

GENETIC ANALYSIS OF MALE AND FEMALE FERTILITY USING LONGITUDINAL BINARY DATA

by

TRAVIS AVERILL

(Under the Direction of Romdhane Rekaya)

ABSTRACT

A permanent effect and two random regression models were used to analyze insemination events. In all three cases, longitudinal threshold models were implemented. In the first study, outcomes of insemination events in the first 250 d after calving were analyzed using a permanent effect model. The posterior mean (SD) of additive, service sire and permanent effect variance was 0.034 (0.006), 0.009 (0.001) and 0.171 (0.013), respectively. The PM (SD) of the estimated heritability was 0.028 (0.005). In the second study, all insemination events were used and a quadratic function and Ali -Schaeffer model were employed to model the additive genetic effects. With random regression models, genetic variance and heritability for insemination success increased with time. Genetic correlations between successive inseminations were positive and high and decreased with the increase of the interval between inseminations. Model comparison based on Bayes factor showed a slight superiority of the quadratic regression model.

INDEX WORDS: Male and female fertility, longitudinal binary data, random regression, dairy

GENETIC ANALYSIS OF MALE AND FEMALE FERTILITY USING LONGITUDINAL
BINARY DATA

by

TRAVIS AVERILL
B.S., MICHIGAN STATE UNIVERSITY, 2002

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment
of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2004

© 2004

TRAVIS AVERILL

All Rights Reserved

GENETIC ANALYSIS OF MALE AND FEMALE FERTILITY USING LONGITUDINAL
BINARY DATA

by

TRAVIS AVERILL

Major Professor: Romdhane Rekaya

Committee: J. Keith Bertrand
Ignacy Misztal

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2004

DEDICATION

I would like to dedicate this work to my Mom and Dad for their support and guidance they have provided over the years.

ACKNOWLEDGEMENTS

I would like to express my great appreciation to Romdhane Rekaya for his assistance, friendship, and guidance. It has been an experience and honor to work with him these last two years. I would also like to thank Dr. Bertrand for the experience and knowledge he has provided. Also, I would like to thank Dr. Misztal for his perspective on things.

I would like to express special thanks to graduate students of the Animal and Breeding group: Kelly, Matt, Robyn, Oseni, Jarmila, Nathan, Chyong-Huoy, Shogo, Jesus, Andreas, and Kath. These past two years have been an interesting experience and journey.

In addition, I would like to thank Mike Kelly, for his assistance and friendship. A special thanks to Linda and Dr. Heusner for their friendship and assistance. Also, I would like to thank Dr. McPeake for his assistance, friendship, and guidance.

Finally, a special thanks to my family. Without the support and guidance, goals would never have been achieved. Mom and Dad thanks for being there through the lows and highs. Also, I would like to thank James for his support and guidance provided over the years.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER	
1 INTRODUCTION	1
2 REVIEW OF LITERATURE	3
3 GENETIC ANALYSIS OF MALE AND FEMALE FERTILITY USING LONGITUDINAL BINARY DATA	19
4 RANDOM REGRESSION MODELS FOR MALE AND FEMALE FERTILITY EVALUATION USING LONGITUDINAL BINARY DATA	37
5 CONCLUSIONS.....	65

LIST OF TABLES

	Page
Table 3.1: A descriptive summary of the edited data	35
Table 3.2: Summary of posterior distributions of the variance components and heritability	36
Table 4.1: Summary of posterior distributions of the genetic (co)variances between the three random coefficients, and permanent and service sire variances: Model 1	56
Table 4.2: Posterior means of genetic correlations between random coefficients and associated standard errors: Model 1	57
Table 4.3: Posterior means of genetic variances (SD) of insemination success, genetic correlations between specific time points, and associated standard deviations: Model 1	58
Table 4.4: Summary of posterior distributions of the genetic (co)variances between the five random coefficients, and permanent and service sire variances: Model 2	60
Table 4.5: Posterior means of genetic correlations between random coefficients and associated standard errors: Model 2	61
Table 4.6: Posterior means of genetic variances (SD) of insemination success, genetic correlations between specific time points, and associated standard deviations: Model 2	62
Table 4.7: Probability of conception after first insemination, probability of no conception after 3 inseminations and probability of no conception after 365 days for the best and worst five cows: Model 1	64

LIST OF FIGURES

	Page
Figure 4.1: Heritability of insemination success as a function of time using model 1.....	59
Figure 4.2: Heritability of insemination success as a function of time using model 2.....	63

CHAPTER 1

INTRODUCTION

Long time selection for production responses has deteriorated some secondary or non-production traits such as fertility or health conditions as a result of their antagonist relationships. These traits are not only of economic interest but also of environmental and animal welfare importance. Thus, improvement in or at least the maintenance of herd fertility has become a major objective in genetic improvement programs worldwide. Poor fertility and cows with low milk yield are the two major reasons for voluntary culling. Further, milk production depends at least in part on fertility, because the more frequently a cow calves, the greater the amount of milk produced in her lifetime.

The outcome of an insemination event depends on both male and female fertility, and it has been well documented that several factors, such as herd management, heat detection practices, and genetics can improve reproductive performance of dairy cows. Although the heritability of fertility traits is generally low, ranging from one to ten percent depending on definition of the trait and the methodology used for analysis, there is a consensus that sufficient genetic variability exists and can be exploited to improve reproductive performance. Several types of traits are used in fertility evaluation, ranging from binary (discrete) responses to continuous or interval traits. Consequently, depending on the trait definition, different models and methodologies have been developed and implemented to analyze reproductive performances.

Current models and techniques used in fertility studies have concentrated on evaluating female and male fertility separately. More importantly, not all sources of variation are accounted

for using any of the three types of traits mentioned before. Especially, there is no easy way to account appropriately for the service bull effect or male fertility. In the majority of cases, cows need more than one insemination per conception and at best, it is only possible to account for the last service bull used ignoring previous ones. Furthermore, it has been reported that almost 50% of the usable data is being discarded by considering only first service data because at least half of the cows have repeat insemination data available. Additionally, a high proportion of fertility data is usually discarded as a result of the inconsistencies in data recording that necessitate stringent editing. Thus, including the repeated records will increase the amount of information leading to more accurate sire evaluations.

Another issue of interest is the sequence of service sires used for cows having more than one insemination. For two cows having the same number of inseminations and sired by the same bulls but in different order is not accounted for with actual models, which may bias the prediction of both male and female fertility. However, by using all insemination events, all service sires and the order they were used can be accommodated easily. For a cow that conceived after 3 inseminations, her vector of responses will be (0, 0, 1). Similarly, a cow that conceived after 1 insemination will have a vector of binary response of (1). An alternative idea consists in modeling all insemination events for each cow as repeated binary responses allowing on one hand to use all available information and on another hand to account properly for male and female fertility.

The main objective of this study was to develop and implement a longitudinal threshold model for analysis of the binary outcome of all insemination events. Repeatability and random regression based models will be implemented and compared using Bayesian Information criteria.

