Genetic evaluation of calving to first insemination using natural and artificial insemination mating data

K. A. Donoghue, R. Rekaya, J. K. Bertrand, and I. Misztal

Animal and Dairy Science Department, University of Georgia, Athens 30602-2771

ABSTRACT: Mating and calving records for 51,084 first-parity heifers in Australian Angus herds were used to examine the relationship between probability of calving to first insemination (CFI) in artificial insemination and natural service (NS) mating data. Calving to first insemination was defined as a binary trait for both sources of data. Two Bayesian models were employed: 1) a bivariate threshold model with CFI in AI data regarded as a trait separate from CFI in NS data and 2) a univariate threshold model with CFI regarded as the same trait for both sources of data. Posterior means (SD) of additive variance in the bivariate analysis were similar: 0.049 (0.013) and 0.075 (0.021) for CFI in AI and NS data, respectively, indicating lack of heterogeneity for this parameter. A similar trend was observed for heritability in the bivariate analysis, with posterior means (SD) of 0.025 (0.007) and 0.048 (0.012) for AI and NS data, respectively. The posterior means (SD) of the additive covariance and corresponding genetic correlation between the traits were 0.048 (0.006) and 0.821 (0.138), respectively. Differences were observed between posterior means for herd-year variance: 0.843 vs. 0.280 for AI and NS data, respectively, which may reflect the higher incidence of 100% conception rates within a herd-year class (extreme category problem) in AI data. Parameter estimates under the univariate model were close to the weighted average of the corresponding parameters under the bivariate model. Posterior means (SD) for additive, herd-year, and service sire variance and heritability under the univariate model were 0.063 (0.007), 0.56 (0.029), 0.131 (0.013), and 0.036 (0.007), respectively. These results indicate that, genetically, cows with a higher probability of CFI when mated using AI also have a high probability of CFI when mated via NS. The high correlation between the two traits, along with the lack of heterogeneity for the additive variance, implies that a common additive variance could be used for AI and NS data. A single-trait analysis of CFI with heterogeneous variances for herd-year and service sire could be implemented. The low estimates of heritability indicate that response to selection for probability of calving to first insemination would be expected to be low.

Key Words: Beef Cattle, Female Fertility, Threshold Models

Introduction
Previous research investigating reproductive performance under natural service matings has focused on the continuous traits of calving date and days to calving (e.g., Buddenberg et al., 1990; Meyer et al., 1990; Johnston and Bunter, 1996). These studies found that both traits were lowly heritable but had a clear economic interpretation. Studies have also used information from artificial insemination mating data to evaluate reproductive performance (e.g., Evans et al., 1999; Doyle et al., 2000). Traits that describe female fertility measures arising from artificial insemination data conform well to the threshold model and also exhibit genetic variation.

Little attention has been given to genetic evaluations of fertility that incorporate both natural service and artificial insemination information. Such an evaluation requires trait definitions such that information from both types of matings can be fairly combined and compared. A binary trait (calving to first insemination) that evaluates the probability that a female will produce a live calf from her first opportunity to do so in artificial insemination data could be compa-
rable to a female that produces a calf as a result of becoming pregnant within the first 21 d of the breeding season in natural mating data. For a complete analysis of the trait, full mating information on every female should be available, including management of estrus for artificial insemination matings.

The objective of this study was to develop methodology to combine both types of mating data into a joint genetic evaluation, and provide one genetic value to report for fertility. Genetic parameters for calving to first insemination were estimated under bivariate and univariate models during the course of this study.

Materials and Methods

Data

Bivariate Analysis. The data consisted of mating and calving records for first-parity females from Angus herds in Australia. Females had either an artificial insemination (AI) or a natural service (NS) record, but not both types of records. Only females having their first mating record between 270 and 625 d of age were included in the analysis. Before editing, there were 53,094 mating records from first-calf heifers born between 1987 and 2000 available in the database. Females with incomplete records (n = 541) and whose mating records resulting in multiple births (n = 274) were removed during edits. Incomplete records included those in which the information on sex of calf and identification of mating sire were missing. In addition, mating records whose gestation length was more than two SD lower than the mean (by sex of calf) were considered outliers (Kadarmideen and Coffey, 2001) and were removed during edits (n = 471). Gestation length (GL) was computed using AI data as the difference between the insemination date and subsequent calving date, and averaged by sex of calf. Average (SD) GL were 279.2 d (5.2 d) and 280.3 d (5.2 d) for female and male calves, respectively. Finally, all herd-year groups containing only a single record (n = 724) were removed from the data set. After editing, the AI and NS data sets consisted of 16,358 and 34,726 records, respectively. A total of 2,239 and 4,465 sires were represented in the AI and NS data sets, respectively, with 1,658 sires having progeny in both data sets. The structure of the data sets is shown in Table 1.

