

## ORIGINAL ARTICLE

## Comparing sire and dam estimates of heritability: jackknife and likelihood approaches

DA Roff

Department of Biology, University of California, Riverside, CA, USA

Three estimates of heritability are available from the half-sib pedigree design: the sire, dam and genotypic estimates. Because of its significantly smaller standard error, the genotypic estimate is preferred provided that there are no non-additive effects that inflate the estimate. I present two methods to test for such effects: these are a *t*-test of the paired sire and dam pseudovalues from the jackknife procedure and the likelihood ratio test from the animal model. Both methods are shown to be valid tests for significant dominance and/or maternal effects. SPLUS

coding for the implementation of the jackknife method is provided. Unless sample sizes are very large, the power of the tests is low and hence caution is advised in the use of the genotypic estimate following a nonsignificant test. An approximate power analysis can be done using the data from the jackknife method but the estimated power is typically a substantial underestimate of the true power and its use is not recommended.

*Heredity* (2008) 100, 32–38; doi:10.1038/sj.hdy.6801048; published online 5 September 2007

**Keywords:** animal model; half-sib; heritability; jackknife; maternal effect

## Introduction

One of the most common methods of estimating heritabilities or genetic correlations is the half-sib (also called the North Carolina 1) design in which each male is mated to several females (Roff, 1997; Lynch and Walsh, 1998). Variance components are estimated by a nested analysis of variance, giving three possible estimates of heritability, called the sire, dam and genotypic estimates (Becker, 1992). Epistatic effects are assumed to be negligible and maternal effects may be due to either nuclear or cytoplasmic sources (Roach and Wulff, 1987; Mousseau and Fox, 1998). The sire estimate is typically the preferred estimate as it is free from possible maternal and/or dominance variance (Kearsey and Pooni, 1996). The dam and genotypic estimates are potentially inflated by either of these sources of variance. Because its standard error is approximately half that of the sire or dam estimates, in the absence of these effects, the best estimates of additive genetic variance and heritability are the genotypic estimates (the mean of the sire and dam estimates).

Most papers estimating parameters by the half-sib design either present only the sire estimate or both the sire and dam estimate and then draw conclusions concerning the presence of maternal and/or dominance variance by a qualitative examination of the absolute difference between the two estimates (for example, p. 171, Falconer, 1989). Despite the fact that the genotypic estimate has a significantly smaller standard error, it is rarely used. A primary reason for this reluctance is the problem of deciding if the dam estimate is confounded

by maternal and/or dominance components, or if the sire and dam estimates are significantly different. In the present paper, I present two methods for statistically testing for a difference between the sire and dam estimate and use simulation to determine both the type I error rate for the tests and their power. The methods I consider are the jackknife using the variance components from a nested analysis of variance and the likelihood ratio test utilizing the animal model.

## Methods

## The jackknife method for testing between sire and dam estimates

The logic and methodology of the jackknife are explained in detail in Roff (2006) and I present here only the method of application. Because the method is the same for either variances, covariances, heritabilities or correlations, I describe the application in regard to the heritability. The procedure is as follows:

- (1) Estimate the sire and dam heritabilities, say  $\hat{h}_S^2$  and  $\hat{h}_D^2$ , respectively, by nested analysis of variance or restricted maximum likelihood,
- (2) Delete a sire group from the data set and recalculate the two components, say  $\hat{h}_{S-1}^2$  and  $\hat{h}_{D-1}^2$ , respectively
- (3) Construct the sire and dam pseudovalues ( $S_{S1}$ ,  $S_{D1}$ , respectively),  $S_{S1} = n\hat{h}_S^2 - (n-1)\hat{h}_{S-1}^2$  and  $S_{D1} = n\hat{h}_D^2 - (n-1)\hat{h}_{D-1}^2$  where  $n$  is the number of sire families.
- (4) Re-enter the deleted sire family, delete the next family and calculate the next pair of pseudovalues. Repeat this process, deleting each sire family once to produce  $n$  pseudovalues.
- (5) The distribution of pseudovalues should be the same as that of the heritabilities and hence the sire and

Correspondence: Professor DA Roff, Department of Biology, University of California, Riverside, CA 92521, USA.

E-mail: derek.roff@ucr.edu

Received 4 June 2007; revised 11 July 2007; accepted 13 July 2007; published online 5 September 2007

dam estimates can be compared using a paired *t*-test of the pseudovalues (see Appendix A for SPLUS coding).