CHAPTER 2

REVIEW OF LITERATURE

During the last few decades, the dairy industry has changed dramatically in the United States and worldwide. According to the USDA (National Agricultural Statistics Service, 2004), a 41% decrease in the number of dairy operations occurred between 1993-2002, yet the number of dairy cows decreased only by four percent from 1994-2003. During the same period, total milk production increased by 11%, which translates to 16% increase in milk production per cow. The increase of milk production was the result of a combination of improved management and nutrition, intense genetic selection and biotechnology (Rajala-Schultz and Frazer, 2003). However, during the same period, little attention has been given to health and fertility traits (Pryce et al., 2004), which has led to serious deterioration of these traits due to their antagonist relationships with milk yield (A-Ranberg et al., 2003). Poor reproductive performance has led to substantial economic losses as a result of prolonged calving intervals, increased insemination and veterinary costs, higher culling rates, and increased replacement costs (A-Ranberg et al., 2003). Freeman (1984) reported that reproductive failures account for approximately 16% of disposals of all dairy cows in the United States. Thus, maintaining reasonable reproductive efficiency for dairy cows is becoming a challenging problem of great importance for the profitability of the dairy industry (Hayes et al., 1992; Britt, 1985).

Poor reproductive efficiency is a worldwide problem affecting the dairy industry. Although some attempts were made to link poor reproductive efficiency to the increase of Holstein genes in dairy populations, the high conception rates of non-lactating Holstein heifers

(70-80%), during a period when milk production per cow increased by 218%, indicates that the problem is more complex and that milk yield plays a major role (Beam and Butler, 1999).

Many studies have reported an antagonistic relationship between fertility traits and milk production (Abdallah and McDaniel, 2000; Roxstrom et al., 2001; A-Ranberg et al., 2003). In fact, their estimated genetic correlation ranged from 0.4 to 0.8. Thaller (1998) concluded that an expected decrease of three percent in reproductive efficiency per generation is due to the antagonistic relationship with milk yield. However, milk yield is not the only factor affecting reproductive efficiency. In fact, Lucy (2001) concluded that inbreeding has increased substantially in the United States Holsteins population since 1980 and might play a crucial role in fertility. Additionally, Wolfenson et al. (2000) argued that reproductive performance in lactating dairy cows is extremely sensitive to heat stress. Oseni et al. (2004) reported similar results.

Until recently, fertility traits were not seriously accounted for in most breeding programs for several reasons including the lack of unified definition for reproductive performance, lack of efficient recording system, and the theoretical and computational complexity in modeling and analyzing such data. Pryce et al. (2004) stated that the limitation in using insemination data is, in part, because of the considerable variation in the recording quality.

The outcome of an insemination depends on both male and female fertility, and it has been well documented that several factors, including herd management, heat detection practices, and genetics play a major role in reproductive performance. Although the heritability of fertility traits are low, ranging from one to ten percent depending on the definition of the trait and the methodology used for analysis, there is consensus that sufficient genetic variability exists, and it can be exploited to improve reproductive performances.

For the joint evaluation of male and female fertility, the definition of the trait is crucial. Consequently, it is appropriate to discuss the most frequent definitions used to assess fertility. Basically there are three type of traits used for reproductive performance evaluation: interval traits, discrete traits and count traits. Depending on the trait definition, different models and methodologies have been developed and implemented to analyze reproductive performance.

Interval traits

Interval or continuous traits are most commonly used for fertility evaluation, in part because their simplicity and availability at a large scale. Further, their analysis can be accommodated easily using existing standard tools, particularly mixed linear model methodology. Most interval traits such as days to first service (DFS), calving interval (CI), and days open (DO) are likely to be biased by management decisions influenced by the potential yield or season of calving of individual cows (Stott et al., 1999; Butler and Smith, 1989; Darwash et al., 1997). For the interval from first insemination to conception, high producing cows tend to have more opportunities for re-insemination in the case of failure of conception. Mayne et al. (2002) reported that herds with high heat detection rate had significantly shorter calving intervals and significantly lower 305-day protein yields, less body condition loss after calving, and significantly smaller negative energy balance. They concluded that calving interval shorter than 380 days is achievable by minimizing negative energy balance in early lactation, good heat detection, and early insemination of cows after calving.

Given the non-uniformity and the effects of management decisions, interval traits (e.g. calving interval, days to first insemination, days open) tend to have low heritabilities (0.003 to 0.12), however, most estimates were around 0.04, which makes the identification of more fertile genotypes difficult and have consequently negated selection for fertility (Darwash et al. 1997).

Interval from calving to first breeding is regarded as one of the most important practical measures of reproductive performances. It was argued that economic losses due to poor fertility are the result of prolonged calving interval, increased insemination cost, reduced returns from calves born and forced replacement in the event of culling (Esselmont and Peeler, 1993). Using calving interval as a measure of fertility presents a problem because only animals that survive to the next lactation have a calving interval. Therefore, evaluation based on this trait alone could be biased as a result of culling of low fertility cows. To deal with the culling problem, it was recommended (Roxstrom and Strandberg, 2002; Olori et al. 2002) that calving interval should be treated as a censored trait and analyzed jointly with survival scores to take into account the nonrandom scoring of calving interval.

Weller (1989) used a large data set from the Israeli Holstein population to estimate the genetic parameters of days to first breeding in the first and second parities. Heritability of days to first breeding was 0.048 and 0.031 for first and second parity, respectively. The genetic and environmental correlation between days to first breeding in the first and second parity was 0.732 and 0.061, respectively. The genetic correlation between days to first breeding and production traits ranged between 0.2 and 0.3. However, these results have to be interpreted with caution because the nonrandom selection applied to the data (only cows with records in both parities were considered).

Using Holstein data from New Zealand, the heritability for calving interval was 0.017 (Grosshans et al. 1996). This estimate is relatively low compared with other estimates. However, it is still within the range of reported estimates for the trait. The genetic correlations with production traits ranged from 0.026 to 0.22. Although lower than reported estimates, it still reflects the antagonistic relationship between reproductive performance and production traits. In

a more recent study, Olori et al. (2002) used single and multiple-trait models to analyze calving interval, survival rate and milk yield. The estimated heritability for calving interval in the three-trait model was 0.04 and the genetic correlation with milk yield was 0.4. The distribution of calving interval proofs from the single and multiple-trait models were relatively similar.

However, a left shift on the distribution of calving interval proves suggests that analyzing calving interval alone, or along with survival rate without correcting for milk yield may result in overestimation of breeding values for calving interval. Furthermore, the trend indicated a steady increase in calving interval at the rate of 0.14 days per year from 1984 to 1995.

Hansen et al. (1983) stated that the relationships between 305-day yield and days open are largely due to the influence of gestation length on yield. However, voluntary delay of breeding has a major impact on DO leading to some unreasonable values for the trait. Thus, large DO reflect not only the reproductive performance but also preferential husbandry. Conceptually, observed DO values cannot be used directly for fertility genetic evaluation. In fact, several restrictions based on truncation or thresholds have been developed to minimize the impact of management decisions on the trait. Currently, the USDA uses a threshold on large DO at 250 days. Oseni et al. (2004) showed that different values for the threshold have resulted in major changes in the estimated genetic parameters. On the other hand, DO present several advantages compared to other fertility measurements. DO is a continuous trait for which analysis tools are widely available, it is easy and cost effective to measure on a large scale and has a higher heritability. Moore et al. (1990) argued that days open would be preferred due to the lack of correlation with age and weight at calving, higher heritability and more variation than days to first breeding.

Heritability estimates for days to first breeding and days open are relatively higher than other reproductive traits. Available estimates ranged from almost zero to 0.15 and even higher. However, the majority of estimates ranged between 0.03 and 0.05 (Brotherstone et al. 2002; Raheja et al. 1989a; Dematawewa et al. 1998; Hoeschele 1991).

