m,$$
$$\mathrm{MS}_{\mathsf{M}} = \widehat{\sigma}^2_{\mathsf{R}} + w_1\,\widehat{\sigma}^2_{\mathsf{F}} + w_2\,\widehat{\sigma}^2_{\mathsf{M}},$$
$$\mathrm{MS}_{\mathsf{F}} = \widehat{\sigma}^2_{\mathsf{R}} + w_3\,\widehat{\sigma}^2_{\mathsf{F}},$$
$$\mathrm{MS}_{\mathsf{R}} = \widehat{\sigma}^2_{\mathsf{R}}.$$

The quantity $\widehat{\mathsf{L}}_{(1-\frac{\alpha}{2})}$ is defined to be zero if $(w_3\mathrm{MS}_{\mathsf{M}})/(w_2\mathrm{MS}_{\mathsf{F}}) < \mathcal{F}_{\frac{\alpha}{2};\mathrm{df}_{\mathsf{M}},\mathrm{df}_{\mathsf{F}}}$. The upper confidence limit, $\widehat{\mathsf{U}}_{(1-\frac{\alpha}{2})}$, was estimated using equation (7) by replacing $\frac{\alpha}{2}$ with $1 - \frac{\alpha}{2}$ in the $\mathcal{F}$ quantile points. The lower $\widehat{\mathsf{L}}_{(1-\frac{\alpha}{2})}$ and upper $\widehat{\mathsf{U}}_{(1-\frac{\alpha}{2})}$ confidence limit together define the two-sided (1–α) confidence interval $\widehat{\mathsf{CI}}_{(1-\alpha)}$. For balanced data and under model (5) these confidence limits reduce to those of Graybill and Wang (1979).

We performed parametric simulations to evaluate the estimator for heritability and its confidence limits. The population value of h^2 is a function of the cumulative distribution function K_ψ, which is determined by the model given in equation (5) and a parameter set ψ. Equations (2)–(4) indicate how the causal parameters of the set ψ are related to the observable variance components σ^2_{M}, σ^2_{F} and σ^2_{R}. The sampling design $\mathcal{D}$ characterizes the individual samples drawn from the distribution K_ψ by indicating the number of replicates at the different and nested levels of the model. REML solutions for $\widehat{\sigma}^2_{\mathsf{M}}, \widehat{\sigma}^2_{\mathsf{F}}$, and $\widehat{\sigma}^2_{\mathsf{R}}$ were obtained with the `lme4`-package (Bates 2005) in `R` (`R` Development Core Team 2006) and used to calculate $\widehat{\mathsf{h}^2}$ according to equation (6). For comparison, variance components were also estimated by analysis of variance based on unweighted sums of squares, ANOVAuw (Burdick and Graybill 1992) because the confidence interval estimator in equation (7) was developed for this method. From R repetitions of the data simulation based on the distribution K_ψ and the sampling design $\mathcal{D}$ we obtained $\widehat{\mathsf{h}^2}^*_1, \ldots, \widehat{\mathsf{h}^2}^*_r, \ldots, \widehat{\mathsf{h}^2}^*_R$. The quality of the estimator $\widehat{\mathsf{h}^2}$ for a particular parameter set ψ and sampling design $\mathcal{D}$ was evaluated by estimating its bias and variance. The bias was estimated as

$$\mathsf{B}_R(\widehat{\mathsf{h}^2}) = \frac{1}{R}\sum_{r=1}^{R} \widehat{\mathsf{h}^2}^*_r - \mathsf{h}^2, \tag{8}$$

and the variance as

$$\mathsf{V}_R(\widehat{\mathsf{h}^2}) = \frac{1}{R-1}\sum_{r=1}^{R}\left(\widehat{\mathsf{h}^2}^*_r - \frac{1}{R}\sum_{r=1}^{R}\widehat{\mathsf{h}^2}^*_r\right). \tag{9}$$

The quality of the upper and lower confidence limits and the two-sided confidence interval (7) was described by the empirical error rate, EER, which is the proportion of the

estimated confidence intervals not containing the population parameter h^2,

$$\begin{aligned}
\mathrm{EER}_R(\widehat{\mathrm{L}}_{(1-\frac{\alpha}{2})}) &= \frac{1}{R}\sum_{r=1}^{R} I\left\{h^2 \le \widehat{\mathrm{L}}^*_{(1-\frac{\alpha}{2})r}\right\},\\
\mathrm{EER}_R(\widehat{\mathrm{U}}_{(1-\frac{\alpha}{2})}) &= \frac{1}{R}\sum_{r=1}^{R} I\left\{h^2 \ge \widehat{\mathrm{U}}^*_{(1-\frac{\alpha}{2})r}\right\},\\
\mathrm{EER}_R(\widehat{\mathrm{CI}}_{(1-\alpha)}) &= \mathrm{EER}_R(\widehat{\mathrm{L}}_{(1-\frac{\alpha}{2})}) + \mathrm{EER}_R(\widehat{\mathrm{U}}_{(1-\frac{\alpha}{2})}).
\end{aligned} \tag{10}$$

The indicator I{condition} is equal to 1 if the condition is true and 0 otherwise.

