



## Comparison of Genetic Parameter Estimates of Total Sperm Cells of Boars between Random Regression and Multiple Trait Animal Models

S.-H. Oh\* and M. T. See<sup>1</sup>

North Carolina A&T State University, Greensboro, NC, USA

**ABSTRACT :** The objective of this study was to compare random regression model and multiple trait animal model estimates of the (co) variance of total sperm cells over the active lifetime of AI boars. Data were provided by Smithfield Premium Genetics (Rose Hill, NC). Total number of records and animals for the random regression model were 19,629 and 1,736, respectively. Data for multiple trait animal model analyses were edited to include only records produced at 9, 12, 15, 18, 21, 24, and 27 months of age. For the multiple trait method estimates of genetic and residual variance for total sperm cells were heterogeneous among age classifications. When comparing multiple trait method to random regression, heritability estimates were similar except for total sperm cells at 24 months of age. The multiple trait method also resulted in higher estimates of heritability of total sperm cells at every age when compared to random regression results. Random regression analysis provided more detail with regard to changes of variance components with age. Random regression methods are the most appropriate to analyze semen traits as they are longitudinal data measured over the lifetime of boars. (**Key Words :** Genetic Correlation, Boar, Semen, Random Regression Model, Multiple Trait Animal Model)

### INTRODUCTION

Artificial insemination plays an important role in animal breeding by allowing greater utilization of genetically superior sires. The opportunity for genetic improvement of male fertility traits has been previously shown (Brandt and Grandjot, 1998; Oh et al., 2006). However, appropriate statistical methods for analysis of semen traits in pigs has not been extensively studied. Total sperm cells per ejaculate are longitudinal data whereas volume changes over age. In previous studies, this type of data was analyzed by multiple trait methods choosing the most important time points as separate traits. Because of the number of potential observations over the lifetime of a boar, it would be difficult to thoroughly analyze this type of data due to computational limits. Semen data have also been analyzed similarly to growth curves ignoring genetic effects (Morant and Gnanasakthy, 1989), or were considered simple repeated measurements ignoring time dependency.

In many cases, the assumption of a univariate repeated model is not appropriate while a full multivariate model with the number of traits equal to the number of ages would

result in a highly over-parameterized analysis (Meyer and Hill, 1997). Random regression models (RRM) developed by Meyer (1998) have been extensively applied to the test-day model analysis of milk yield of dairy cattle (Olori et al., 1999; Strabel and Misztal, 1999; Meyer, 2000) and have also been fitted to weight data of pigs (Huisman et al., 2002). RRM provide a method for analyzing independent components of variation that reveal specific patterns of change over time.

Evidence has been reported that changes in animal performance with increasing age are influenced by genetic factors. Animal breeders are interested in genetic parameters that describe the change of traits over time. Analysis of these changes can be undertaken using repeatability (Henderson, 1984), multiple trait (Reents et al., 1995) or random regression models. Random regression allows for the calculation of (co) variances at every age (Meyer, 1998). Multiple trait animal models have traditionally been used for traits measured over time by defining observations at distinct ages as different traits. However, computational requirements need to explain the number of traits equal to the number of ages (Meyer and Hill, 1997).

Accordingly, records collected over ages are often analyzed as repeated measurements or as different traits that are separated by specific intervals. However, it is the object

\* Corresponding Author: S.-H. Oh. Tel: +1-336-334-7672, Fax: +1-336-334-7288, E-mail: soh@ncat.edu

<sup>1</sup> North Carolina State University, Raleigh, NC, USA.

Received July 9, 2007; Accepted February 27, 2008

of interest how much RRM is better than multi-trait analysis. The objective was to compare the (co) variance of total sperm cells (TSC;  $\times 10^9$ ) over the active lifetime of AI boars between RRM and a multi-trait animal model.