Oseni et al. (2004) studied DO using five different upper bounds (150, 200, 250, 300, 365 days). A large data set including DO records from several states was used. By relaxing the upper bound from 150 to 365 days, DO mean increased by 25 to 50% depending on the state. Similarly, the residual and genetic variances for DO increased with the increase of the upper bound. The estimated heritability ranged from 0.03 to 0.06. Genetic correlations between days open and production traits appear to be antagonist with estimates ranging from close to zero (Weller 1989; Raheja et al., 1989a; Raheja et al., 1989b) to relatively high 0.53-0.68 (Dematawewa and Berger 1998). In addition, Marti et al. (1994) studied the effect of milk yield and parity on DO. The regression coefficient of days open on milk production indicated an increase of around one and half days in DO for each additional 100 Kg in milk yield.

Binary traits

The recent theoretical and computational developments in the analysis of discrete and binary data have made the use of these traits for reproductive performance evaluation possible. Although these traits were analyzed for a period of time as continuous responses using existing mixed linear methodology, in clear violation of their distributional assumptions, the threshold liability model is becoming the standard tool for discrete data analysis especially after the papers by Best et al. (1995) and Sorensen et al. (1995). There are several binary traits used as a measure of fertility. However, conception rate (CR) and non-return rate (NRR) after a fixed number of days (e.g. 70 or 90 days) and success or failure of an insemination were the most widely used

binary traits found in the literature. All these traits are in some sense connected and tend to have lower heritability especially when analyzed using mixed linear models. Compared to interval traits, binary responses are usually measured early during the breeding season which will reduce the impact of environmental and management effects and allow for early genetic evaluation (Thaller, 1998). Non-return rate has mostly been used as a measure of male fertility, and it depends on complete recording of all subsequent inseminations. Furthermore, a calving date is not required. More recently, Weigel and Rekaya (2000) used non-return at day 70 (NR70) and 90 (NR90) for the joint evaluation of male and female fertility. The cow being inseminated was considered as the additive genetic effect (female fertility) and the service bull (male fertility) was considered as a random environmental effect (representing a combination of genetic and permanent environmental effects). They used a linear and threshold model for estimating genetic parameters for NR and confirmed pregnancy at day 60 and 90 in dairy herds in California and Minnesota. The linear model estimates were lower than the threshold model estimates except for phenotypic variance in Minnesota herds. They argued that the phenotypic variance in Minnesota herds could be the result of small herd size in Minnesota compared to California herds and to the use of herd-season as contemporary group instead of herd-month. The correlations between NR and confirmed pregnancy at day 60 (CP60) were high, ranging from 0.88 to 0.92 and lower between NR and CP90 (0.66 to 0.84) indicating a potential error (misclassification) of pregnancy status. Further, the lowest NR was observed in winter (November to February) and the highest in spring (May to June). Schnyder and Stricker (2002) used a bivariate linear animal model to estimate variance components for days to first breeding and non-return rate in three breeds (Holsteins, Red and Whites, and Brown Swiss). The estimated genetic variance and service sire variance ranged from 0.007 to 0.011 and 0.0013 to 0.0023, respectively. Heritability estimates

were less than four percent for all three breeds with the lowest of 0.03 for the Red and Whites breed. The genetic and residual correlation between NR and interval from calving to first insemination ranged from 0.22 to 0.31 and -0.06 to 0.04 , respectively. A-Ranberg (2002) reported a 3.7 and 2.8% increase in NR rate for heifers and first lactation cows, respectively during a twenty-year period (1979-2000). His estimates of service sire variance were practically null ($0.0005 - 0.0008$), and the obtained heritabilities were much smaller ($0.012 - 0.015$) than those reported by Schnyder and Stricker (2002).

The major problem with NR as a measure of reproductive performance was highlighted by Mayne et al. (2002) who reported that the average conception rate to first AI was 37.1% or 16% less than the CR estimated from the 60 day NRR in the same herds. The difference illustrates the unreliability of accepting the 60 day NRR as proof of a successful mating. Taylor et al. (1985) used conception rate as a measure of fertility performance and found that both the additive (0.016) and service sire (0.005) variances were low compared to those obtained using NR. Further, their results suggest higher conception rate in fall months compared with winter months, and the largest difference, 6.1% was observed between October and January.

Jansen (1986) used a linear model to estimate genetic parameters for NR56 and conception rate for heifers and cows. Heritability estimates were relatively similar for NR56 and CR, although few discrepancies were observed. NR56 heritability ranged from 0.1 to 2.3%. Parity one had the highest heritability for NR56 and parity two had the highest heritability for CR. Hodel et al. (1995) analyzed NR on heifers and cows using a bivariate model and reported higher service sire variance, additive variance and heritability using cow data. They concluded that the maximum fertility appears to be achieved approximately 120 days after calving and that inseminating cows before 45 days after calving is not advisable, because the uterus requires a

longer recovery period. Clay and McDaniel (2001) reported that cows bred within 50 days after calving are expected to have 5.5% greater chance of being rebred within 70 days than cows bred between days 70 to 79 after calving. Similarly, cows bred at more than 139 days were expected to have 3.3% less chance of being rebred within 70 days than cows bred at 70 to 79 days.

Count Traits

The only count trait used, as a measurement of fertility, was the number of services to conception (NS). If inseminations are conducted at regular intervals, it reflects the ability of the cow to start cycling after calving and her potential of becoming pregnant. However, it is seldom the case that inseminations are carried out at regular intervals. Furthermore, censoring is a major problem when analyzing NS as too many cows have incomplete records. Additionally, NS is not a continuous trait and its analysis requires special methodology (poisson models) and software. As a result, only a few studies have looked at this trait. Raheja et al. (1989a) reported that NS ranged from 1.54 to 1.55 for the first three lactations. Heritability estimates increased with lactation number from 0.03 to 0.06 for first and third lactation, respectively. In another study, Raheja et al. (1989b) used a mixed linear model to analyze NS as a measure of female fertility and NR as a measure of male fertility. The estimated genetic and phenotypic correlation between these two traits was -0.09 and -.012, respectively.

Dematawewa and Berger (1998) used a repeatability model for NS across lactations and reported estimates of heritability of 0.028 and repeatability of 0.083. Heritability estimates from across parity analyses ranged from 0.01 (second lactation) to 0.11.

Foulley et al. (1987) were first in developing a poisson model for analysis of count data. Tempelman and Gianola (1996) used such a model for analysis of NS in Holstein heifers. They reported an estimated heritability of 2.6%, similar to estimates of other discrete fertility traits.