The evaluation was restricted to particular combinations of sampling designs $\mathcal{D}$ (Table 1) and parameter sets ψ (Table 2). A sampling design is described by the number of males, females, progenies, and missing values. The sampling designs were chosen to cover a large range of different male numbers, M. Although most empirical studies start with a balanced design, random loss of observations often leads to unbalanced data. To evaluate the effect of unbalancedness, in some sampling designs (Table 1) a fixed proportion of individuals was randomly deleted from each simulated data set. Therefore the proportion of missing males depends on the number of progenies per male and the proportion of missing females on the number of progenies per female. The sampling designs $\mathcal{D} = 1, \ldots, 8$ can be compared to $1m, \ldots, 8m$ to see the impact of 50% missing values on the estimator and its accuracy. The sampling designs $\mathcal{D} = 1m, \ldots, 8m$ can be compared to 5, ... , 12, with the same number of individuals, but in the second group without missing values and half the number of males in the first group. Within the group of $\mathcal{D} = 1, \ldots, 4$; 5, ..., 8; 9, ..., 12; and 13, ... , 16 all but the number of males was kept constant to study the impact of male number.

The parameter sets ψ = 1, ... , 15 were chosen to represent all possible combinations of a high and a low additive, dominance, maternal, and environmental variance component. This allows us to evaluate if the estimator for heritability itself or its confidence limits fail under certain conditions. The parameter sets ψ = 15, ... , 20 were chosen to evaluate the influence of the additive genetic variance on the estimator with all other parameters being constant.

The parameter set, ψ, can be replaced by empirical estimates from an experiment $\widehat{\psi}$ and the cumulative distribution function K_ψ replaced by $\mathrm{K}_{\widehat{\psi}}$. The same statistics can then be calculated to evaluate the quality of the estimator for these particular parameters.

Table 1 Sampling designs $\mathcal{D}$: Determined by the number of males, M the number of females per male, F and the number of progenies per male–female combination, P

Sampling design $\mathcal{D}$	Male number M	Female number F	Progeny number P	Missing (%)	Number of individuals
1	200	6	4	0	4,800
2	200	6	2	0	2,400
3	200	3	4	0	2,400
4	200	3	2	0	1,200
5	100	6	4	0	2,400
6	100	6	2	0	1,200
7	100	3	4	0	1,200
8	100	3	2	0	600
9	50	6	4	0	1,200
10	50	6	2	0	600
11	50	3	4	0	600
12	50	3	2	0	300
13	25	6	4	0	600
14	25	6	2	0	300
15	25	3	4	0	300
16	25	3	2	0	150
1 m	200	6	4	50	2,400
2 m	200	6	2	50	1,200
3 m	200	3	4	50	1,200
4 m	200	3	2	50	600
5 m	100	6	4	50	1,200
6 m	100	6	2	50	600
7 m	100	3	4	50	600
8 m	100	3	2	50	300

Missing indicates the percentage of randomly missing individuals. The last column shows the total number of individuals (sample size)

Table 2 Parameter sets, ψ, used for parametric simulations

Parameter set, ψ	Variance components				Heritability h^2
	Additive, σ^2_{Add}	Dominance, σ^2_{Dom}	Maternal, σ^2_{Mat}	Environmental, σ^2_{Env}	
1	100	100	100	100	0.25
2	10	100	100	100	0.032
3	100	10	100	100	0.323
4	100	100	10	100	0.323
5	100	100	100	10	0.323
6	10	10	100	100	0.045
7	10	100	10	100	0.045
8	10	100	100	10	0.045
9	100	10	10	100	0.455
10	100	10	100	10	0.455
11	100	100	10	10	0.455
12	10	10	10	100	0.077
13	10	10	100	10	0.077
14	10	100	10	10	0.077
15	100	10	10	10	0.769
16	33	100	100	100	0.1
17	129	100	100	100	0.3
18	300	100	100	100	0.5
19	700	100	100	100	0.7
20	2700	100	100	100	0.9

Determined by the additive genetic variance, the dominance genetic variance, the maternal variance, and the microenvironmental variance. The corresponding narrow-sense heritability is given in the last column

The parametric simulation of data sets as described in the previous section not only allows us to evaluate different estimators (e.g. $\widehat{h^2}$), but also to calculate the realized variation of an estimator in a given sample. After parametric simulations we know for every observation the size of the male, female and residual effect, and we can calculate the corresponding realized sampling variances $\tilde{\sigma}^2_M, \tilde{\sigma}^2_F$ and $\tilde{\sigma}^2_R$. The realized sampling variance of the male effect, $\tilde{\sigma}^2_M$, is calculated as

$$\tilde{\sigma}_M = \frac{1}{M-1}\sum_{m=1}^{M}(m_m - \bar{m})^2, \qquad (11)$$

and the realized sampling variances of the female and residual effects can be calculated in the same way. In empirical investigations only the phenotypic value can be observed and therefore the phenotypic variance has to be partitioned into $\widehat{\sigma}^2_M, \widehat{\sigma}^2_F$, and $\widehat{\sigma}^2_R$ as described for equation (5). For a given sample the realized heritability, $\tilde{h}^2$, can be calculated from the realized sampling variances, $\tilde{\sigma}^2_M, \tilde{\sigma}^2_F$, and $\tilde{\sigma}^2_R$, according to equation (6) in the same way as estimated heritability $\widehat{h^2}$ can be estimated from $\widehat{\sigma}^2_M, \widehat{\sigma}^2_F$, and $\widehat{\sigma}^2_R$. The variance of the realized heritability, $V(\tilde{h}^2)$, can be calculated from equation (9) by replacing $\widehat{h^2}$ by $\tilde{h}^2$. Whereas $V(\tilde{h}^2)$ represents only the variance due to the sampling, $V(\widehat{h^2})$ represents the variance due to sampling plus the variance due to estimation (variance partitioning). Thus comparing them allows us to assess what proportion of the estimated variance, $V(\widehat{h^2})$, is due to sampling and what proportion is due to estimation.