Because of its smaller standard error, if no significant difference is found between the two estimates, one might use the genotypic estimate. However, before doing so, it is important to consider the power of the test to detect a biologically meaningful difference. Power analysis of a *t*-test is straightforward and an approximate power analysis can be carried out using the estimated standard error from the paired *t*-test (provided there are sufficient sire families to give a reasonable estimate of the standard error). However, the estimation of power will be an underestimate since a nonsignificant test, given that there really is a difference, will occur only when the standard error is unusually large relative to the mean. Thus, it is important to determine if the estimation of power is seriously biased downwards.

#### The likelihood ratio approach using the animal model

In contrast to the approach employed in the foregoing section of determining variance components from variation among families (sires and dams), the animal model uses the individual as the unit of estimation. A detailed description of the animal model and its application is given by Kruuk (2004). In the present context, maternal and/or dominance variance is estimated by including Dam as a random term in the model. The significance of this term is then estimated from the log-likelihood for the model without the Dam term (say  $LL_{NoDam}$ ) with the likelihood of the model including the Dam term (say  $LL_{Dam}$ ):  $2(LL_{Dam} - LL_{NoDam})$  is distributed as  $\chi^2$  with 1 d.f.

#### Statistical tests and programs used

Because non-additive effects increase the variance estimated by the Dam component, the test for a dam effect is one-tailed. The jackknife method was implemented using Fortran, which permitted a relatively large number of iterations (for example, 11 000 for the type 1 error simulations) for each set of parameter combinations. Variance components for the animal model were estimated using the SPLUS add-on of ASReml (Butler *et al.*, 2007). Due to logistical constraints, the number of iterations per parameter set was set at 1000 for the animal model. Analysis using the animal model was restricted to variance components, whereas the jackknife approach permitted the analysis of both variances and heritabilities.

#### Description of simulation model

Individual trait values were simulated by the equation

$$X_{ijk} = \left( \frac{Y_i + Z_{ij}}{2} \right) + W_{ijk} + e_{ijk} + M_{ij} \quad (1)$$

where  $X_{ijk}$  is the trait value for the *k*th individual from dam family *j* and sire family *i*;  $Y_i$  is a random normal variate unique to sire family *i*,  $N(0, V_A)$ , with mean 0 and additive genetic variance,  $V_A$ ;  $Z_{ij}$  is a random normal variate,  $N(0, V_A)$ , unique to dam family *j*;  $W_{ijk}$  is a random normal variate,  $N(0, V_A/2)$ , unique to each individual;  $e_{ijk}$  is the environmental variance distributed as  $N(0, 1)$ ; and  $M_{ij}$  is the maternal and/or dominance component (for simplicity, I shall refer to this compo-

nents as the 'maternal' effect or component), distributed as  $N(0, V_M)$ . The additive genetic variance,  $V_A$ , was determined by setting the environmental variance,  $V_E$ , equal to 1 and using the relationship  $V_A = h^2/(1-h^2)$ , where  $h^2$  is the heritability in the absence of maternal effects. Maternal genetic variance was set as a fraction of the additive genetic variance,  $V_M = MV_A$ .

The sire,  $h_S^2$ , and dam,  $h_D^2$  heritabilities are

$$\begin{aligned} h_S^2 &= V_A/(1 + V_A + V_M) \\ h_D^2 &= (V_A + 4V_M)/(1 + V_A + V_M) \end{aligned} \quad (2)$$

Using the relationships between  $h^2$  and  $V_A$  and between  $V_M$  and  $V_A$ , it can be shown that the ratio of the dam to sire heritability estimates is  $1 + 4M$ .