Longitudinal analysis of all insemination events

Longitudinal analysis of all insemination events during a breeding season was never implemented to the best of our knowledge. With respect to analysis of the three type of traits mentioned earlier, not all sources of variation are accounted for by the various models employed. Especially, there is no easy way to account appropriately for the service sire effect or male fertility. In the majority of cases, cows need more than one insemination per conception and at best, it is only possible to account for the last service sire used ignoring previous ones. For many years, DRMS Raleigh has provided a regional evaluation for male fertility through the Expected Relative Conception Rate (ERCR) system using only 70-day nonreturn of first breedings (Clay and McDaniel, 2001; Weigel, 2000). However, Weigel (2000) reported that almost 50% of the usable data is discarded by considering only first service data because at least half of the cows have repeat insemination data available. The same author concluded that the use of such information is desirable. Furthermore, a high proportion of the fertility data is usually discarded as a result of the inconsistencies in data recording that necessitate stringent editing. Thus, including repeated records will increase the amount of information leading to more accurate sire evaluations. Another issue of interest is the sequence of service sires used for cows having more than one insemination. For two cows having the same number of inseminations and sired by the same bulls but in different order, such sequence of sires used is not accounted for with actual models, which may bias the prediction for both male and female fertility. For a cow that conceived after 3 inseminations, her vector of responses will be (0, 0, 1). Similarly, a cow that conceived after 1 insemination will have a vector of binary response of (1). An alternative idea consists in modeling the number of inseminations per cow as repeated binary responses allowing the use all available information to account properly for the male and female fertility.

CONCLUSION

Fertility is a complex trait that is becoming more and more important for many genetic improvement programs in dairy cattle worldwide. Long-term single trait selection for increased milk production had a negative impact on several secondary traits including fertility.

Most interval traits used as a measure of fertility are influenced by management decisions and production level, which potentially could lead to non-ignorable bias. Some discrete or binary traits are less affected by management decisions and could potential help improve reproductive performances in dairy cattle. However, the theoretical and computational complexities associated with their use and the non-availability of nation wide recording of those traits could limit their extensive use.

Reproductive performance depends on male and female fertility. As such, these two components have been modeled jointly. Unfortunately, neither the interval traits nor the binary traits can accommodate these two components in a joint analysis. Therefore, longitudinal data is the only choice available now for the joint analysis of male and female fertility.

REFERENCES

- Abdallah, J. M., and B. T. McDaniel. 2000. Genetic parameters and trends of milk, fat, days open and body weight after calving in North Carolina experimental herds. *J. Dairy Sci.* 83:1364-1370.
- A.-Ranberg, I. M., G. Klemetsdal, B. Heringstad, M. Svendsen, and T. Steine. 2002. Breeding values for daughter fertility in Norwegian dairy cattle; data quality and model validation. *INTERBULL Bulletin* 29:200-204.

- A.-Ranberg, I. M., B. Heringstad, G. Klemetsdal, M. Svendsen, and T. Steine. 2003. Heifer fertility in Norwegian dairy cattle: Variance components and genetic change. *J. Dairy Sci.* 86:2706-2714.
- Beam, S. W., and W. R. Butler. 1999. Effects of energy balance on follicular development and first ovulation in postpartum dairy cows. *J. Reprod. Fert. Suppl.* 54:411-424.
- Best, N. G., M. K. Cowles, and S. K. Vines. 1995. CODA Manual version 0.30. MRC Biostatistics Unit, Cambridge, UK. 41pp.
- Britt, J. H. 1985. Enhanced reproduction and its economic implications. *J. Dairy Sci.* 68:1585-1592.
- Brotherstone, S., G. Banos, and M. P. Coffey. 2002. Evaluation of yield traits for the development of a UK fertility index for dairy cattle. 7th World Congress on Genetics Applied to Livestock Production. 01-28.
- Butler, W. R. and R. D. Smith. 1989. Interrelationships between energy balance and postpartum reproductive function in dairy cattle. *J. Dairy Sci.* 72:767-783.
- Clay, J. S., and B. T. McDaniel. 2001. Computing mating bull fertility from DHI nonreturn data. *J. Dairy Sci.* 84:1238-1245.
- Darwash, A. O., G. E. Lamming, and J. A. Woolliams. 1997. Estimation of genetic variation in the interval from calving to postpartum ovulation of dairy cows. *J. Dairy Sci.* 80:1227-1234.
- Dematawewa, C. M. B., and P. J. Berger. 1998. Genetic and phenotypic parameters for 305-day yield, fertility, and survival in Holsteins. *J. Dairy Sci.* 81:2700-2709.
- Esslemont, R. J., and E. J. Peeler. 1993. The scope for raising margins in dairy herds by improving fertility and health. *British Veterinary J.* 149:537-547.

- Foulley, J. L., D. Gianola, and S. Im. 1987. Genetic evaluation of traits distributed as Poisson-binomial with reference to reproductive characters. *Theoretical and Applied Genetic*. 73:870-877.
- Freeman, E. A. 1984. Sire selection. Secondary traits: sire evaluation and the reproductive complex. *J. Dairy Sci.* 67:449.
- Grosshans, T., Z. Z. Xu, L. J. Burton, and D. L. Johnson. 1996. Genetic parameters for fertility traits in seasonal dairy cattle. *Dairy Cattle Breeding in New Zealand*.
- Hansen, L. B., A. E. Freeman, and P. J. Berger. 1983. Variances, repeatabilities, and age adjustments of yield and fertility in dairy cattle. *J. Dairy Sci.* 66:281-292.
- Hayes, J. F., R. I. Cue, and H. G. Monardes. 1992. Estimates of repeatability of reproductive measures in Canadian Holsteins. *J. Dairy Sci.* 75:1701-1706.
- Hodel, F., J. Moll, and N. Kuenzi. 1995. Analysis of fertility in Swiss Simmental cattle – Genetic and environmental effects on female fertility. *Livest. Prod. Sci.* 41:95-103.
- Hoeschele, I. 1991. Additive and nonadditive genetic variance in female fertility of Holsteins. *J. Dairy Sci.* 74:1743-1752.
- Jansen, J. 1986. Direct and maternal genetic parameters of fertility traits in Friesian cattle. *Livest. Prod. Sci.* 15:153-164.
- Lucy, M. C. 2001. Reproductive loss in high-producing dairy cattle: where will it end? *J. Dairy Sci.* 84:1277-1293.
- Marti, C. F., and D. A. Funk. 1994. Relationship between production and days open at different levels of herd production. *J. Dairy Sci.* 77:1682-1690.

- Mayne, C. S., M. A. McCoy, S. D. Lennox, D. R. Mackey, M. Verner, D. C. Catney, W. J. McCaughey, A. R. G. Wylie, B. W. Kennedy, F. J. Gordon. 2002. Fertility of dairy cows in Northern Ireland. *Veterinary Record*. 150:707-713.
- Moore, R. K., B. W. Kennedy, L. R. Schaeffer, and J. E. Moxley. 1990. Relationships between reproduction traits, age and body weight at calving, and days dry in first lactation Ayrshires and Holsteins. *J. Dairy Sci.* 73:835-842.
- National Agricultural Statistics Service 2004. USDA-NASS Milk production and milk cows. Online: <http://www.usda.gov/nass/aggraphs/milk1.htm>.
- Olori, V. E., T. H. E. Meuwissen, and R. F. Veerkamp. 2002. Calving interval and survival breeding values as measure of cow fertility in a pasture-based production system with seasonal calving. *J. Dairy Sci.* 85:689-696.
- Oseni, S., I. Misztal, S. Tsuruta, and R. Rekaya. 2004(submitted). Genetic parameters for days open and pregnancy rate for US Holsteins using different editing criteria. *J. Dairy Sci.*
- Pryce, J. E., M. D. Royal, P. C. Garnsworthy, and I. L. Mao. 2004. Fertility in the high-producing dairy cow. *Livest. Prod. Sci.* 86:125-135.
- Raheja, K. L., E. B. Burnside, and L. R. Schaeffer. 1989a. Relationships between fertility and production in Holstein dairy cattle in different lactations. *J. Dairy Sci.* 72:2670-2678.
- Raheja, K. L., K. Nadarajah, and E. B. Burnside. 1989b. Relationship of bull fertility with daughter fertility and production traits in Holstein dairy cattle. *J. Dairy Sci.* 72:2679-2682.
- Rajala-Schultz, P. J., and G. S. Frazer. 2003. Reproductive performance in Ohio dairy herds in the 1990s. *Anim. Reprod. Sci.* 76:127-142.