Results

The difference between the true unobservable value in a population and an estimate of it, the statistical error, can be partitioned into a systematic and a random component, the bias and the variance of an estimator. For all the investigated heritability estimates (Tables 3, 4, 5, 6), the square root of the variance $\sqrt{V_R(\widehat{h^2})}$ was considerably larger than the bias $B_R(\widehat{h^2})$, so the statistical error of heritability estimates is mainly due to the random component, as expected. Consequently, in empirical work we have primarily to decrease the variance of heritability, $V_R(\widehat{h^2})$, to minimise the error.

The variance of heritability estimates, $V_R(\widehat{h^2})$, was influenced by several factors. The comparison of different *sampling designs*, $\mathcal{D}$, showed that the number of males had

Table 3 Heritability, h^2, variance partitioning with REML

Design, Parameter set	Heritability				95% confidence interval, $\widehat{CI}_{95\%}$					
		Bias	$\sqrt{\text{Variance}}$		Expectation			Empirical error rate[a]		
$\mathcal{D}, \psi$	h^2	$B_R(h^2)$	$\sqrt{V_R(\widehat{h^2})}$	$\sqrt{V_R(\tilde{h}^2)}$	Lower $E_R(\hat{L})$	Upper $E_R(\hat{U})$	Length $E_R(\hat{U}-\hat{L})$	Lower $EER_R(\hat{L})$	Upper $EER_R(\hat{U})$	Two-sided $EER_R(\widehat{CI})$
1, 1	25.0	0.0	6.3	3.2	14.3	38.2	23.9	2.5	2.6	5.1
2, 1	25.0	0.0	7.1	3.2	12.6	40.4	27.8	2.6	2.8	5.3
3, 1	25.0	–0.3	10.5	3.2	7.1	47.0	40.0	1.8	2.9	4.7
4, 1	25.0	0.0	12.2	3.2	5.3	52.0	46.7	1.9	1.3	3.2
5, 1	25.0	0.0	8.4	3.2	10.7	44.6	33.9	2.1	2.5	4.6
6, 1	25.0	0.0	10.0	3.2	8.7	47.9	39.2	2.3	2.4	4.7
7, 1	25.0	0.2	13.8	3.2	4.3	57.9	53.6	2.1	0.0	2.1
8, 1	25.0	–0.3	16.7	3.2	3.1	63.7	60.6	2.0	0.0	2.0
9, 1	25.0	–0.4	11.8	4.5	6.8	54.4	47.6	2.1	2.3	4.4
10, 1	25.0	–0.1	13.8	4.5	5.4	59.5	54.1	2.2	0.2	2.4
11, 1	25.0	0.6	18.7	4.5	2.6	72.6	70.0	1.9	0.0	1.9
12, 1	25.0	1.6	21.7	4.5	2.1	80.1	77.9	2.0	0.0	2.0
13, 1	25.0	0.1	16.4	7.1	4.4	70.7	66.2	2.2	0.0	2.2
14, 1	25.0	0.5	19.0	7.1	3.5	75.9	72.4	2.5	0.0	2.5
15, 1	25.0	2.8	24.9	7.1	2.1	88.7	86.6	2.8	0.0	2.8
16, 1	25.0	3.7	27.7	7.1	1.6	94.2	92.6	2.0	0.0	2.0
1*m*, 1	25.0	–0.3	7.1	3.2	10.9	41.6	30.7	2.0	2.0	4.0
2*m*, 1	25.0	0.0	10.0	3.2	6.0	50.0	44.0	1.7	1.3	3.0
3*m*, 1	25.0	–0.1	13.0	3.2	3.6	56.3	52.7	1.5	0.0	1.5
4*m*, 1	25.0	1.5	18.4	3.2	1.3	76.3	75.0	0.9	0.0	0.9
5*m*, 1	25.0	–0.2	10.5	3.2	7.2	49.9	42.7	2.0	1.9	3.8
6*m*, 1	25.0	0.1	14.1	3.2	3.5	61.7	58.3	1.5	0.0	1.5
7*m*, 1	25.0	1.3	17.6	3.2	2.3	71.2	68.9	1.7	0.0	1.7
8*m*, 1	25.0	2.1	23.9	3.2	0.8	90.4	89.5	0.9	0.0	0.9

Characteristics of the estimator and its 95% confidence intervals based on parametrically simulated data sets, $R = 3{,}332$, according to sampling designs $\mathcal{D} = 1, \ldots, 8m$ (Table 1) and parameter set $\psi = 1$ (Table 2). The nominal error rate is $\alpha = 5\%$ for the two-sided confidence interval and $\alpha = 2.5\%$ for the lower and upper limits. All values are given as percentages

[a] Confidence intervals for empirical error rates are always shorter than $EER \pm 1.5$

a strong influence on $V_R(\widehat{h^2})$ (Tables 3, 4). For example doubling the number of males approximately halved $V_R(\widehat{h^2})$. In designs with a constant number of individuals and a constant number of males but a different number of females and progenies (e.g. $\mathcal{D} = 2, 3) V_R(\widehat{h^2})$ was smaller when the number of females was higher and the number of progenies low. The variance of realized heritability, $V_R(\tilde{h}^2)$, was for a given parameter set only determined by the number of males (Table 3, 4). Therefore, $V_R(\tilde{h}^2)$ was the same up to the third digit for sampling designs within the groups $\mathcal{D} = 1, \ldots, 4; 5, \ldots, 8; 9, \ldots, 12;$ and 13, ... , 16. As one would anticipate on general grounds, one can conclude that the number of males determined the variance of realized heritability, $V_R(\tilde{h}^2)$, whereas the number of replicates within a male determined the difference between variance of estimated heritability, $V_R(\widehat{h^2})$, and variance of realized heritability, $V_R(\tilde{h}^2)$. Therefore replications within a male can only improve estimation up to a certain point, beyond which the number of males becomes limiting.