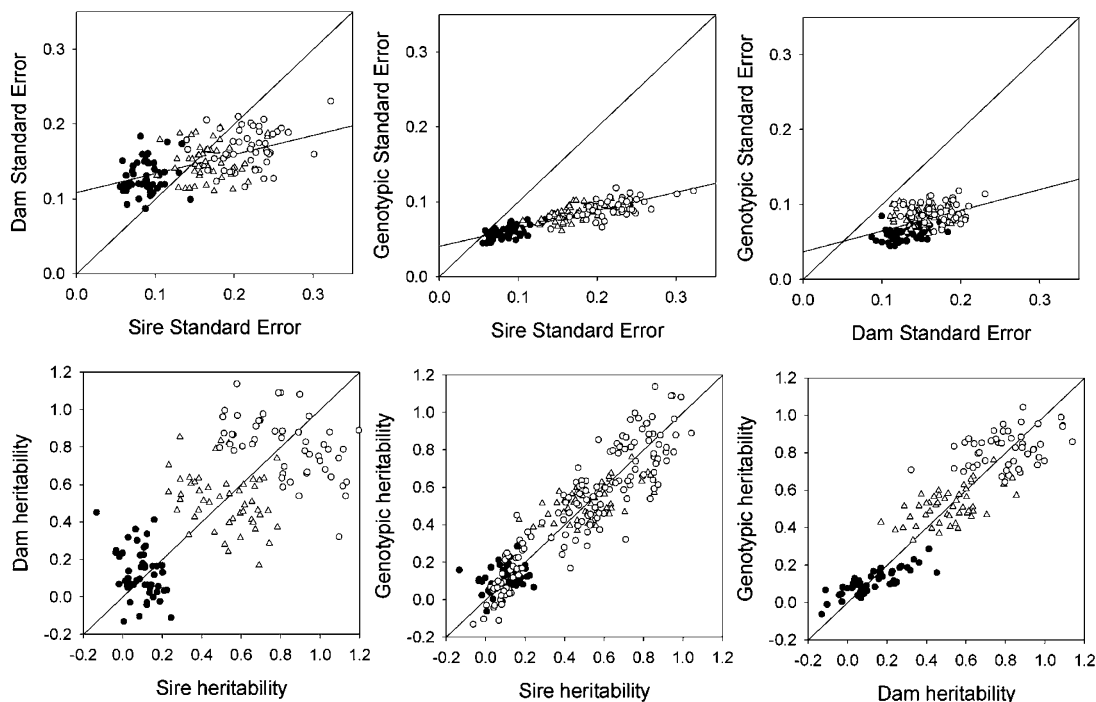
Simulations were run with 3 dams per sire, 5 offspring per dam and 20, 50 or 100 sires. Heritabilities were set at 0.1, 0.5 or 0.8 in the absence of maternal effects. As noted above, maternal genetic variance was set as a fraction of the additive genetic variance: 0, 0.125, 0.25 and 0.5, which corresponds to a dam heritability estimate,  $h_D^2$  in relation to the sire estimate,  $h_S^2$  of  $h_D^2 = h_S^2, 1.5h_S^2, 2h_S^2$  or  $3h_S^2$ , respectively. The inclusion of maternal effects reduced the sire heritabilities at the three values of maternal effects to 0.099, 0.471 and 0.727. In the case of no maternal effects, the simulations determine if the type 1 error rate is the required 5%, while the other cases are used to examine the power of the test to detect a maternal effect and the efficacy of estimating power given a nonsignificant test.

## Results

Even with the considerable sample size of 750 individuals (50 sires, 3 dams/sire, 5 offspring/dam) and no maternal effect, there is substantial discrepancy between the sire and dam estimates of heritability (Figure 1). At low heritability (0.1) the sire estimate of the standard error is smaller than the dam, but the situation reverses as heritability increases, the two being approximately the same at 0.5 and reversed at 0.8 (Figure 1). The genotypic estimates of heritability are consistently close to the correct value and the standard errors are substantially reduced illustrating the merit of using the genotypic estimate when non-additive effects can be discounted.

#### No maternal effects (type 1 error rate)

Both the jackknife method and the likelihood ratio test produced type I error rates close to the standard 5%. At heritabilities less than 0.8, the jackknife method significantly underestimates the type 1 error rate for both heritabilities and additive genetic variances (Table 1). While the absolute difference is small relative to the required 0.05 (0.03–0.04 rather than 0.05), it is statistically significant. For the largest heritability (0.8) the jackknife method produces a type 1 error rate for the test between sire and dam heritabilities that is not significantly different from the required 5%, whereas the jackknife method for the additive genetic variances slightly, but significantly, overestimates the type 1 error rate (Table 1). The likelihood ratio test gives similar results to the jackknife method (Table 1), although the smaller number of replications (1000) resulted in a slightly more erratic



**Figure 1** Comparison of the three types of heritability estimates and their standard errors for sample sizes of 50 sires, 3 dams/sire and 5 offspring per dam (50 replicates per combination). Heritabilities set at 0.1 (closed circle), 0.5 (open triangle) and 0.8 (open circle).

**Table 1** Effect of varying the number of sires ( $N$ ) and heritability ( $h^2$ ) on the type 1 error rates for a half-sib pedigree design

N	$h^2$	ANOVA <sup>a</sup>		Animal model
		$h^2$	$V_A$	$V_A$ (ANOVA $h^2$ , ANOVA $V_A$ ) <sup>b</sup>
20	0.1	<b>0.0321</b>	<b>0.0300</b>	<b>0.0222 (0.038, 0.038)</b>
50	0.1	<b>0.0375</b>	<b>0.0365</b>	<b>0.0374 (0.036, 0.035)</b>
100	0.1	<b>0.0385</b>	<b>0.0380</b>	0.0572 (0.051, 0.051)
20	0.5	<b>0.0448</b>	<b>0.0419</b>	0.05421 (0.047, 0.047)
50	0.5	<b>0.0444</b>	<b>0.0448</b>	0.0506 (0.042, 0.045)
100	0.5	<b>0.0473</b>	0.0484	<b>0.0667 (0.050, 0.051)</b>
20	0.8	0.0515	0.0524	0.0566 (0.051, 0.048)
50	0.8	0.0496	<b>0.0537</b>	<b>0.0210 (0.035, 0.040)</b>
100	0.8	0.0511	<b>0.0553</b>	0.0587 (0.056, 0.060)

Abbreviation: ANOVA, analysis of variance.