- Roxstrom, A., E. Strandberg, B. Berglund, U. Emanuelson and J. Philipsson. 2001. Genetic and environmental correlations among female fertility traits, and between the ability to show oestrus and milk production in dairy cattle. *Acta. Agric. Scand., Sect. A, Anim. Sci.* 51:192-199.
- Roxstrom, A., and E. Strandberg. 2002. Genetic analysis of functional, fertility-, mastitis-, and production-determined length of productive life in Swedish dairy cattle. *Livest. Prod. Sci.* 74:125-135.
- Schnyder, U., and C. Stricker. 2002. Genetic evaluation for female fertility in Switzerland. *INTERBULL Bulletin.* 29:138-141.
- Sorensen, D. A., S. Andersen, D. Gianola, and I. Korsgaard. 1995. Bayesian inference in threshold using Gibbs sampling. *Genet. Sel. Evol.* 27:229-249.
- Stott, A. W., R. F. Veerkamp, and T. R. Wassell. 1999. The economics of fertility in the dairy herd. *Anim. Sci.* 68:49-57.
- Taylor, J. F., R. W. Everett, and B. Bean. 1985. Systematic environmental, direct, and service sire effects on conception rate in artificially inseminated Holstein cows. *J. Dairy Sci.* 68:3004-3022.
- Tempelman, R. J. and D. Gianola. 1996. A mixed effects model for overdispersed count data in animal breeding. *Biometrics.* 52: 265-279.
- Thaller, G. 1998. Genetics and breeding for fertility. *INTERBULL Bulletin.* 18:55-61.
- Weigel, K. A. 2000. Toward national fertility evaluations. *Proc 18th Technical Conference on Artificial Insemination and Reproduction.*
- Weigel, K. A. and R. Rekaya. 2000. Genetic parameters for reproductive traits of Holstein cattle in California and Minnesota. *J. Dairy Sci.* 83:1072-1080.

Weller, J. I. 1989. Genetic analysis of fertility traits in Israeli dairy cattle. *J. Dairy Sci.* 72:2644-2650.

Wolfenson, D., Z. Roth, and R. Meidan. 2000. Impaired reproduction in heat-stressed cattle: basic and applied aspects. *Anim. Reprod. Sci.* 60-61:535-547.

CHAPTER 3

GENETIC ANALYSIS OF MALE AND FEMALE FERTILITY USING LONGITUDINAL BINARY DATA¹¹

Averill, T., R. Rekaya, and K. Weigel

¹ Averill, T., R. Rekaya, and K. Weigel. Submitted to J. Dairy Science (2004).

ABSTRACT

A longitudinal Bayesian threshold analysis of insemination events during the first 250 d after calving of first parity Holsteins was carried out. The outcome of an insemination event was treated as a binary response of either a success (1) or a failure (0). Thus, all breeding information for a cow, including all service sires, was included, thereby allowing for a joint evaluation of male and female fertility. An edited data set of 297,827 insemination records from 151,758 first lactation cows was used. On the liability scale, the model included the systematic effects of herd-year of insemination, technician, month of insemination, and regressions on age of service sire, three test days in the first 100 days of lactation (early milk yield) and days in milk at insemination. The random effects in the model were the additive breeding value, the permanent effect of the cow and the service sire effect. Posterior mean (SD) of the dispersion parameters in the model were 0.034 (0.006), 0.009 (0.001) and 0.171 (0.013) for the additive, service sire and permanent environmental variances, respectively. The residual variance was fixed to one, as a result of the non-identifiability of the threshold model. The posterior mean (SD) of heritability was 0.028 (0.005). This point estimate of heritability is well within the range of available estimates for the trait. Thus, these estimates suggest that some genetic variation exists that can potentially be used to improve, or at least avoid further deterioration of reproductive performance. The estimate of the regression coefficient on age of service sire was 0.001 indicating better fertility among older bulls. However, this result has to be interpreted with caution given the preferential use of proven bulls on well-managed cows (as opposed to problem breeders). The estimate of the regression coefficient was negative (-0.005) for early milk yield, as expected, and positive (0.003) for days in milk at insemination. This suggests that high producing cows are less likely to conceive at the beginning of lactation.

Key words: Male and female fertility, longitudinal binary data, dairy

INTRODUCTION

The main goal of dairy producers is to maximize returns from milk production while minimizing costs (Freeze et al., 1992). This forces dairy producers to maintain reproductive performance while obtaining a high level of milk production. However, long-term selection for milk yield has caused deterioration in some non-production or “secondary” traits as a result of antagonist genetic relationships. As a result, more cows are being culled involuntarily as a consequence of poor fertility (Hansen et al., 1983; Lopez-Gatius et al., 2002; Weller, 1989). This decline in fertility has become a major problem in the dairy industry, causing significant economic losses (Clay et al., 2001; Taylor et al. 1985; Olori et al., 2002) due to additional insemination costs, decreased milk production per day of herd life, and decreased intensity of selection (Hodel et al., 1995). Thus, improvement in, or at least the maintenance of, herd fertility has become a major objective in genetic improvement programs worldwide. The outcome of an insemination depends on both male and female fertility, and it has been well documented that several factors, such as herd management, heat detection practices, and genetics can improve reproductive performance. Although the heritability of fertility is low, ranging from one to ten percent depending on definition of the trait and the methodology used for its analysis, there is a consensus that sufficient genetic variability exists, and this can be exploited to improve reproductive performance.

Several types of traits are used in fertility evaluation, ranging from binary (discrete) responses to continuous or interval traits. Consequently, depending on the trait definition, different models and methodologies have been implemented to analyze reproductive performance. Raheja et al. (1989b) used a mixed linear model to study the relationships between

bull fertility with daughter fertility and production traits in Holstein cattle. Three fertility traits (days from calving to first breeding, days open, and number of AI services per conception) and three production traits (breed class average-milk, breed class average-fat, and breed class average-fat %) in the first two lactations were used. Breed Class Average (BCA) is an index combining milk, fat and protein used by all milk recording programs across Canada. It consists of comparing a cow's actual or projected 305 day production to the BCA standard for a cow of the same breed, age at calving and month of calving. The estimated genetic correlations (based on the correlations of estimated breeding values) between female fertility, male fertility, and production traits ranged between -0.56 to 0.13 . Weigel and Rekaya (2000) estimated genetic parameters associated with male and female fertility traits using linear and threshold models. Non-return rate and confirmed pregnancy after 60 and 90 days defined as binary traits were considered. The proportion of phenotypic variance explained by service bull effect ranged from 0.005 to 0.008 . In all cases, estimates based on a threshold model were higher. Heritability estimates for female fertility (i.e., the proportion of variance due to the animal being inseminated) ranged from 1.4 to 3.1% . In a recent study, Brotherstone et al. (2002) investigated the relationships between three fertility traits (calving interval, number of inseminations, and conception rate to first insemination) and nine production traits (daily milk, fat plus protein, or milk energy yield at the 3rd test; daily milk, fat plus protein, or milk energy yield at the test nearest to the body condition score date, and, 305-d lactation milk, fat plus protein, and milk energy yield). They concluded that heritabilities of fertility traits were very similar and ranged from 0.03 to 0.04 . However, the genetic correlations between fertility and production traits were antagonist, ranging from -0.34 to 0.43 . The genetic correlation between calving interval and 305-

d milk energy yield (0.43) suggests that an increase in 305-d milk yield is associated with an increase in calving interval.