The influence of the *parameter set*, ψ, on the variance of heritability was small compared to the influence of the design (Tables 5, 6). To compare the variance of heritability among different parameter sets one also needs to take into account that heritability can only take values between zero and one. Heritabilities that are close to zero or one can essentially only vary in one direction, making the variance less useful as a measure of uncertainty.

The influence of the *variance partitioning method* on the variance of estimated heritability, $V_R(\widehat{h^2})$, was rather small. As expected, for balanced data there was no

Table 4 Heritability, h^2, variance partitioning with ANOVA of unweighted sums of squares

Design, Parameter set	Heritability				95% confidence interval, $\widehat{CI}_{95\%}$					
		Bias	$\sqrt{\text{Variance}}$		Expectation			Empirical error rate[a]		
$\mathcal{D}, \psi$	h^2	$B_R(h^2)$	$\sqrt{V_R(\widehat{h^2})}$	$\sqrt{V_R(\tilde{h}^2)}$	Lower $E_R(\hat{L})$	Upper $E_R(\hat{U})$	Length $E_R(\hat{U}-\hat{L})$	Lower $EER_R(\hat{L})$	Upper $EER_R(\hat{U})$	Two-sided $EER_R(\widehat{CI})$
1, 1	25.0	0.0	6.3	3.2	14.3	38.2	23.9	2.5	2.6	5.1
2, 1	25.0	0.0	7.1	3.2	12.6	40.4	27.8	2.6	2.8	5.3
3, 1	25.0	–0.3	10.5	3.2	7.1	47.0	39.9	1.8	2.9	4.7
4, 1	25.0	0.0	12.2	3.2	5.3	51.9	46.6	1.9	2.2	4.1
5, 1	25.0	0.0	8.4	3.2	10.7	44.6	33.9	2.1	2.5	4.6
6, 1	25.0	0.0	10.0	3.2	8.7	47.9	39.2	2.3	2.4	4.7
7, 1	25.0	0.2	13.8	3.2	4.3	57.7	53.3	2.1	2.0	4.1
8, 1	25.0	–0.3	16.7	3.2	3.1	63.0	59.9	2.0	2.5	4.4
9, 1	25.0	–0.4	11.8	4.5	6.8	54.4	47.5	2.1	2.3	4.4
10, 1	25.0	–0.1	13.8	4.5	5.4	59.3	53.9	2.2	2.7	5.0
11, 1	25.0	0.6	18.7	4.5	2.6	71.3	68.7	1.9	2.5	4.4
12, 1	25.0	1.6	21.7	4.5	2.1	77.5	75.4	2.0	2.3	4.3
13, 1	25.0	0.1	16.4	7.1	4.4	70.0	65.6	2.2	2.9	5.1
14, 1	25.0	0.5	19.0	7.1	3.5	74.8	71.2	2.5	2.5	5.0
15, 1	25.0	2.8	24.9	7.1	2.1	84.8	82.7	2.8	2.2	5.0
16, 1	25.0	3.7	27.7	7.1	1.6	87.7	86.1	2.0	2.4	4.4
1*m*, 1	25.0	–1.9	7.7	3.2	9.6	39.9	30.3	1.6	3.7	5.3
2*m*, 1	25.0	–2.1	11.4	3.2	5.2	47.7	42.4	1.6	3.1	4.7
3*m*, 1	25.0	–6.3	13.4	3.2	2.0	49.3	47.3	0.6	5.8	6.4
4*m*, 1	25.0	–3.1	20.2	3.2	1.4	68.4	67.0	1.4	4.1	5.5
5*m*, 1	25.0	–1.9	11.0	3.2	6.3	48.0	41.7	1.5	4.1	5.6
6*m*, 1	25.0	–2.0	15.2	3.2	3.2	58.7	55.6	1.9	2.6	4.5
7*m*, 1	25.0	–4.1	17.9	3.2	1.6	63.8	62.1	1.1	4.1	5.2
8*m*, 1	25.0	–0.4	26.3	3.2	1.3	81.0	79.6	2.0	2.8	4.7

Characteristics of the estimator and its 95%-confidence intervals based on parametrically simulated data sets, R = 3,332, according to sampling designs $\mathcal{D} = 1, \ldots, 8m$ (Table 1) and parameter set ψ = 1 (Table 2). The nominal error rate is α = 5% for the two-sided confidence interval and α = 2.5% for the lower and upper limits. All values are given as percentages

[a] Confidence intervals for empirical error rates are always shorter than EER ± 1.5

difference and for highly unbalanced data (50% missing values, $\mathcal{D} = 1m, \ldots, 8m$) the estimates from restricted maximum likelihood (REML) variance partitioning tended to be slightly smaller than from unweighted analysis of variance (ANOVAuw) (Tables 3, 4).

The confidence interval estimators of of Sen et al. (1992) give very reliable results for heritability estimates, as shown by our parametric simulations (Tables 3, 4, 5, 6). Over a wide range of different sampling designs, $\mathcal{D} = 1, \ldots, 16, 1m, \ldots, 8m$, (Table 1) and a wide range of different parameter sets, ψ = 1, ... , 20, (Table 2) the empirical error rate was close to the nominal error rate, even for the strongly unbalanced data sets, $\mathcal{D} = 1m, \ldots, 8m$.