## MATERIALS AND METHODS

### Data source

Records of total sperm cells per ejaculate ( $n = 19,629$ ) from 834 boars were provided by Smithfield Premium Genetics (Rose Hill, NC). One thousand seven hundred and thirty six individuals were included in the pedigree file. Boars represented three breeds and were housed in AI stations. Each AI station was similar in number of boars of each breed and management. Thirty-four collectors recorded these data over 5 years (1998 to 2002) with approximately one-half of all records in 2000. Data were distributed evenly across seasons. Total sperm cells per ejaculate were determined by multiplying semen volume, measured as the weight of the ejaculate volume, by total concentration as measured using a self-calibrating photometer. Observations were removed when the number of data at a given age of boar classification time point was less than 10, or total sperm cells were missing, zero or less than zero. Differences between boar collection date and birth date were used to provide each record with a fixed age of boar classification in weeks for RRM. When a boar had two observations during one week of age the record closest to the whole week was utilized.

For the multiple trait analyses, records were edited to include only records produced at 9, 12, 15, 18, 21, 24, and 27 months of age and were used as separate traits. Number of observations at 9, 12, 15, 18, 21, 24, and 27 months of age were 305, 413, 370, 306, 248, 200 and 109, respectively. Number of animals with valid records was 750. Frequency of records was highest at 12 months of age, decreasing gradually over time.

### Multiple trait animal model

Least square means were estimated for the fixed effects of breed and AI station, and differences within fixed effects were compared by least significant differences using the PDIFF option in SAS 8.01 (SAS Institute Inc., Cary, NC). The statistical model included fixed effects of year-season, breed, collector, and AI station.

Variance components for the multiple trait analyses were estimated by derivative free REML using MTDFREML (Boldman et al., 1995). Fixed effects for the model were year-season, breed, collector and AI station. Convergence was considered to have been reached when the variance of the -2 log likelihood in the simplex was less than  $1 \times 10^{-9}$ . To obtain convergence, as well as standard

errors for parameter estimates, the seven ages classifications of total sperm cells were evaluated in five trait analyses. Therefore, twenty-one combinations of five trait analyses were conducted. The results presented are the means for each parameter estimate, and standard deviations were considered as standard errors.

The multiple trait model was as follows:

$$Y_{ijklmn} = \mu + A_i + YS_j + B_k + C_l + F_m + e_{ijklmn}$$

where,  $\mu$  is overall mean,  $A_i$  is the random additive genetic effect of  $i^{\text{th}}$  animal,  $YS_j$  is the fixed effect of  $j^{\text{th}}$  year-season,  $B_k$  is the fixed effect of  $k^{\text{th}}$  breed,  $C_l$  is the fixed effect of  $l^{\text{th}}$  collector,  $F_m$  is the fixed effect of  $m^{\text{th}}$  AI station, and  $e_{ijklmn}$  is measurement error. The vector presentation of this model is:  $Y = Xb + Zu + e$  where  $Y$  is the vector of observations for all traits,  $b$  is a vector of common fixed effects due to year-season, breed, collector and AI station,  $u$  is a vector of random genetic effects,  $e$  is a vector of residuals,  $X$  and  $Z$  are incidence matrices relating observations to the fixed and animal effects and  $E [y' u' e'] = [b' X' 0' 0']'$ . Variances of the random variables were:

$$V \begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} G_o \otimes A & 0 \\ 0 & R_o \otimes I \end{bmatrix}$$

Where,  $\otimes$  denotes a direct product operation,  $G_o$ , and  $R_o$  are genetic and residual covariance matrices, with order equal to the number of traits in the analysis, and  $A$  is the numerator relationship matrix.

### Random regression model

Random regression procedures are fully described in Oh et al. (2006). Parameters were estimated for total sperm cells by age of boar classification under a random regression model using DxMRR (Meyer, 1998). The analysis model included breed, collector and year-season as fixed effects; additive genetic effects, permanent environmental effect of boar and measurement error as random effects. Random regression models were fitted to evaluate all combinations of first- through seventh-order polynomial covariance functions for fixed age of boar classification, additive genetic and permanent environmental effects. This resulted in the evaluation of 343 models. Goodness of fit for models was tested using log likelihood value, Akaike's Information Criterion (AIC) and Schwarz Criterion (SC) (Oh et al., 2006).