Grosshans et al. (1996) used a sire model and restricted maximum likelihood methodology to analyze eleven fertility traits and three production traits (milk yield, fat yield, and protein yield) and concluded that the heritability of fertility traits ranged from 0.007 (number of service per conception) to 0.134 (age at calving). Genetic correlations between milk production and fertility traits, except age at first calving, were low (-0.248 to 0.289). However, the correlations of age at calving with production traits were higher and positive (antagonist) being 0.209, 0.447, and 0.704 with milk yield, fat yield and protein yield, respectively, indicating an increase in age at calving with an increase in milk production.

Kadarmideen et al. (2000) estimated genetic parameters for various disease traits and conception after first insemination in UK dairy population using linear and threshold models. They concluded that the threshold model yields slightly higher estimates. In fact, the heritability of the fertility trait was 0.01 and 0.012 using linear and threshold models, respectively.

Most of the previous research on dairy cattle fertility has focused on separate analyses of female and male fertility. More importantly, not all sources of variation are accounted for in many analyses, particularly when only one record per cow is used. Weigel (2000) reported that almost 50% of the usable data is discarded by considering only first services, because at least half of the cows have repeated insemination data available. The same author concluded that the use of such information is desirable. Furthermore, a high proportion of the fertility data is usually discarded as a result of inconsistencies in data recording that necessitate stringent editing. Thus, including the repeated records will increase the amount of information leading to more accurate sire evaluations. However, when using a single record per cow, there is no easy way to

account appropriately for the service sire effect because only one service sire will be accounted for. In the majority of cases, cows need more than one insemination per conception, and it is inappropriate to account for one of the service bulls and ignore all others. Another issue of interest is the sequence of service sires for cows having more than one insemination. For two cows having the same number of inseminations to the same bulls (but in a different sequence), differences in sequence are not accounted for with most models, and this may bias the prediction of both male and female fertility. An alternative method consists of modeling the number of inseminations per cow as repeated binary responses, thereby allowing the use of all available information and accounting properly for all factors affecting male and female fertility.

The objective of this study was to develop and implement a longitudinal binary model for the genetic evaluation of male and female fertility while making use of all available information.

MATERIALS AND METHODS

The data consisted of insemination records from dairy herds in California, generated between 2002 and 2003 and provided by AgriTech Analytics. The trait of interest was the outcome of all insemination events (1 = success, 0 = failure). Thus, if a cow conceived after three inseminations, her response vector was (0, 0, 1). If a cow conceived after one insemination, her response vector was (1). Only first parity cows were included in the analysis of 297,823 records. Data editing consisted of keeping only pregnant cows (those with a successful insemination as verified by a veterinary pregnancy examination) and non-pregnant cows that had at least three insemination events within an opportunity period of 250 days after calving. Further, cows with inconsistent identification, more than one successful insemination per lactation, or an unrealistic interval between consecutive inseminations were removed. After editing, the data consisted of a total of 151,758 cows with 297,823 records, averaging 1.96

inseminations per cow. The data also included a total of 2179 technicians, 3190 AI service sires, and 990 herd-year contemporary groups. The averages for number of inseminations per technician and service sire were 137 and 93, respectively. A more detailed summary of the data is presented in Table 3.1.

The pedigrees of these cows had a high proportion of missing or unknown sires, due to a high percentage of cows in California having missing sire identification. After matching the identification numbers in the data and pedigree files provided by AgriTech Analytics with the sire pedigrees from the National Association Animal Breeders, the international sire identification numbers were determined and used in building the pedigree. In general, the pedigree file was of poor quality and consisted of 159,250 animals.

Statistical model and implementation

The basic latent variable model for the analysis of cross-sectional binary responses in an animal breeding context has been available for more than twenty years (Gianola, 1982). The longitudinal threshold model, as presented by Gianola and Rekaya (2000) and Heringstad et al. (2001), is an extension of the cross-sectional model. As with the basic threshold model, it consists of postulating the existence of unobserved continuous random variable that relates to the observed binary or discrete response.

Let $\mathbf{y}_i = (y_{it_1}, y_{it_2}, \dots, y_{it_l})'$ be a $n_i \times 1$ vector of binary responses for animal, ($i = 1, 2, \dots, q$) observed at times $t_1, t_2, \dots, t_{n_i}$. As in the cross-sectional analysis, the binary response observed at a time t_j related to an underlying random variable satisfying:

$$y_{ij} = \begin{cases} 1 & \text{if } l_{ij} > T \\ 0 & \text{otherwise} \end{cases}$$

where the subscript j represents time t_j and will be used as such from here on, and T is a threshold value. Further, it is assumed that

$$l_{ij} \sim N(\mu_{ij}, \sigma_e^2) \quad [1]$$

The probability of observing a positive case (success) is:

$$\begin{aligned} p_{ij} &= pr(l_{ij} > T | \mu_{ij}) \\ &= 1 - pr(l_{ij} < T | \mu_{ij}) \\ &= 1 - \Phi\left(\frac{T - \mu_{ij}}{\sigma_e}\right) \end{aligned} \quad [2]$$

where Φ is the cumulative distribution function of standard normal. It is clear from [2] that it is not possible to infer separately μ_{ij} , T and σ_e^2 . Hence, some restrictions are placed on two of the three model parameters. A common choice is to set $T=0$, and $\sigma_e^2=1$, leading to:

$$p_{ij} = pr(l_{ij} > T | \mu_{ij}) = 1 - \Phi(-\mu_{ij}) = \Phi(\mu_{ij}) \quad [3]$$

where μ_{ij} can be linearly related to a set of systematic and random effects.

Furthermore, a mixed linear model can be used to express the relationship between liability and μ_{ij} . In matrix notation, the model can be written as:

$$\mathbf{l} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{W}_1\mathbf{p} + \mathbf{W}_2\mathbf{s} + \mathbf{e}$$

where $\mathbf{l}$ is a vector of unobserved liabilities; $\boldsymbol{\beta}$ is the vector of systematic effects (herd-year of insemination, technician, month of insemination and regression coefficients on age of service sire (months), early milk yield (mean of test day records within the first 100 days of lactation) and days in milk to insemination); $\mathbf{u}$ is vector of additive breeding value; $\mathbf{p}$ is vector of permanent effects; $\mathbf{s}$ is vector of service sires; $\mathbf{e}$ is the vector of residual terms, and $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{W}_1$, $\mathbf{W}_2$ are known incidence matrices with the appropriate dimensions.

The reason for using herd-year instead of herd-year-month as contemporary group was the small number of records in the latter and, consequently, the large number of classes containing all successes or all failures.