The method used to partition the variance components had large consequences for the confidence intervals. If REML methods were used to partition the variance, the upper confidence limits tended to be conservative (Tables 3, 4, 5) and hence also the two-sided empirical error rate was rather conservative. In contrast ANOVAuw estimators provided empirical error rates closer to the nominal error rates and the average length of the confidence intervals was shorter (Tables 4, 5, 6). This would suggest ANOVAuw to be superior to partition the variance, but a closer look at the distribution of confidence interval lengths showed the opposite (Figs. 1, 2). The median and the inter-quartile range, the length of the box, was almost identical for both methods, but the ANOVAuw method produced many confidence intervals that were much too short and therefore unreliable. Our simulations showed that the confidence intervals from REML solutions vary less. Especially for small sampling designs, the REML method should be preferred for partitioning the variance components, although the confidence intervals tended to

Table 5 Heritability, h^2, variance partitioning with REML

Design, Parameter set	Heritability				95% confidence interval, $\widehat{CI}_{95\%}$					
		Bias	$\sqrt{\text{Variance}}$		Expectation			Empirical error rate[a]		
$\mathcal{D}, \psi$	h^2	$B_R(h^2)$	$\sqrt{V_R(\widehat{h^2})}$	$\sqrt{V_R(\tilde{h^2})}$	Lower $E_R(\hat{L})$	Upper $E_R(\hat{U})$	Length $E_R(\hat{U}-\hat{L})$	Lower $EER_R(\hat{L})$	Upper $EER_R(\hat{U})$	Two-sided $EER_R(\widehat{CI})$
6, 1	25.0	0.0	10.0	3.2	8.7	47.9	39.2	2.3	2.4	4.7
6, 2	3.2	1.7	5.5	0.0	0.2	23.3	23.1	2.1	0.0	2.1
6, 3	32.3	0.1	10.5	4.5	14.0	57.2	43.3	2.1	1.7	3.8
6, 4	32.3	0.1	9.5	4.5	15.6	55.0	39.3	2.2	1.8	4.1
6, 5	32.3	0.2	11.0	4.5	13.4	58.0	44.6	2.2	1.7	3.9
6, 6	4.5	1.5	6.3	0.0	0.3	25.4	25.1	2.2	0.0	2.2
6, 7	4.5	1.0	5.5	0.0	0.3	21.3	20.9	2.8	0.0	2.8
6, 8	4.5	1.6	6.3	0.0	0.3	26.7	26.4	2.4	0.0	2.4
6, 9	45.5	–0.1	11.0	5.5	26.5	70.3	43.9	2.0	2.1	4.1
6, 10	45.5	–0.1	12.6	6.3	23.1	74.1	51.0	2.3	2.2	4.5
6, 11	45.5	0.1	11.0	6.3	25.8	71.6	45.8	1.9	2.0	3.9
6, 12	7.7	0.4	6.3	0.0	0.8	23.9	23.1	2.0	0.0	2.0
6, 13	7.7	1.2	8.4	0.0	0.6	32.6	32.1	2.5	0.0	2.5
6, 14	7.7	0.5	6.3	0.0	0.7	26.0	25.3	2.2	0.0	2.2
6, 15	76.9	–0.4	13.0	8.9	52.1	97.0	44.8	2.6	2.2	4.7
6, 16	10.0	0.3	7.7	0.0	1.2	29.9	28.7	2.5	0.0	2.5
6, 17	30.0	0.1	10.5	4.5	12.6	54.1	41.5	2.5	2.0	4.5
6, 18	50.0	–0.2	11.8	6.3	28.8	76.9	48.1	2.0	2.0	4.0
6, 19	70.0	–0.3	13.0	8.4	46.0	94.1	48.1	2.1	2.2	4.3
6, 20	90.0	–2.9	11.4	8.4	63.1	99.4	36.3	2.5	2.4	4.9

Characteristics of the estimator and its 95%-confidence intervals based on parametrically simulated data sets, R = 3,332, according to sampling designs $\mathcal{D} = 6$ (Table 1) and parameter set ψ = 1 (Table 2). The nominal error rate is α = 5% for the two-sided confidence interval and α = 2.5% for the lower and upper limits. All values are given as percentages

[a] Confidence intervals for empirical error rates are always shorter than EER ± 1.5

be slightly conservative compared to those from ANOVAuw.

As outlined in the introduction all methods for partitioning the variance have deficiencies if data are unbalanced. Because exact results on the influence of unbalanced data on the estimators of heritability and their confidence intervals are unavailable, parametric simulations are a valuable alternative. Our simulations showed that even for highly unbalanced data sets the resulting estimates are accurate. By comparing the sampling designs, $\mathcal{D} = 5, \ldots, 12$ to $\mathcal{D} = 1m, \ldots, 8m$, with the same number of individuals, unbalanced data did not provide less accurate estimates than balanced data. Considering the variance of estimated heritability and the length of the confidence interval, we found that the designs $\mathcal{D} = 1m, 3m, 5m$, and $7m$ with 50% missing values provided even more accurate estimates of heritability than sampling designs $\mathcal{D} = 5, 7, 9,$ and 11. The four progenies per female made these designs very robust to randomly missing values and the high number of males allowed an accurate estimation of heritability. Comparing designs with only two progenies per female, balanced data, $\mathcal{D} = 6, 8, 10, 12$, provided slightly better results than unbalanced data, $\mathcal{D} = 2m, 4m, 6m, 8m$.