Difference between the sire and dam estimates tested using the jackknife method (ANOVA) or the animal model.

<sup>a</sup>Estimates based on 11 000 replicates. Critical upper and lower values of  $P$  (not corrected for multiple tests) are 0.0470 and 0.0530. Entries in bold font are significantly different from 0.05.

<sup>b</sup>Estimates based on 1000 replicates. Values for the jackknife methods based on only the 1000 replicates used for the Animal model are shown in parentheses. Critical upper and lower values of  $P$  (not corrected for multiple tests) are 0.0402 and 0.0598.

set of estimates, as is also evident in the jackknife estimates based on the same 1000 replications.

#### Maternal effects present (power analysis)

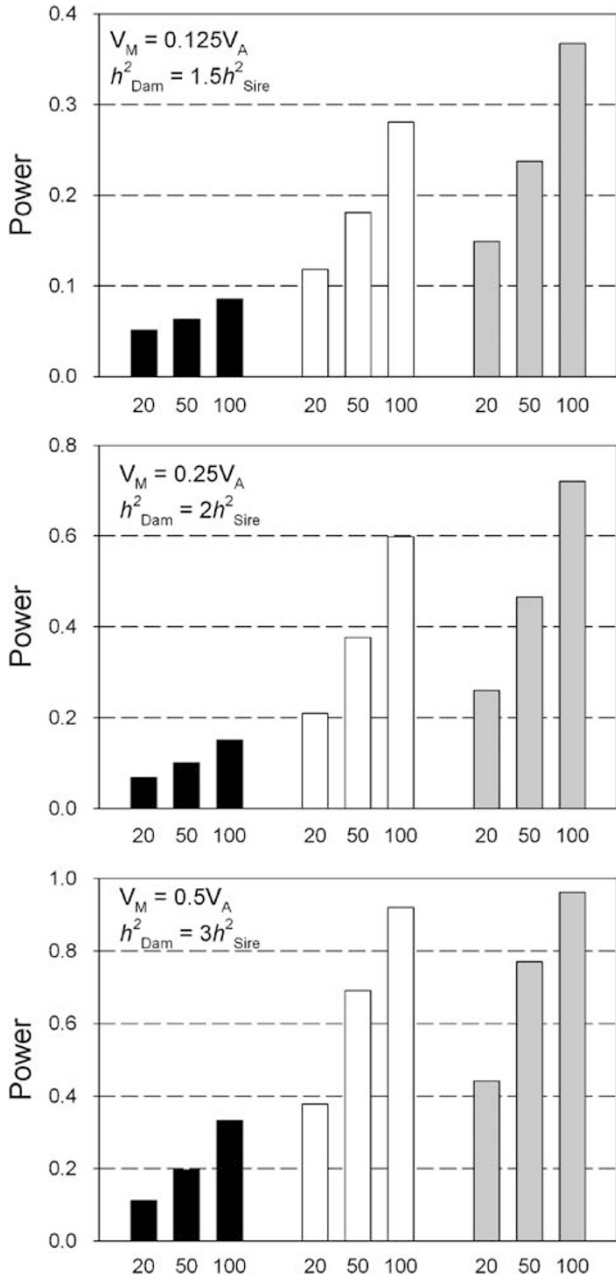
Analysis using the jackknife shows that, not surprisingly, power increases with the size of the maternal variance relative to the additive genetic variance, sample size and heritability (Figure 2: each estimate was made from 10 000 replications). High heritabilities and large sample sizes are required to raise the power above 60%: only in

the case of the largest maternal variance ( $V_M=0.5V_A$ ), the largest number of sires (100) and medium to large heritabilities (0.471, 0.727) does the power exceed 80% (Figure 1). To examine the likelihood test, I ran 1000 replications at  $V_M=0.125V_A$ ,  $h^2=0.5$  ( $=0.471$  with maternal effects) and number of sires at all three sizes ( $N=20, 50, 100$ ). The results agree well with those from the jackknife ( $N=20$ ,  $P_{\text{Jack}}=0.11$ ,  $P_{\text{Like}}=0.09$ ;  $N=50$ ,  $P_{\text{Jack}}=0.19$ ,  $P_{\text{Like}}=0.17$ ;  $N=100$ ,  $P_{\text{Jack}}=0.29$ ,  $P_{\text{Like}}=0.22$ ).