Based on the assumptions made earlier, the conditional distribution of liabilities given the model parameters was:

$$p(\mathbf{l}|\boldsymbol{\beta}, \mathbf{u}, \mathbf{p}, \mathbf{s}) \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{W}_1\mathbf{p} + \mathbf{W}_2\mathbf{s}, \mathbf{I}) \quad [4]$$

For a full Bayesian implementation of [4], prior distributions for the model parameters are required. To avoid a potentially improper posterior distribution, the following priors were assumed:

A normal distribution with mean zero and a large variance was assumed as a prior for the systematic effects, $\boldsymbol{\beta}$:

$$p(\beta_i) \sim N(0, 10^4) \quad [5]$$

Multivariate normal distributions were assumed for all random effects in the model:

$$\begin{aligned} p(\mathbf{u} | \sigma_u^2) &\sim N(\mathbf{0}, \mathbf{A}\sigma_u^2) \\ p(\mathbf{p} | \sigma_p^2) &\sim N(\mathbf{0}, \mathbf{I}\sigma_p^2) \end{aligned} \quad [6]$$

$$p(\mathbf{s} | \sigma_s^2) \sim N(\mathbf{0}, \mathbf{I}\sigma_s^2)$$

where σ_u^2 , σ_p^2 and σ_s^2 were the additive, permanent environmental, and service sire variances, respectively.

For the three variances, a flat bounded prior was assumed:

$$p(\sigma_k^2) \sim U[0, 1]$$

where $U(.)$ is the uniform distribution and $k = \{u, p, s\}$

The augmented joint posterior distribution is obtained as the product of densities in [4]-[6], and all conditional posterior distributions follow easily. Albert and Chib (1993) and Sorensen et al. (1995) give all needed conditional distributions. The fully conditional posterior distributions of $\beta, \mathbf{u}, \mathbf{p}, \mathbf{s}, l, \sigma_u^2, \sigma_s^2$ and σ_p^2 are all in closed form, being normal for $\beta, \mathbf{u}, \mathbf{p}, \mathbf{s}$, truncated normal for each l_i , and scaled inverted chi-square for σ_u^2, σ_s^2 and σ_p^2 . A detailed derivation of these conditional distributions can be found in Heringstad et al. (2001) and Rekaya et al. (2000).

Convergence diagnostics were assessed using the method of Raftery and Lewis (1992) and visual inspection of parameter trace plots. The required length of the burn-in period was less than 6,000 iterations for all parameters. Thus, a total single chain length of 100,000 iterations of the Gibbs sampler was used, with a conservative burn-in period of 25,000 iterations. The remaining 75,000 iterations were retained without thinning for post-Gibbs analysis.

RESULTS AND DISCUSSION

A summary of the posterior distributions of the variance components and heritability of insemination success is presented in Table 3.2. The posterior mean (SD) of the additive variance was 0.034 (0.006). Although small, the posterior mean of the genetic variance indicates the presence of some genetic variability between cows that can be used to improve, or at least to avoid further deterioration of, the reproductive performance of Holstein cows. Weigel and Rekaya (2000), in a study of nonreturn rate at 60 and 90 days after insemination used a threshold model and obtained similar estimates.

The posterior mean (SD) of the service sire variance was 0.009 (0.001). This point estimate tends to be slightly higher than those reported in the majority of fertility studies (Hodel et al., 1995, Taylor et al., 1985). A.-Ranberg et al. (2003) using a multiple trait mixed linear

model for 56-day nonreturn rate in heifers, reported estimates for service sire variance close to zero (lower than 0.0005). Although their estimates are not directly comparable with those obtained in this study, it seems that a higher service sire variance is obtained when repeat services, as well as the order in which they occur, are considered in the model. This is in part because service bulls have different potential of making cows pregnant as indicated by the service bull variance. Further, the order in which bulls were used can lead to different outcome of insemination events and subsequently to the variation in the data. Hansen et al. (1983) concluded that service sire variance was less than one percent for most of the fertility traits considered in their study. These estimates indicate some differences between service sires on their ability to make cows pregnant. Furthermore, given the binary nature of a large number of fertility traits, a small difference in service sire fertility could be the reason for a successful (or failed) insemination event.

The posterior mean of the permanent environmental variance was 0.171 (0.013). To the best of our knowledge, no previous longitudinal study of insemination events has been conducted, so estimates of the permanent environmental variance cannot be compared with literature values. This point estimate is higher than the additive and service sire estimates in this study.

The posterior mean of the heritability was 0.028 (0.005). Although relatively low, this estimate is within the range of reported values for this parameter in fertility studies. The literature estimates ranged from 0.026 to 0.04 for fertility traits measured as number of services per conception (Bar-Anan et al., 1985, Raheja et al., 1989a). Together with the estimate of the genetic variance, it seems that selection for a successful outcome of an insemination event is possible.

The estimate of the regression coefficient on age of the service sire was 0.001, indicating a higher fertility for older (proven) bulls. However, this result has to be interpreted with caution given the preferential use of older bulls. Thus, young bulls tend to be mated with less fertile cows due to lower semen prices. Furthermore, there is the potential of a non-linear relationship between insemination success and age of the service sire that cannot be accommodated by a simple regression.

The posterior mean of the regression coefficient on early milk yield was -0.005. Although expected, (cows with higher milk production have a lower chance of a successful insemination) this estimate was contradictory to the results found by Weigel and Rekaya (2000), who concluded that milk yield had no effect on 60-day nonreturn rate. Dematawewa and Berger (1998) found positive correlations of 0.53 and 0.63 between milk yield and days open and number of services per conception, respectively. This antagonistic relationship between milk yield and fertility traits was not supported by Weller (1989) who reported no correlation between conception status and milk yield.

The estimate of the regression coefficient on days in milk at insemination was 0.003. This positive regression indicates that cows being bred shortly after calving are less likely to become pregnant. Weigel and Rekaya (2000) reported a similar result for the effect of milk yield on 60-day nonreturn rate. In addition, it is possible that the ability of a cow to get pregnant could vary over time. Thus, a random regression model that contemplates a more complex relationship between days in milk and insemination success is more appropriate.

CONCLUSIONS

The use of a longitudinal model allows for the inclusion of all breeding information within an opportunity period. It avoids the elimination of useful information and makes possible

the joint analysis of male and female fertility. Furthermore, it accounts for all service sires, as well as the order in which they were used, leading to potentially more precise estimates. Estimates of genetic parameters indicate the possibility of genetic improvement of reproductive performance through selection. This study is being extended to handle incomplete or “censored” records, including breeding events of non-pregnant cows after 250 days, using a random regression model. Such a model will also allow for better modeling of the effects of age of service sire and days in milk on the insemination outcome. Furthermore, it will allow the derivation of new selection criteria other than the single breeding value, such as time to conception and number of inseminations to conception. Finally, although survival analysis can handle the kind of data used in this study, we believe that a longitudinal model is practically more appropriate, because it allows the use of an animal model and can be incorporated into a joint genetic evaluation with other traits in the selection index in a straightforward manner.

INTERPRETATIVE SUMMARY

The heritability of successful insemination (conceived or not conceived) and variance components were estimated with data collected from AgriTech Analytics from the years 2002 to 2003. Hence heritability of successful insemination was estimated to be 0.028, and 0.034 for additive effect, 0.009 for service sire effect and 0.171 for permanent environment effect. Survival analysis can handle the kind of data used in this study, and a longitudinal model is practically more appropriate, because it allows the use of an animal model and can be incorporated into a joint genetic evaluation with other traits in the selection index in a straightforward manner.