Calving to first insemination as recorded in the database.

Table 1. Descriptive statistics of the datasets for the bivariate and single trait analyses

<table>
<thead>
<tr>
<th>Variable</th>
<th>AI</th>
<th>NS</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of records</td>
<td>16,358</td>
<td>34,726</td>
<td>51,084</td>
</tr>
<tr>
<td>No. of sires</td>
<td>2,239</td>
<td>4,465</td>
<td>5,047</td>
</tr>
<tr>
<td>No. of herd × year classes</td>
<td>1,136</td>
<td>2,014</td>
<td>3,150</td>
</tr>
<tr>
<td>No. of service sires</td>
<td>687</td>
<td>3,504</td>
<td>3,969</td>
</tr>
<tr>
<td>CFI frequency, %d</td>
<td>79.3</td>
<td>53.0</td>
<td>61.5</td>
</tr>
</tbody>
</table>

\(a\) Artificial insemination mating data.

\(b\) Natural service mating data.

\(c\) Combined (AI and NS) data.

\(d\) Calving to first insemination.

Model

Bivariate Analysis. A bivariate threshold model was used, with both AICFI and NSCFI as binary traits possessing an underlying bivariate normal distribution. These models have been implemented in the animal breeding field using marginal maximum likelihood or empirical Bayes methods by several researchers (Foulley et al., 1983; Hoeschele et al., 1986), and are an extension of the univariate threshold model (Gianola and Foulley, 1983).

A mixed linear model was used for analyses of the underlying liability for both AICFI and NSCFI. In matrix notation, the model can be written as follows:

\[
\lambda = X\beta + Z_s s + Z_u u + e
\]

where \(\lambda\) is a vector of unobserved liabilities; \(\beta\) is the vector of systematic effects (herd × year effects, month of mating effects, linear and quadratic covariates for age at mating); \(s\) is the vector of service sire effects; \(u\) is the vector of additive genetic values; \(e\) is the vector of residual terms; and \(X, Z_s, Z_u\) are known incidence matrices.

Conditionally on the model parameters, it was assumed that the sampling distribution of liabilities was

\[
p(\lambda|\beta, s, u, R_0) = N(X\beta + Z_s s + Z_u u, I \otimes R_0)
\]

where \(R_0\) is a \(2 \times 2\) variance-covariance matrix with the following structure:
Given the well-known nondefinability problem of the threshold models, at least two restrictions are needed. In this study, the threshold and the residual variances were arbitrarily set to 0 and 1, respectively. Further, all females had either an AI or a NS record; none had both traits measured. Consequently, the residual covariance cannot be inferred and it was set to 0, leading to

$$R_0 = \begin{bmatrix} \sigma^2_{e1} & \sigma_{e12} \\ \sigma_{e12} & \sigma^2_{e2} \end{bmatrix}$$

A normal distribution was used as prior for the effect of service sire:

$$p(s \mid S) \sim N(0, I \otimes S)$$

where $S = \begin{bmatrix} \sigma^2_{s1} & 0 \\ 0 & \sigma^2_{s2} \end{bmatrix}$ is the service sire (co)variance matrix. The following uniform bounded priors were assigned to the diagonal elements of matrix $S$:

$$p(\sigma^2_{si}) \sim U[0, 1] \quad \text{for } i = 1, 2$$

A multivariate normal distribution was used as prior for the animal effects:

$$p(u \mid A, G_0) \sim N(0, A \otimes G_0)$$

where $G_0 = \begin{bmatrix} \sigma^2_{g11} & \sigma_{g12} \\ \sigma_{g12} & \sigma^2_{g22} \end{bmatrix}$ is the additive (co)variance matrix, and $A$ is the additive relationship matrix between animals. The following uniform bounded priors were assigned to the elements of matrix $G_0$:

$$p(\sigma^2_{ui}) \sim U[0, 1] \quad \text{for } i = 1, 2$$

and

$$p(\sigma_{ui2}) \sim U[-\sqrt{\sigma^2_{g11} \sigma^2_{g22}}, \sqrt{\sigma^2_{g11} \sigma^2_{g22}}]$$