In each case in which a nonsignificant result was obtained, I estimated the power to detect the true difference using the standard error estimated from the pseudovalues: except when the heritability or sample size is low, the estimated power is far below the true value (Figure 3). These results suggest that a *post hoc* power analysis is unlikely to be informative.

The low power of the tests (Figure 2) indicates that if a dam effect is not detected caution should be exercised in assuming no effect and using the genotypic estimate. To examine the consequences of this action, we need to consider the error incurred (Figure 4). The absolute percentage error between the real heritability value and the estimated value is very similar for the sire and genotypic estimates when maternal effects are relatively low ( $0.125V_A$ , triangles in Figure 3), is generally higher for the genotypic estimate for the intermediate maternal effect ( $0.25V_A$ , circles in Figure 3) and is consistently higher for the highest maternal effect ( $0.5V_A$ , squares in Figure 3).

Given that there is a maternal effect but that the statistical tests (jackknife or likelihood) are not significant, what is the probability that the true value is not included in the confidence regions of the sire or genotypic estimates? Confidence regions were estimated

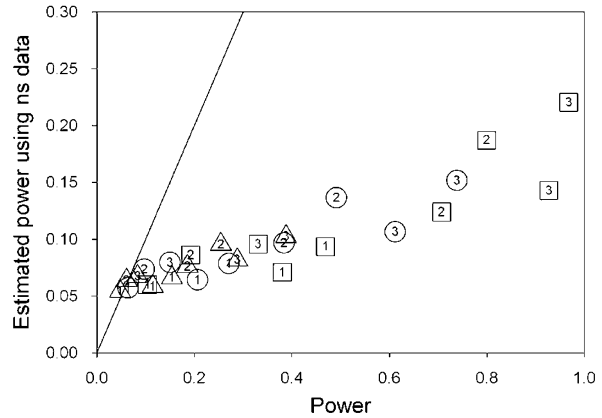


**Figure 2** Power analyses for various combinations of sire number ( $x$  axis), heritability (black bars=0.099, open bars=0.471, grey bars=0.727) and maternal variance (different panels). Results for each combination based on 10 000 replicates.

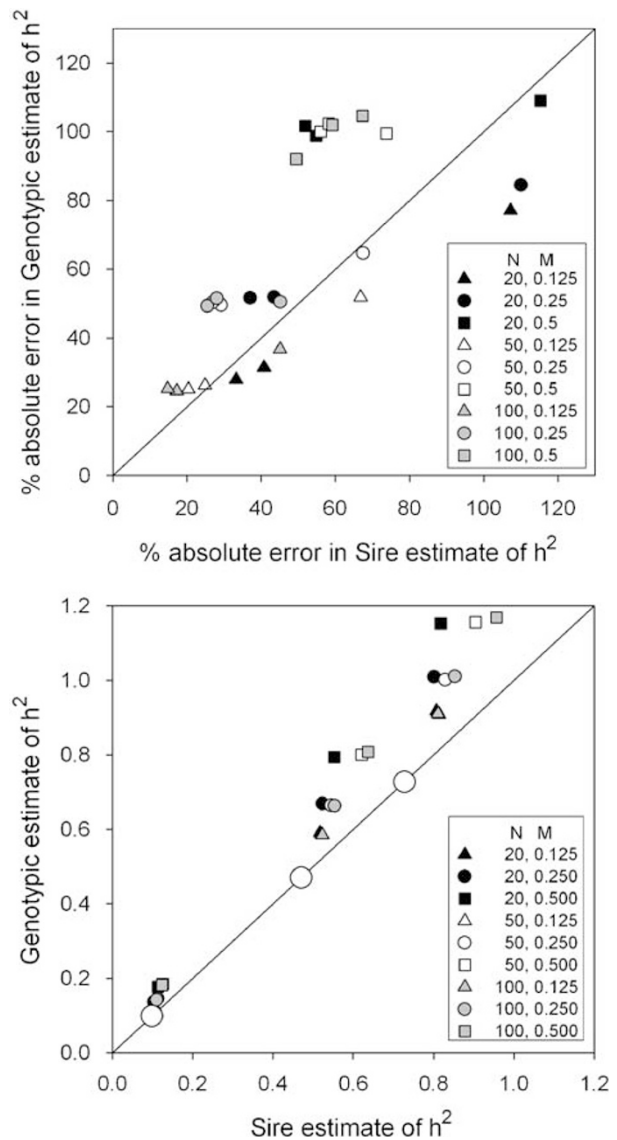
as  $\pm 2$  s.e. The confidence interval estimated using the sire component excludes the true value relatively few times (2–3%) except when the maternal effect is very large ( $M=0.5$ ), the number of sires is large (100) and the heritability is moderate to large (0.471, 0.727; Figure 5). On the other hand, the confidence limits estimated using the genotypic component has a high probability of excluding the true value (Figure 5).