$$AIC = -2\log L + 2 \times p$$

$$SC = -2\log L + p \times \log(N - r(X))$$

**Table 1.** Number of total sperm cell records and simple statistics by age of boar classification for multiple trait analyses

Age (month)	9	12	15	18	21	24	27
Number of records	305	413	370	306	248	200	109
Mean ( $\times 10^{-9}$ )	85.16	106.3	114.3	117.9	115.5	116.4	125.8
SD	36.42	40.21	43.89	44.81	44.83	44.19	42.06
CV	42.77	37.85	38.41	38.02	38.83	37.98	33.43

**Table 2.** Least squares means of total sperm cells ( $\times 10^{-9}$ ) for breed and farm by age of boar

Age (month)	9	12	15	18	21	24	27
Breed 1	69.89 <sup>a</sup>	97.22 <sup>a</sup>	99.47 <sup>a</sup>	101.5 <sup>a</sup>	107.8 <sup>a</sup>	123.3 <sup>a</sup>	105.3 <sup>a</sup>
Breed 2	76.11 <sup>ab</sup>	105.7 <sup>ab</sup>	104.2 <sup>a</sup>	110.3 <sup>ab</sup>	109.9 <sup>a</sup>	116.7 <sup>a</sup>	112.1 <sup>a</sup>
Breed 3	79.36 <sup>bc</sup>	108.8 <sup>bc</sup>	117.7 <sup>c</sup>	122.8 <sup>bc</sup>	130.0 <sup>c</sup>	142.0 <sup>c</sup>	109.8 <sup>a</sup>
AI station 1	81.80 <sup>a</sup>	109.8 <sup>a</sup>	118.9 <sup>a</sup>	120.2 <sup>a</sup>	120.2 <sup>a</sup>	122.4 <sup>a</sup>	120.0 <sup>a</sup>
AI station 2	68.43 <sup>a</sup>	97.97 <sup>a</sup>	95.37 <sup>b</sup>	102.9 <sup>a</sup>	111.6 <sup>a</sup>	132.2 <sup>a</sup>	98.15 <sup>a</sup>

H0: LSMean (i) = LSMean (j) (p<0.05).

Where, p is the number of parameters estimated and r(X) is the rank of the coefficient matrix of fixed effects.

The general model is

$$y_{ij} = F_{ij} + \sum_{n=0}^{k_F-1} \beta_n \phi_n(w_{ij}) + \sum_{n=0}^{k_A-1} \alpha_{in} \phi_n(w_{ij}) + \sum_{n=0}^{k_P-1} \delta_{in} \phi_n(w_{ij}) + \epsilon_{ij}$$

where  $y_{ij}$  is the  $j^{\text{th}}$  record from the  $i^{\text{th}}$  animal,  $w_{ij}$  is the standardized (-1 to 1) age at recording,  $\phi_n(w_{ij})$  is the  $n^{\text{th}}$  Legendre polynomial of age,  $F_{ij}$  is a set of fixed effects,  $\beta_n$  are the fixed regression coefficients to model the population mean,  $\alpha_{in}$  are the random regression coefficients for additive genetic effects, and  $\delta_{in}$  are the random regression coefficients for permanent environmental effects, respectively.  $k_F$ ,  $k_A$ , and  $k_P$  denote the corresponding orders of fit.

In matrix notation,

$$y = Xb + Za + Cp + e$$

Where,

y: vector of  $N$  observations measured on  $ND$  animals

b: vector of fixed effects (including  $F_{ij}$  and  $\beta_n$ )

a: vector of  $kA \times NA$  additive genetic random regression coefficients

p: vector of  $kR \times ND$  permanent environmental random regression coefficients

e: vector of  $N$  measurement errors

X, Z and C: corresponding design matrices

$kA$  and  $kR$ : the order of fit for a and p and corresponding genetic and permanent environmental covariance function A and R.

$$V \begin{bmatrix} a \\ p \\ e \end{bmatrix} = \begin{bmatrix} K_A \otimes A & 0 & 0 \\ 0 & K_P \otimes I & 0 \\ 0 & 0 & R \end{bmatrix}$$

$K_A$  and  $K_P$  are the matrices of coefficients of the covariance function for additive genetic and permanent environmental effects. A is the numerator relationship matrix, and I is an identity matrix. It is assumed that all measurement errors are equal.