Discussion

Our study has shown that heritability estimates are generally highly uncertain. Even in large experiments, 95%-confidence intervals for narrow-sense heritability cover large parts of the possible range between zero and one. For example, a sample size of more than 4,800 individuals is needed to estimate heritability with a 95%-confidence interval of length 0.25. Because our samples were generated under the assumptions of independent and normally distributed effects, empirical investigations may need even more replicates. If biological assumptions are violated, e.g. through selection during the experiment, further uncertainty will be added to the estimates. Our investigation furthermore showed that the high variability of heritability

Table 6 Heritability, h^2, variance partitioning with ANOVA based on unweighted sums of squares

Design, Parameter set	Heritability				95% confidence interval, $\widehat{CI}_{95\%}$					
		Bias	$\sqrt{\text{Variance}}$		Expectation			Empirical error rate[a]		
$\mathcal{D}, \psi$	h^2	$B_R(h^2)$	$\sqrt{V_R(\widehat{h^2})}$	$\sqrt{V_R(\tilde{h^2})}$	Lower $E_R(\hat{L})$	Upper $E_R(\hat{U})$	Length $E_R(\hat{U}-\hat{L})$	Lower $EER_R(\hat{L})$	Upper $EER_R(\hat{U})$	Two-sided $EER_R(\widehat{CI})$
6, 1	25.0	0.0	10.0	3.2	8.7	47.9	39.2	2.3	2.4	4.7
6, 2	3.2	1.7	5.5	0.0	0.2	21.4	21.2	2.1	1.8	3.9
6, 3	32.3	0.1	10.5	4.5	14.0	57.2	43.3	2.1	1.7	3.8
6, 4	32.3	0.1	9.5	4.5	15.6	55.0	39.3	2.2	1.8	4.1
6, 5	32.3	0.2	11.0	4.5	13.4	58.0	44.6	2.2	1.7	3.9
6, 6	4.5	1.5	6.3	0.0	0.3	23.8	23.5	2.2	1.8	4.0
6, 7	4.5	1.0	5.5	0.0	0.3	20.2	19.9	2.8	2.0	4.7
6, 8	4.5	1.6	6.3	0.0	0.3	25.0	24.7	2.4	2.0	4.4
6, 9	45.5	–0.1	11.0	5.5	26.5	70.3	43.9	2.0	2.1	4.1
6, 10	45.5	–0.1	12.6	6.3	23.1	74.1	51.0	2.3	2.2	4.5
6, 11	45.5	0.1	11.0	6.3	25.8	71.6	45.8	1.9	2.0	3.9
6, 12	7.7	0.4	6.3	0.0	0.8	23.5	22.7	2.0	2.5	4.4
6, 13	7.7	1.2	8.4	0.0	0.6	31.2	30.6	2.5	1.9	4.4
6, 14	7.7	0.5	6.3	0.0	0.7	25.3	24.6	2.2	2.5	4.7
6, 15	76.9	–0.4	13.0	8.9	52.1	97.0	44.8	2.6	2.2	4.7
6, 16	10.0	0.3	7.7	0.0	1.2	29.4	28.2	2.5	2.2	4.7
6, 17	30.0	0.1	10.5	4.5	12.6	54.1	41.5	2.5	2.0	4.5
6, 18	50.0	–0.2	11.8	6.3	28.8	76.9	48.1	2.0	2.0	4.0
6, 19	70.0	–0.3	13.0	8.4	46.0	94.1	48.1	2.1	2.2	4.3
6, 20	90.0	–2.9	11.4	8.4	63.1	99.4	36.3	2.5	2.4	4.9

Characteristics of the estimator and its 95% confidence intervals based on parametrically simulated data sets, $R = 3{,}332$, according to the design $\mathcal{D} = 6$ (Table 1) and parameter set $\psi = 1, \ldots, 20$ (Table 2). The nominal error rate is $\alpha = 5\%$ for the two-sided confidence interval and $\alpha = 2.5\%$ for the lower and upper limits. All values are given as percentages

[a] Confidence intervals for empirical error rates are always shorter than $EER \pm 1.5$

estimators is mainly the result of sampling and of estimating variance components. It is not caused by biological factors, and should not be confused with variability among populations or environments. Although it is well known that the estimation of the variance needs more replicates than the estimation of the arithmetic mean, the required increase in replication is often underestimated.

Given this large variation it is clear that reliable confidence intervals are needed to interpret heritability estimates. A confidence interval is usually called reliable if it covers the true, unobservable value in the population with the stated probability. Several methods for calculating confidence intervals for heritability estimators have been proposed. The sampling variance of broad-sense or family-mean heritability, both calculated from two variance components, has received considerable attention, especially for balanced data (Osborne and Paterson 1952; Knapp et al. 1985, 1989; Knapp and Bridges 1987; Koots and Gibson 1996; Visscher 1998; Burch and Harris 2005). For unbalanced data and under normality assumptions, Harville and Fenech (1985) developed a method to calculate exact confidence intervals on a ratio of two variance components, which allows one to give exact confidence intervals for broad-sense heritability. Graybill and Wang (1979) described a method for calculating confidence intervals for a ratio involving three variance components, based on balanced data, but it has been used only rarely to calculate confidence intervals for narrow-sense heritability (but see Collaku and Harrison 2005). Sen et al. (1992) extended the method of Graybill and Wang (1979) to unbalanced data, but their method has not been used for calculating confidence intervals for heritability. In this study we showed that the proposed confidence intervals for heritability are very reliable over a large range of biologically relevant combinations of parameters. Even for strongly unbalanced data, e.g. with 50% randomly generated missing values, the method presented provides reliable results, whereas other methods (e.g. Graybill and Wang 1979) fail completely (data not presented). This result is especially useful in practice because it shows that resources