### Discussion

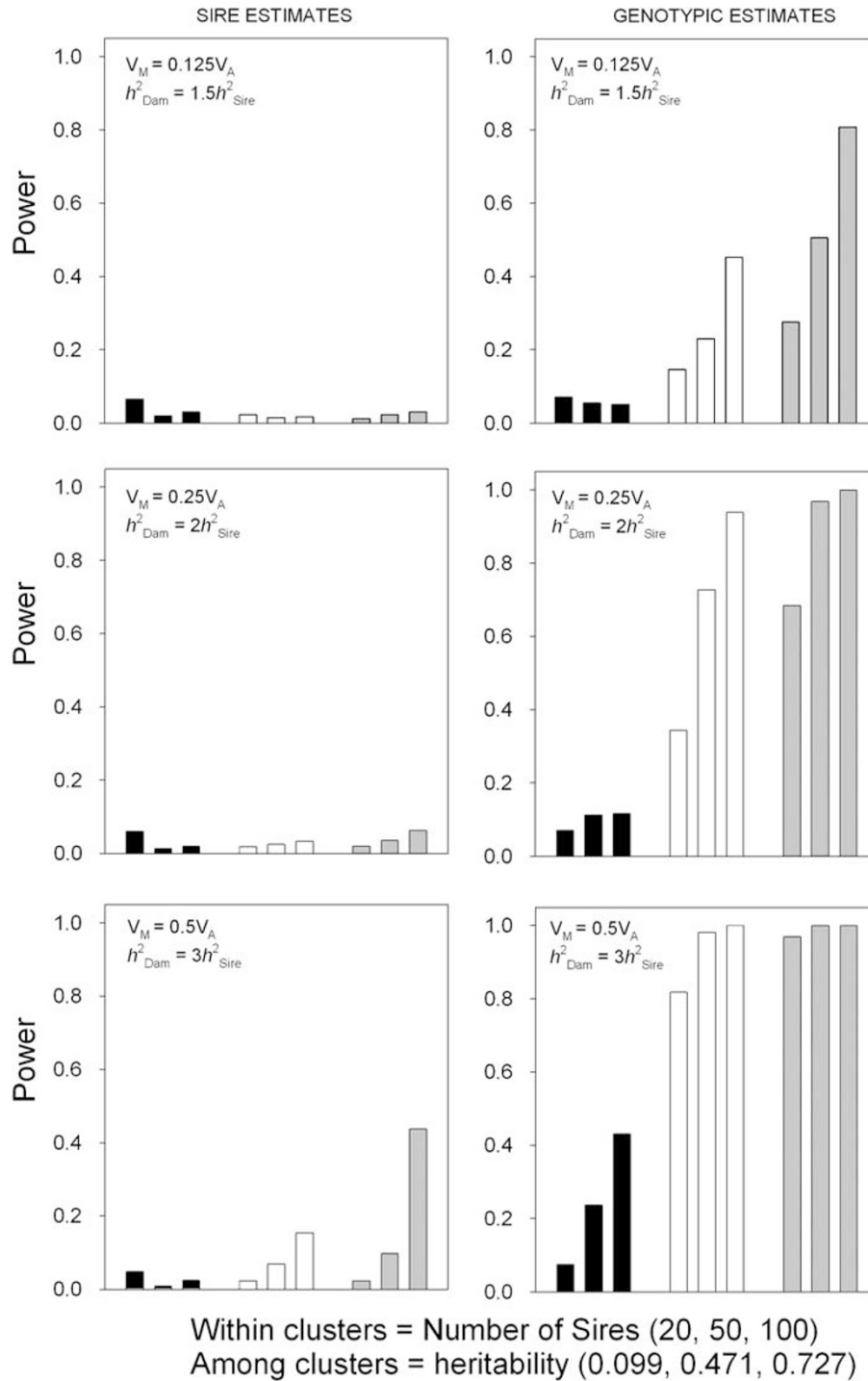
Both the jackknife and likelihood approaches are appropriate for testing the difference between sire and dam



**Figure 3** A comparison of the estimated power following a nonsignificant test with the true power. Numerals indicate heritabilities (0.099, 0.471, 0.727, respectively) and symbols indicate the maternal variance as a fraction of the additive genetic variance (triangle=0.125, circle=0.25, square=0.5). Solid line shows 1:1 relationship.