## RESULTS AND DISCUSSION

### Multiple trait animal model analyses

Mean and standard deviation of total sperm cells increased with age (Table 1). Coefficient of variation was the highest at 9 months of age and lowest at 27 months of age, however, there was not much difference among ages (Table 1). Breed 3 showed significantly more total sperm cells than both breeds 1 and 2 only at 15, 21 and 24 months of age (p<0.05; Table 2). At 9, 12 and 18 months of age, breed 3 had more total sperm cells than breed 1 but not breed 2 (p>0.05) as shown in Table 2. Also, there was no breed effect observed at 27 months of age in the comparison of total sperm cells produced. Artificial insemination stations did not differ (p>0.05) except at 15 months of age. There was no AI station by age interaction (p>0.05).

Estimates of genetic variance (Table 3) for total sperm cells were lower for boars 9, 12 and 15 months of age. The genetic variance for total sperm cells was also much higher at 24 months as compared to all other ages, which may be due to sampling of records. Estimates of permanent environmental variance (Table 3) for total sperm cells were close to zero and did not differ across age classifications. Estimates of residual variance (Table 3) for total sperm cells were lower at 9, 12 and 27 months and higher at 15, 18 and 21 months. A significantly lower residual variance was observed for total sperm cells at 24 months. Again, this may be due to sampling of records. These results suggest that both genetic and residual variance for total sperm cells over age classifications are heterogeneous and possibly different traits.

Heritability estimates for total sperm cells were similar

**Table 3.** Estimates of genetic variance ( $\sigma_a^2$ ), permanent environmental variance ( $\sigma_{pe}^2$ ), residual variance ( $\sigma_e^2$ ) and heritability ( $h^2$ ) across age classifications for total sperm cells over the active lifetime of AI boars

Age (month)	9	12	15	18	21	24	27
$\sigma_a^2 \pm SE$	320±57	390±60	439±47	487±88	512±96	1,388±163	633±153
$\sigma_{pe}^2 \pm SE$	3±2	7±7	9±10	8±6	9±10	18±13	11±9
$\sigma_e^2 \pm SE$	799±55	970±60	1,249±44	1,277±77	1,170±77	341±160	906±144
$h^2 \pm SE$	0.28±0.05	0.29±0.04	0.26±0.03	0.27±0.05	0.30±0.05	0.79±0.09	0.41±0.09

**Table 4.** Estimates of residual (above diagonal) and genetic (below diagonal) correlation between age classifications for total sperm cells over the active lifetime of AI boars

Age (month)	9	12	15	18	21	24	27
9	-	0.30±0.01	0.16±0.02	0.14±0.18	0.25±0.16	0.63±0.17	0.32±0.10
12	0.73±0.07	-	0.27±0.04	0.06±0.05	0.21±0.13	0.82±0.10	0.41±0.08
15	0.77±0.07	0.87±0.13	-	0.25±0.08	0.31±0.06	0.38±0.16	0.17±0.07
18	0.50±0.49	0.77±0.18	0.71±0.23	-	0.30±0.08	0.24±0.12	0.22±0.08
21	0.16±0.53	0.52±0.42	0.62±0.22	0.62±0.29	-	0.71±0.07	-0.06±0.12
24	-0.05±0.14	0.18±0.26	0.29±0.16	0.32±0.09	0.37±0.11	-	0.31±0.09
27	0.21±0.05	0.44±0.22	0.50±0.14	0.49±0.12	0.52±0.12	0.54±0.05	-

across age classifications with the exception of 24 months of age (Table 3). Heritability of total sperm cells at 24 months of age was high because of high genetic variance and the low residual variance at that age. This may be due in part to selection of records for specific age points. Heritability estimates in this study were similar to those reported in the literature. Masek et al. (1977) estimated 0.24 as repeatability using two-factorial hierarchical analysis of variance. Du Mesnil du Buisson et al. (1978) reported that the heritability for the number of spermatozoa produced per ejaculate in comparable collection rate conditions was 0.35, even though the standard deviations were too high to affirm interpreting the results. Huang and Johnson (1996) estimated repeatability of total number of sperm (billion) as 0.26 for three collections per week, and 0.16 for daily collections. On the other hand, Brandt and Grandjot (1998) reported that heritability and repeatability of number of sperm cells was 0.24 and 0.46 on average, respectively.