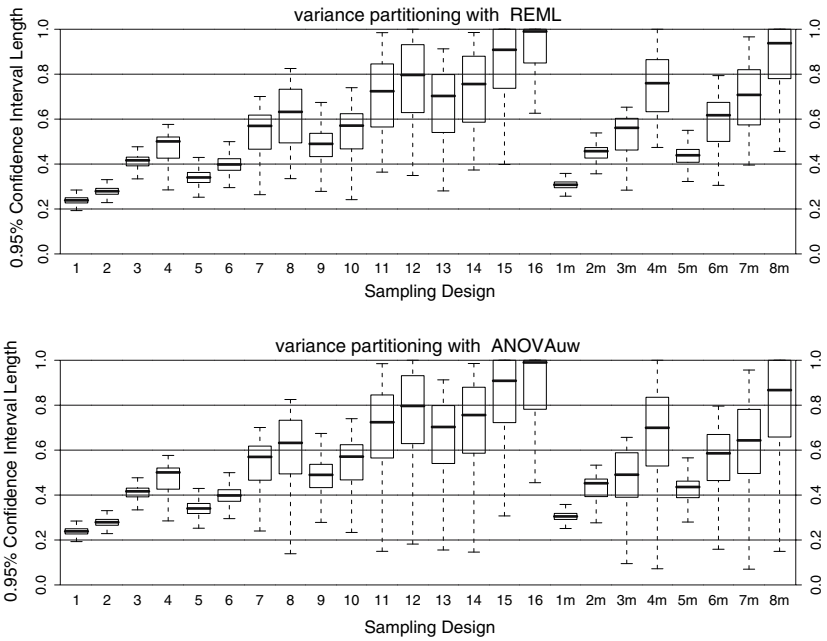

Fig. 1 Length of confidence intervals: The box plots show the 95%-confidence interval lengths of 3,332 parametrically simulated data sets according to design $\mathcal{D} = 1, \ldots, 8m$ (Table 1) and parameter set $\psi = 1$ (Table 2). Variance components were estimated using restricted maximum likelihood, REML (*above*), and analysis of variance with unweighted sums of squares, ANOVAuw (*below*) (The length of the box shows the interquartile range and the maximum length of each whisker is 1.5 times the interquartile range.)

should be allocated to increase the overall sample size instead of aiming to keep the overall survival of individuals high, and thereby the data balanced. Furthermore, parametric simulations using a planned sampling design together with a plausible parameter set allow us to optimize the design a priori by taking into account the expected rate of randomly dying individuals. This makes the range of application for the proposed method very broad, and it is clearly preferable to alternatives with rather restrictive assumptions. Parametric simulation allows us to find the optimal experimental design that directly minimises the variance of heritability. Therefore they are preferable to the oft-cited suggestions of Robertson (1959), which are based on the assumption that the variance of the female and male variance components should be minimised to the same extent. Our simulations for the given parameter set show that the number of males should be considerably larger than the number of females per male and progenies per female. Contrary to what is often observed in the literature, our simulations also showed that without missing values the number of females per male should always be larger than the number of progenies per female.

Unfortunately, empirical studies rarely report reliable confidence intervals for heritability, making it difficult to

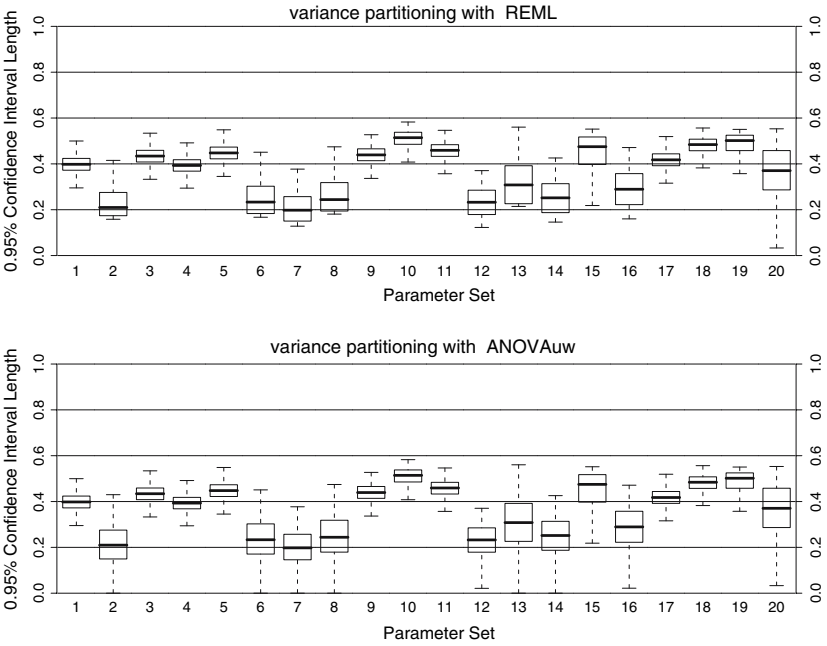

Fig. 2 Length of confidence intervals: The box plots show the 95%-confidence interval lengths of 3,332 parametrically simulated data sets according to design $\mathcal{D} = 6$ (Table 1) and parameter set $\psi = 1, \ldots, 20$ (Table 2). Variance components were estimated using restricted maximum likelihood, REML (*above*), and analysis of variance with unweighted sums of squares, ANOVAuw (*below*) (The length of the box shows the interquartile range and the maximum length of each whisker is 1.5 times the interquartile range.)

determine their accurateness and impossible to interpret their magnitude. In the quantitative genetic literature very often statistical tests are used to determine if the additive genetic variance is significantly larger than zero. If an estimate is "significant" it is interpreted without considering its uncertainty. Often studies are reported where heritability estimates are compared differing in values of only 0.2 but having confidence intervals longer than 0.9 (see e.g. studies cited by Mousseau and Roff 1987; Roff and Mousseau 1987). We obtained these confidence intervals using the method proposed here applied to the sampling design and parameters reported in the respective studies. This shows clearly the need to report confidence intervals in order to interpret the estimates.

Even larger numbers of replicates will be needed to estimate differences among environments or populations, or to estimate the rate at which quantitative genetic parameters evolve. However, to empirically investigate the validity of the quantitative approach, these are unfortunately the critical quantities.