**Figure 4** A comparison of the sire and genotypic estimates of heritability when there is a maternal effect. Results based on 500 replicates/combination.  $N$  = Number of sires,  $M = V_A/V_M$ .



**Figure 5** A comparison of the proportion of times the confidence limits fails to include the true heritability value when there is a maternal effect. Results based on 500 replicates/combination.

estimates in a half-sib pedigree design. At the lowest heritability (0.1) used the type 1 error rate is somewhat underestimated (0.03–0.04), meaning false positives occur in less than the usual 5% of cases, but not enough to invalidate the test. The type 1 error rate for the test for the difference in variance components appears to be slightly overestimated at the highest heritability used (0.055 instead of the required 0.05).

Given that the genotypic estimate of heritability has a less erratic behaviour and is more precise with a

markedly lower standard error than either the sire or dam estimate (Figure 1), the above results suggest that this estimate should be used following a test for a difference between the sire and dam estimates. However, because of the relatively low power of the test (Figure 2), conclusions assuming the lack of non-additive effects should be made with caution. The low power, unless sample sizes are very large, is consistent with power analyses of tests for differences in genetic parameters between populations (Shaw, 1991).

The jackknife method has two potential advantages over the animal model: first, because it is a simple paired *t*-test it is possible to do at least some rough calculations on the power of the test should no significant difference be found between sire and dam estimates. Unfortunately, estimates of power in the case of nonsignificant results considerably underestimate the true power when sample sizes and/or heritability are large (Figure 3). The second potential advantage of the jackknife method is that it is readily programmed in SPLUS (see Appendix A), R or SAS, whereas the animal model is extremely difficult and will typically require a dedicated package such as ASREML, which is not inexpensive (of course the animal model can be used for a wide range of pedigree structures and hence is more flexible than the nested analysis of variance approach).

I suggest that all three heritability estimates should be reported and either the jackknife or likelihood method be used to test for a difference between the sire and dam estimate. If the difference is significant, one has evidence of non-additive effects and the sire estimate is preferred. If the sire estimate exceeds the dam estimate, then one can use the genotypic estimate because non-additive effects will *increase* the dam estimate not the sire estimate. If the sire estimate is not significantly less than the dam estimate, the course of action is less clear due to the low power of these tests. The choice will often depend upon whether the researcher considers it likely that non-additive effects are present. For example, based on previous studies, it would be extremely unwise to assume the absence of maternal effects in propagule size or early growth rate, but reasonable to assume little effect on adult morphological structures (see reviews in Mousseau and Fox, 1998).

## Appendix A

SPLUS coding for the analysis of a half-sib data set, with a test of the difference between sire and dam estimates

```
# The program assumes that
# 1) the data are in a file called Data (this can be anything if changed in
#   main program)
# 2) the Sire coding is SIRE
# 3) the Dam coding is DAM
# 4) The trait is coded as Trait
# DAM is coded as 1,2,3: 1,2,3 etc
#####Function to calculate heritability#####
H2.Estimator <- function(DATA)
{
# Estimate Variance components
  Var1 <- varcomp(formula=Trait~SIRE/DAM, data=DATA, method="reml")
# Estimate Heritability
  h2.Sire <- 4*Var1$variances[1]/sum(Var1$variances)
  h2.Dam <- 4*Var1$variances[2]/sum(Var1$variances)
  h2.SD <- 2*(Var1$variances[1]+Var1$variances[2])/sum(Var1$variances)
# Output Heritability
  H2 <- c(Prop,h2.Sire,h2.Dam, h2.SD)
  return(H2)
}
#####End of Function#####
```

## Acknowledgements

This research was supported by a NSF grant DEB-0445140. I am very grateful to Elizabeth King for her insightful comments on a previous version of this paper.