Genetic and residual correlations between measures of total sperm cells at different ages averaged 0.64 and 0.30, respectively (Table 4). Genetic correlations between adjacent ages were higher than those between more distant ages. Decreasing genetic correlations with increasing age may also be due to the limited amount of data and the selection of records to defined age ranges. Huisman et al. (2002) reported a similar observation in an evaluation of pig body weights.

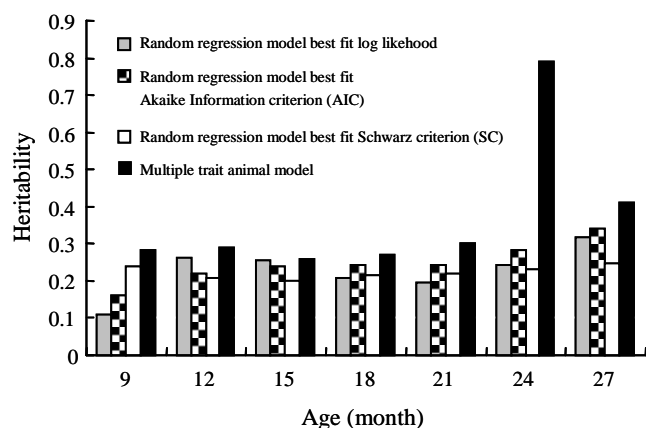
#### Random regression model analysis

Results from the random regression analysis are presented in Oh et al. (2006). In brief the random regression model, fitting 6th, 5th, and 7th order for fixed, additive genetic and permanent environmental effects showed the

largest log likelihood value. This model was the 4th best fitting model based on AIC and the 52nd best fitting model based on SC. AIC showed best fit when, respectively, 6th, 4th, and 7th order fixed, additive genetic and permanent environmental effects were fitted. This was the 3rd best-fitting model based on log likelihood and 20th best-fitting model based on SC. Schwarz Criterion showed the best fit when 4th, 2nd, and 7th order polynomials were fitted for fixed, additive genetic and permanent environmental effects, respectively. This model was ranked with the 10th best fit by log likelihood and 2nd best-fitting model by AIC. Based on the conservative nature of SC and the relative ranking by the other criterion, this model may be the best overall fit. Heritability estimates for total sperm cells over weeks of age ranged from 0.27 to 0.48. Standard deviations tended to decrease from 33 weeks of age to about 45 weeks, maintained consistent intervals by 100 weeks of age and then increased rapidly. This increase in variance closely follows the numbers of total sperm cells records over age.

#### Comparison between models

Estimates of genetic parameters from both multiple trait and random regression methods would indicate that measures of total sperm cells at different ages are genetically different traits. Figure 1 shows the comparison of heritability estimates between the three best-fit models determined from random regression model analysis and the evaluation of seven ages by multiple trait animal model analyses. The results are similar except for the heritability at 24 months of age from the multiple trait method that had very high genetic variance. Other than 24 months, the results are consistent but it appears that the multiple trait method resulted in a higher estimate of heritability of total



**Figure 1.** Comparison of heritability estimates of total sperm cells by age of boar classification from random regression and multiple trait animal models.

sperm cells at each age. These higher estimates may be due to the reduced amount of data used in the multiple trait method or possibly the age classifications selected. The ability to accurately estimate genetic correlations between different ages is reduced by limiting records to specific ages. Therefore, multiple trait methods that do not include all available data may not be the most appropriate method for analyzing longitudinal data. However, this method may be improved with sufficient numbers of records at each age and if availability of computer resources allowed for more age classifications. Random regression analysis provides much more detail with regard to the changes of the variance components with age. Genetic correlations between total sperm cells at different ages were larger for adjacent ages. RRM with comparatively high order polynomials for fixed, additive genetic and permanent environmental effects provided the best fit.