The joint posterior density is proportional to the product of the density of the conditional distribution of the observation times the joint prior density. Draws from the conditional posterior distribution of all of the parameters were obtained using a Gibbs sampler with data augmentation (Sorensen et al., 1995). The joint posterior was augmented with the liabilities for each female. After augmentation, all of the fully conditional posterior distributions of all model parameters can be derived as described by Albert and Chib (1993) and Sorensen et al. (1995). These distributions are normal for the location parameters; truncated normal for the liabilities; and scaled-inverted Wishart distributions for the dispersion parameters. Liabilities were sampled from their truncated normal distribution using inverse cumulative distribution function technique (Devroye, 1986).

**Single- Trait Analysis.** A univariate version of the above bivariate model was implemented. Similar priors (after adjustment of dimensions) were assumed to the model parameters. A full description of the derivation and implementation using a Bayesian approach via MCMC methods can be found in Heringstad et al. (2001).

**Results and Discussion**

For all analyses, convergence was assessed using methodology presented by Raftery and Lewis (1992).
The required length of the burn-in period was always less than 2,500 iterations for all parameters. Thus, 75,000 iterations of the sampler were run with a conservative 25,000 iterations discarded as burn-in. All remaining 50,000 iterations were retained without thinning for post-Gibbs analysis.

There was a large difference in the percentage of females calving to their first insemination between the AI and NS data sets (79 vs. 53%). It is reasonable to assume that higher levels of management could be associated with slightly higher first insemination conception rates in AI data. Due to the cost and higher level of management required to implement an AI program, these breeders may be employing more intensive techniques in order to maximize pregnancy rates. However, closer examination of the AI data revealed a high incidence of extreme category problem (ECP) at the herd-year level (all observations within a class falling into the same category). For example, 50% of herd-year classes contained observations that fell into the same category in AI data, compared to 22% in the NS data. This suggests that the large difference in percentage of females calving to their first insemination observed between the two types of matings is most likely due to incomplete recording in the AI data; that is, only successful inseminations have been reported in some herds.

Summaries of the posterior distributions of (co)variance components, heritabilities, and genetic correlation from the bivariate analysis are presented in Table 2. The posterior mean of the additive variance in the AI data (0.049) tended to be slightly lower than in NS data (0.075). Despite this trend, the posterior mean of additive variance in the AI data was within the high posterior density (HPD) interval (95%) for the NS data. These results indicate that there was no major statistical evidence of heterogeneity of variance for the additive effect in these two data sets. The posterior mean of the additive covariance between AICFI and NSCFI was 0.048, and the corresponding genetic correlation was 0.821. These results suggest a high, positive genetic correlation between CFI for AI matings and the same traits under NS matings. A large positive genetic correlation means that, genetically, cows with a higher probability of calving as a result of the first insemination when mated using AI also have a high probability of calving as a result of the first insemination when mated via natural service. The lack of statistical evidence of heterogeneity observed for the additive variance, together with the high positive correlation between both traits, indicates that a single-trait analysis with homogenous additive variances is a reasonable assumption. Thus, combining AI and NS data could be undertaken for the traits of calving to first insemination, assuming that other variances in the model did not exhibit heterogeneity.