## References

- Becker WA (1992). *Manual of Quantitative Genetics*. Academic Enterprises: Pulman, WA.
- Butler D, Cullis BR, Gilmour AR, Gogel BJ (2007). *ASReml-S Reference Manual*. Dept of Primary Industries and Fisheries: Brisbane.
- Falconer DS (1989). *Introduction to Quantitative Genetics*. Longmans: New York.
- Kearsey MJ, Pooni HS (1996). *The Genetical Analysis of Quantitative Traits*. Chapman and Hall: London.
- Kruuk LEB (2004). Estimating genetic parameters in natural populations using the 'animal model'. *Phil Trans Roy Soc B: Biol Sci* **359**: 873–890.
- Lynch M, Walsh B (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates: Sunderland, MA.
- Mousseau T, Fox CW (eds). *Maternal Effects as Adaptations*. Oxford University Press: New York, 1998.
- Roach DA, Wulff RD (1987). Maternal effects in plants. *Ann Rev Ecol Syst* **18**: 209–235.
- Roff DA (1997). *Evolutionary Quantitative Genetics*. Chapman and Hall: New York.
- Roff DA (2006). *Introduction to Computer-Intensive Methods of Data Analysis in Biology*. Cambridge University Press: Cambridge.
- Shaw RG (1991). The comparison of quantitative genetic parameters between populations. *Evolution* **45**: 143–151.

```
#####MAIN PROGRAM#####
# The main program will contain coding to construct the relevant data set if
# not already available. I here assume that it has been constructed.
# I assume that the file is called Data with at least columns labelled
# SIRE, DAM, Trait (order not important)
# Set up matrix for storage of pseudovalues
# Make sure SIRE and DAM are random factors
  SIRE      <- factor(Data$Sire)
  DAM      <- factor(Data$Dam)
  is.random(Data$SIRE) <- T
  is.random(Data$DAM)  <- T
# Call function and save observed heritabilities
  H2.OBS    <- H2.Estimator(Data)
# Do nested anova to test for SIRE and DAM effects
  Model.ANOVA <- aov(Trait~SIRE/DAM, data=Data)
  A          <- summary(Model.ANOVA, ssType=3)
# Calculate proper F coefficient for SIRE
  F.coef    <- A$F[1]/A$F[2]
  df1       <- A$Df[1]
  df2       <- A$Df[2]
  P         <- 1 - pf(F.coef, df1, df2) # p-value of stat

#####Enter Jackknife#####
# Get a list of the SIRE codes
  menuTable(varnames = "SIRE", data = Data, print.p=F, save.name = "Sire.Sizes")
# Delete Sires for which counts are zero
  Sire.Sizes <- Sire.Sizes[Sire.Sizes$Count>0,]
  Nos.of.Sires <- length(Sire.Sizes$SIRE)
  Pseudovalues <- matrix(0,Nos.of.Sires,3)
# Now iterate through Sires
  for (Ith.Sire in 1: Nos.of.Sires)
    {
# Get code for SIRE and delete from data set, calling new data Data.minus.ONE
  I      <- Sire.Sizes[Ith.Sire,1]
  menuSubset(data = Data, subset.expression = "SIRE!=I", save.name =
    "Data.minus.ONE", show.p = F)
# Call function and calculate pseudovalues
  H2.minus.ONE <- H2.Estimator(Data.minus.ONE)
  Sire.Pseudo <- H2.OBS[2]*Nos.of.Sires - H2.minus.ONE[2]*(Nos.of.Sires-1)
  Dam.Pseudo <- H2.OBS[3]*Nos.of.Sires - H2.minus.ONE[3]*(Nos.of.Sires-1)
  SD.Pseudo <- H2.OBS[4]*Nos.of.Sires - H2.minus.ONE[4]*(Nos.of.Sires-1)
# Store data
  Pseudovalues[Ith.Sire,] <- c(Sire.Pseudo,Dam.Pseudo,SD.Pseudo)
    } # Next pseudovalue

# Analyse heritability pseudovalues
# Test for difference between SIRE and DAM estimate
  Sire.Dam.test <- t.test(x=Pseudovalues[,1], y=Pseudovalues[,2], mu<0,
    paired=T,var.equal=T, conf.level=.95)
  Y            <- Pseudovalues[,1] - Pseudovalues[,2]
  SE.t        <- sqrt(var(Y)/Nos.of.Sires)
# Calculate Jackknife estimates
  h2.Sire     <- mean(Pseudovalues[,1])
  h2.Dam     <- mean(Pseudovalues[,2])
  h2.SireDam <- mean(Pseudovalues[,3])
  n          <- nrow(Pseudovalues)
  SE.Sire    <- sqrt(var(Pseudovalues[,1])/n)
  SE.Dam     <- sqrt(var(Pseudovalues[,2])/n)
  SE.SireDam <- sqrt(var(Pseudovalues[,3])/n)
##### Output results #####
# Nested ANOVA results and original estimate
  print(A)
  print(c(F.coef,df1,df2,P))
  print(H2.OBS) # Sireh2, Damh2, SireDamh2
# Jackknife estimates and t test between SIRE and DAM
  print(c(h2.Sire,SE.Sire,h2.Dam,SE.Dam))
  print(c(h2.SireDam,SE.SireDam))
  Sire.Dam.test
```