## CONCLUSION

These studies show there is an opportunity for genetic selection on semen traits. Estimates of genetic parameters would indicate that measures of total sperm cells at different ages are genetically different traits. However, the ability to accurately estimate genetic correlations between different ages is reduced if records are limited to specific ages. Therefore, multiple trait methods that do not allow for the inclusion of all available data may not be appropriate for analyzing longitudinal data. This method may be appropriate with sufficient numbers of records at each age and availability of computer resources. Random regression methods are the most appropriate to analyze semen traits as they are longitudinal data measured over the lifetime of boars. Additional work is needed to understand the relative economic importance of semen traits in the development of breeding objectives.

## REFERENCES

- Boldman, K. G., L. A. Kriese, L. D. Van Vleck, C. P. Van Tassel and S. D. Kachman. 1995. A Manual for Use of MTFREML. A set of programs to obtain estimates of variances and covariances (Draft). U.S. Department of Agriculture, Agricultural Research Service.
- Brandt, H. and G. Grandjot. 1998. Genetic and environmental effects on male fertility of AI boars. Proc. 6th World Congr. Genet. Appl. Livest. Prod. 23:527-530.
- Du Mesnil du Buisson, F., M. Paquignon and M. Courot. 1978. Boar sperm production: use in artificial insemination - a review. Livest. Prod. Sci. 5:293-302.
- Henderson, C. R. 1984. Applications of linear models in animal breeding. Univ. of Guelph, Guelph, Canada.
- Huang, Y. T. and R. K. Johnson. 1996. Effect of selection for size of testes in boars on semen and testis traits. J. Anim. Sci. 74:750-760.
- Huisman, A. E., R. F. Veerkamp and J. A. M. van Arendonk. 2002. Genetic parameters for various random regression models to describe the weight data of pigs. J. Anim. Sci. 80:575-582.
- Masek, N., J. Kuciel, J. Masek and L. Maca. 1977. Genetical analysis of indicators for evaluating boar ejaculates. Acta Universitatis Agriculturae, Facultas Agronomica, Brno. 25:133-139.
- Meyer, K. 1998. Modeling 'repeated' records: covariance functions and random regression models to analyse animal breeding data. 6th World Congr. Genet. Appl. Livest. Prod. 25:517-520.
- Meyer, K. 2000. Random regressions to model phenotypic variation in monthly weights of Australian beef cows. Livest. Prod. Sci. 65:19-38.
- Meyer, K. 2001. Estimates of direct and maternal covariance functions for growth of Australian beef calves from birth to weaning. Genet. Sel. Evol. 33:487-514.
- Meyer, K. and W. G. Hill. 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood. Livest. Prod. Sci. 47:185-200.
- Morant, S. V. and A. Gnanasakthy. 1989. A new approach to the mathematical formulation of lactation curves. Anim. Produc. 49:151-162.
- Oh, S.-H., M. T. See, T. E. Long and J. M. Galvin. 2006. Genetic parameters for various random regression models to describe total sperm cells per ejaculate over the reproductive lifetime of boars. J. Anim. Sci. 84:538-545.
- Olori, V. E., W. G. Hill, B. J. McGuirk and S. Brotherstone. 1999. Estimating variance components for test day milk records by restricted maximum likelihood with a random regression animal model. Livest. Prod. Sci. 61:53-63.
- Strabel, T. and I. Misztal. 1999. Genetic parameters for first and second lactation milk yields of Polish Black and White Cattle with random regression test-day models. J. Dairy Sci. 82:2805-2810.
- Reents, R., J. C. M. Dekkers and L. R. Schaeffer. 1995. Genetic evaluation for somatic cell score with a test day model for multiple lactations. J. Dairy Sci. 78:2858-2870.
- Van der Werf, J. H. J., M. E. Goddard and K. Meyer. 1998. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. J. Dairy Sci. 81:3300-3308.