The posterior means (SD) of heritability (h²) for CFI in both traits were low, and ranged from 0.03 to 0.05 (0.007 to 0.012). These heritabilities represent the lower bounds for the trait due to the inclusion of herd-year variance in the total phenotypic variance. By not including herd-year variance in the total phenotypic variance, heritability had an upper bound of 0.044 and 0.056 for AICFI and NSCFI, respectively. As a result of the lower estimate of the additive variance in the AI data set, the posterior means of both lower and upper bounds of h² were slightly smaller than the estimate in the NS data. Early studies of fertility reported heritabilities expressed on the observed scale for traits such as pregnancy, conception, and calving rates and found heifer fertility to be lowly heritable (0.05 to 0.17; Koots et al., 1994). Buddenberg et al. (1990) examined the h² of pregnancy on both the observed and underlying scales in Angus, Hereford, and Polled Hereford females. For first-parity females, these estimates ranged from 0.04 to 0.17 and 0.08 to 0.34 on the observed and under-

<table>
<thead>
<tr>
<th>Trait</th>
<th>Parameterab1</th>
<th>Mean</th>
<th>SD</th>
<th>HPD (95%)a</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICFI</td>
<td>σ²h</td>
<td>0.843</td>
<td>0.095</td>
<td>0.685 to 0.999</td>
</tr>
<tr>
<td></td>
<td>σ²s</td>
<td>0.058</td>
<td>0.014</td>
<td>0.034 to 0.085</td>
</tr>
<tr>
<td></td>
<td>σ²u</td>
<td>0.049</td>
<td>0.013</td>
<td>0.032 to 0.074</td>
</tr>
<tr>
<td></td>
<td>h²</td>
<td>0.025</td>
<td>0.007</td>
<td>0.016 to 0.038</td>
</tr>
<tr>
<td>NSCFI</td>
<td>σ²h</td>
<td>0.280</td>
<td>0.023</td>
<td>0.235 to 0.324</td>
</tr>
<tr>
<td></td>
<td>σ²s</td>
<td>0.195</td>
<td>0.018</td>
<td>0.161 to 0.231</td>
</tr>
<tr>
<td></td>
<td>σ²u</td>
<td>0.075</td>
<td>0.021</td>
<td>0.043 to 0.120</td>
</tr>
<tr>
<td></td>
<td>h²</td>
<td>0.048</td>
<td>0.012</td>
<td>0.031 to 0.077</td>
</tr>
<tr>
<td>AICFI-NSCFI</td>
<td>σu12</td>
<td>0.048</td>
<td>0.006</td>
<td>0.032 to 0.057</td>
</tr>
<tr>
<td></td>
<td>r²</td>
<td>0.821</td>
<td>0.138</td>
<td>0.535 to 0.999</td>
</tr>
</tbody>
</table>

aσ²h = herd-year variance; σ²s = service sire variance; σ²u = additive genetic variance.

b2|h² = σ²h²/(σ²h² + σ²s² + σ²u² + 1).

cσu12 is the additive covariance between the two traits, respectively, and r² is the corresponding correlation.

dUpper and lower bounds of high posterior density interval (95%).

Table 2. Genetic parameter estimates for probability of calving to first insemination for bivariate analysis

The large difference in the percentage of heifers calving to their first insemination in artificial insemination mating data.
Bayesian approach vs. Method R to estimate variance components for heifer pregnancy and obtained $h^2$ estimates of 0.138 ± 0.09, with a 90% confidence interval (CI) of 0.015 to 0.347. The point estimates of $h^2$ in our study are lower than the estimates reported in these studies; however, they are both within the 90% CI reported by Evans et al. (1999). The lower estimates of $h^2$ observed in this study could result from more appropriate analytical procedures for the data, as the majority of ECP in the AI data were herd-year variance, leading to an inflated herd-year variance. The type of heat detection and synchronization treatment used was not recorded in the database, and it is possible that, within a herd-year, heifers were exposed to different heat detections and/or treatments to synchronize estrus. In the future, both heat detection and synchronization treatment should be recorded for AI matings and herd-year groups be redefined to include this information. The higher incidence of ECP in the AI data may also have contributed to differences in herd-year variance between the types of mating. This bias in the estimate for herd-year variance for AI data is most likely positive in magnitude, as the majority of ECP in the AI data were herd-year classes in which all observations had a trait value equal to 1. The different estimates for herd-year variance in AI and NS data sets indicate that some heterogeneity of variance may be occurring, although it is not clear whether this is a function of real differences between the types of mating or a lack of information about management and/or incomplete recording in the AI data (ECP). Further research should be undertaken to investigate whether heterogeneity of herd-year variance is observed when full management information is available for AI data.