In practice, the large number of replicates needed to get accurate estimates limits the use of quantitative genetic methods to organisms that can be bred at high numbers. Surprisingly the quantitative genetic approach is rarely directly criticised for inaccurate estimates (but see Mitchell-Olds and Rutledge 1986). Discussions are focused much more on the question whether the biological assumptions of the approach (e.g. no epistatic effects) are accurate (Barton and Turelli 1989; Roff 2003). Hence more and more complex breeding designs are applied that allow one to estimate additional parameters (Lynch and Walsh 1998; Wolf et al. 2000). From a biological point of view this is very promising. On the other hand, our study has shown that even for rather simple models the sample size for accurate estimation needs to be high. Thus, the general tendency in quantitative genetics to increase model complexity (e.g. from broad-sense heritability, to narrow-sense heritability, and to models with epistasis) makes sense only if the sample size of the experiments is also increased. Thus, for a given sample size, the balance between model complexity and the accuracy of the estimates must be of central importance.

We have shown that the presented method is not only a useful tool to calculate reliable confidence intervals for empirical data with missing values but is also valuable to determine the number of replicates in experimental designs. It will help to improve the accuracy of estimates and thus to decide on the appropriate degree of model complexity.

Acknowledgments We thank Sam Scheiner, Katharina Steinmann and two anonymous reviewers for useful comments on previous versions of this manuscript. TF and TS were supported by the grant 3100-067044 from the Swiss National Science Foundation, and TF by the Roche Research Foundation. This work was partially supported by the Swiss National Science Foundation in the context of the NCCR Plant Survival (`www2.unine.ch/nccr`).

Appendix

Quantitative Genetics in R: The R-package `qgen` is a collection of functions to analyse quantitative genetic data. It is especially helpful to perform parametric resampling of quantitative genetic data sets. Resampling allows *first* to determine a priori the expected variance of an estimator, *second* for a given empirical data set to calculate bootstrap confidence intervals, and *third* to evaluate different estimators and confidence intervals. The structure of the functions was kept very simple which easily allows users to extend it with functions that calculate the statistics of their interest. The organisation of the functions together with some examples is described in the documentation and help pages accompanying the package. The package is available at `http://www.r-project.org/`.

References

Ayers FA, Arnold SJ (1983) Behavioural variation in natural populations. IV. Mendelian models and heritability of a feeding response in the garter snake *Thamnophis elegans*. Heredity 51:405–413

Barton NH, Turelli M (1989) Evolutionary quantitative genetics: how little do we know? Annu Rev Genet 23:337–370

Bates DM (2005) Fitting linear mixed models in R. R News 5

Burch BD, Harris IR (2005) Optimal one-way random effects designs for the intraclass correlation based on confidence intervals. Commun Stat: Theory Methods 34:2009–2023

Burdick RK, Graybill FA (1992) Confidence intervals on variance components. Marcel Dekker, New York

Collaku A, Harrison SA (2005) Heritability of waterlogging tolerance in wheat. Crop Sci 45(2):722–727

Davison AC (2003) Statistical models. Cambridge University Press, Cambridge

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. Prentice-Hall, Englewood Cliffs

Graybill F, Wang C (1979) Confidence-intervals for proportions of variability in 2-factor nested variance component models. J Am Stat Assoc 74:368–374

Harville DA, Fenech AP (1985) Confidence intervals for a variance ratio, or for heritability, in an unbalanced mixed linear-model. Biometrics 41(1):137–152

Knapp SJ, Bridges WC Jr (1987) Confidence-interval estimators for heritability for several mating and experiment designs. Theor Appl Genet 73(5):759–763

Knapp SJ, Stroup WW, Ross WM (1985) Exact confidence-intervals for heritability on a progeny mean basis. Crop Sci 25(1):192–194

Knapp SJ, Bridges WC Jr, Yang MH (1989) Nonparametric confidence interval estimators for heritability and expected selection response. Genetics 121(4):891–898

Koots K, Gibson J (1996) Realized sampling variances of estimates of genetic parameters and the difference between genetic and phenotypic correlations. Genetics 143(3):1409–1416

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates

Mitchell-Olds T, Rutledge JJ (1986) Quantitative genetics in natural plant populations: a review of the theory. Am Nat 127(3):379–402

Mousseau TA, Roff DA (1987) Natural-selection and the heritability of fitness components. Heredity 59:181–197

Osborne R, Paterson WSB (1952) On the sampling variance of heritability estimates derived from variance analyses. Proc Roy Soc Edinb 64:456–461

Robertson A (1959) Experimental design in the evaluation of genetic parameters. Biometrics 15(2):219–216

Roff DA (1997) Evolutionary quantitative genetics. Chapman & Hall, London

Roff D (2003) Evolutionary quantitative genetics: are we in danger of throwing out the baby with the bathwater? Ann Zool Fenn 40(4):315–320

Roff DA, Mousseau TA (1987) Quantitative genetics and fitness—lessons from drosophila. Heredity 58:103–118

Searle SR, Casella G, McCulloch CE (1992) Variance components. Wiley, New York

Sen B, Graybill F, Ting N (1992) Confidence-intervals on ratios of variance components for the unbalanced 2-factor nested model. Biometical J 34:259–274

Stearns SC (1982) Components of fitness. Science (Wash.) 218(4871):463–464

Steppan SJ, Philipps PC, Houle D (2002) Comparative quantitative genetics: evolution of the G matrix. Trends Ecol Evol 17(7):320–327

R Development Core Team (2006) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Visscher P (1998) On the sampling variance of intraclass correlations and genetic correlations. Genetics 149(3):1605–1614

Wolf JB, Brodie ED, Walde MJ (2000) Epistasis and the evolutionary process. Oxford University Press, New York