The posterior means of service sire variance were lower in magnitude than the herd-year variance, but higher than the additive variance for both AI and NS data. As well, the posterior mean was higher in the NS data in comparison to the AI data (0.195 vs. 0.058). This result may reflect greater sire differences in NS data due to variation in libido, reduction in fertility between breeding soundness exam and actual use, and differences due to inexperience or bull aggressiveness.

### Table 3. Genetic parameter estimates for probability of calving to first insemination for single-trait analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>SD</th>
<th>HPD (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{u}^2$</td>
<td>0.560</td>
<td>0.029</td>
<td>0.505 to 0.618</td>
</tr>
<tr>
<td>$\sigma_{s}^2$</td>
<td>0.131</td>
<td>0.013</td>
<td>0.107 to 0.157</td>
</tr>
<tr>
<td>$\sigma_{h}^2$</td>
<td>0.063</td>
<td>0.012</td>
<td>0.042 to 0.088</td>
</tr>
<tr>
<td>$h^2$</td>
<td>0.036</td>
<td>0.007</td>
<td>0.025 to 0.048</td>
</tr>
</tbody>
</table>

- $\sigma_{u}^2$ = herd-year variance; $\sigma_{s}^2$ = service sire variance; $\sigma_{h}^2$ = additive genetic variance.
- $h^2 = \sigma_{h}^2/(\sigma_{u}^2 + \sigma_{s}^2 + \sigma_{h}^2 + 1)$.
- Upper and lower bounds of high posterior density interval (95%).

Further investigation is required to determine whether service sire variances are really heterogeneous and, if so, the nature of these differences.

Summaries of the posterior distributions of the variance components and heritability from the single-trait analysis are presented in Table 3. The posterior mean of the additive variance for CFI was very close to the weighted mean of the point estimates for this parameter in the bivariate analysis. As well, both point estimates from the bivariate analysis were within the HPD (95%) for the parameter in the single-trait analysis. Pearson correlations between posterior means of additive genetic effects for sires with progeny with records were 0.962, 0.970, and 0.968 for AICFI-NSCFI, AICFI-CFI, and NSCFI-CFI, respectively. These correlations indicate that no major reranking of sires would be expected when AI and NS data is combined. These results, in conjunction with the lack of heterogeneity for additive genetic variance observed for the bivariate analysis, indicate that fertility data from AI and NS matings can be combined using a common additive variance for the trait of CFI.

The same trend was observed for heritability for CFI, where the posterior mean was close to the weighted mean of the point estimates for this parameter in the bivariate analysis, and both point estimates from the bivariate analysis were within the HPD (95%) for the parameter in the single-trait analysis. As discussed previously, this heritability represents the lower bound for the trait due to the inclusion of herd-year variance in the total phenotypic variance. By not including the herd-year variance in the total phenotypic variance, heritability for probability of CFI has an upper bound of 0.05. As in the bivariate analysis, both lower and upper bound estimates of $h^2$ for the probability of CFI were lower than literature estimates of heifer fertility (Evans et al., 1999) but were within range of standard error associated with these literature estimates, as discussed previously. These results indicate that the trait, probability of CFI, with AI and NS data combined for a single-trait analysis, is lowly heritable, and response to selection on this trait is expected to be small. The use of records from later parities should be investigated as a means of increasing the accuracy of selection for this trait.
Posterior means for herd-year and service sire variance were also close to the weighted mean of the point estimates for these parameters in the bivariate analysis. However, for both parameters, point estimates from the bivariate analysis were outside the HPD (95%) for the corresponding parameter in the single-trait analysis. These results are most likely a reflection of the lack of information for the AI data, particularly with respect to heat detection and estrus synchronization, as discussed previously. If heterogeneous variances do exist for herd-year and service sire, these need to be accounted for in the model.

Implications

The lack of heterogeneity observed for additive variance in the bivariate analysis, along with the high, positive genetic correlation of calving to first insemination under artificial insemination and natural service matings, indicates a common additive variance could be used. Differences were observed between sources of data for herd-year and service sire variances, which are most likely a reflection of the poor structure for the artificial insemination data. These results imply that a single-trait analysis of calving to first insemination with heterogeneous variances for herd-year and service sire could be implemented. Probability of calving to first insemination had a low heritability in both analyses, indicating that response to selection is expected to be low. The use of repeated records over parities to increase the accuracy of selection should be investigated.

Literature Cited