ACKNOWLEDGEMENTS

We thank the National Association of Animal Breeders for the financial support and AgriTech Analytics for providing the data for this research.

REFERENCES

- A.-Ranberg, I. M., B. Heringstad, G. Klemetsdal, M. Svendsen, and T. Steine. 2003. Heifer fertility in Norwegian dairy cattle: Variance components and genetic change. *J. Dairy Sci.* 86:2706-2714.
- Albert, J., and S. Chib. 1993. Bayesian analysis of binary and polychotomous response data. *J. Am. Statist. Assoc.* 88:669-679.
- Bar-Anan, R., M. Ron, and G. R. Wiggans. 1985. Associations among milk yield, yield persistency, conception, and culling of Israeli Holstein dairy cattle. *J. Dairy Sci.* 68:382-386.
- Brotherstone, S., G. Banos, and M. P. Coffey. 2002. Evaluation of yield traits for the development of a UK fertility index for dairy cattle. 7th World Congress on Genetics Applied to Livestock Production. 01-28.
- Clay, J. S., and B. T. McDaniel. 2001. Computing mating bull fertility from DHI nonreturn data. *J. Dairy Sci.* 84:1238-1245.
- Dematawewa, C. M. B., and P. J. Berger. 1998. Genetic and phenotypic parameters for 305-day yield, fertility, and survival in Holsteins. *J. Dairy Sci.* 81:2700-2709.
- Freeze, B. S., and T. J. Richards. 1992. Lactation curve estimation for use in economic optimization models in the dairy industry. *J. Dairy Sci.* 75:2984-2989.
- Gianola, D. 1982. Theory and analysis of threshold characters. *J. Animal Sci.* 54:1079-1096.

- Gianola, D. and R. Rekaya. 2000. Bayesian random regression models for longitudinal binary data with applications to dairy cattle breeding. Symposium on Selected Topics in Variance Components Analysis. University of Florida, Department of Statistics.
- Grosshans, T., Z. Z. Xu, L. J. Burton, and D. L. Johnson. 1996. Genetic parameters for fertility traits in seasonal dairy cattle. Dairy Cattle Breeding in New Zealand.
- Hansen, L. B., A. E. Freeman, and P. J. Berger. 1983. Variances, repeatabilities, and age adjustments of yield and fertility in dairy cattle. J. Dairy Sci. 66:281-292.
- Heringstad, B. R., R. Rekaya, D. Gianola, G. Klemetsdal, and K. A. Weigel. 2001. Bayesian analysis of liability to clinical mastitis in Norwegian cattle with a threshold model: effects of data sampling method and model specification. J. Dairy Sci. 84:2337-2346.
- Hodel, F., J. Moll, and N. Kuenzi. 1995. Analysis of fertility in Swiss Simmental cattle – Genetic and environmental effects on female fertility. Livest. Prod. Sci. 41:95-103.
- Kadarmideen, H. N., R. Thompson, and G. Simm. 2000. Linear threshold model genetic parameters for disease, fertility and milk production in dairy cattle. Anim. Sci. 71:411-419.
- Lopez-Gatius, F., P. Santolaria, J. Yaniz, J. Rutllant, and M. Lopez-Bejar. 2002. Factors affecting pregnancy loss from gestation day 38 to 90 in lactating dairy cows from a single herd. Theriogenology 57:1251-1261.
- Olori, V. E., T. H. E. Meuwissen, and R. F. Veerkamp. 2002. Calving interval and survival breeding values as measure of cow fertility in a pasture-based production system with seasonal calving. J. Dairy Sci. 85:689-696.

- Raftery, A. E., and S. Lewis. 1992. How many iterations in the Gibbs sampler? Pages 736-773 in Bayesian Statistics 5. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds. Oxford University Press, New York, NY.
- Raheja, K. L., E. B. Burnside, and L. R. Schaeffer. 1989a. Heifer fertility and its relationship with cow fertility and production traits in Holstein dairy cattle. *J. Dairy Sci.* 72:2665-2669.
- Raheja, K. L., K. Nadarajah, and E. B. Burnside. 1989b. Relationship of bull fertility with daughter fertility and production traits in Holstein dairy cattle. *J. Dairy Sci.* 72:2679-2682.
- Rekaya, R., K. A. Weigel, D. Gianola, B. Heringstad, and G. Klemetsdal. 2000. Methods for attenuating bias of variance component estimates in threshold models when herds are small. *J. Dairy Sci.* 83(Suppl. 1):56-57 (Abstr.).
- Sorensen, D. A., S. Andersen, D. Gianola, and I. Korsgaard. 1995. Bayesian inference in threshold using Gibbs sampling. *Genet. Sel. Evol.* 27:229-249.
- Taylor, J. F., R. W. Everett, and B. Bean. 1985. Systematic environmental, direct, and service sire effects on conception rate in artificially inseminated Holstein cows. *J. Dairy Sci.* 68:3004-3022.
- Weigel, K. A. 2000. Toward national fertility evaluations. Proc 18th Technical Conference on Artificial Insemination and Reproduction.
- Weigel, K. A., and R. Rekaya. 2000. Genetic parameters for reproductive traits of Holstein cattle in California and Minnesota. *J. Dairy Sci.* 83:1072-1080.
- Weller, J. I. 1989. Genetic analysis of fertility traits in Israeli dairy cattle. *J. Dairy Sci.* 72:2644-2650.

Table 3.1: A descriptive summary of the edited data.

	Number			
Cows	151,758	Inseminations		
Records	297,823	1 st	2 nd	3+*
Successful	128,889	73,207	34,966	20,716
Failed	168,934	78,551	43,585	46,798
			Mean	SD
Service sires	3,190		93.36	229.75
Technicians	2,179		136.68	279.63

* 3rd and subsequent inseminations

Table 3.2. Summary of posterior distributions of the variance components and heritability for insemination success.

Parameter	Mean	SD	HPD (95%)	
σ_a^2	0.034	0.006	0.023	0.045
σ_s^2	0.009	0.001	0.007	0.011
σ_p^2	0.171	0.013	0.145	0.198
h^2	0.028	0.005	0.019	0.037

σ_a^2 = additive genetic variance; σ_s^2 = service sire variance; σ_p^2 = permanent environmental variance; $h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_s^2 + \sigma_p^2 + 1)$; HPD (95%) = Lower and upper bounds of highest probability density interval.

CHAPTER 4

RANDOM REGRESSION MODELS FOR MALE AND FEMALE FERTILITY EVALUATION USING LONGITUDINAL BINARY DATA²

Averill, T., and R. Rekaya

² Averill, T., and R. Rekaya. To be submitted to *Livestock Production Science*.

ABSTRACT

Longitudinal Bayesian threshold analysis of insemination success was carried out using two random regression models with three (M1) and five (M2) parameters to model the additive genetic values at the liability scale. All insemination events of first parity Holstein cows were used. The outcome of an insemination event was treated as a binary response of either a success (1) or a failure (0). Thus, all breeding information for a cow, including all service sires, was included, thereby allowing for a joint evaluation of male and female fertility. An edited data set of 369,353 insemination records from 210,373 first lactation cows was used. On the liability scale, both models included the systematic effects of herd-year of insemination, technician, month of insemination, and regressions on age of service sire, and on milk yield during the first 100 days of lactation. The random effects in the model were the three or five random regression coefficients specific to each cow, the permanent effect of the cow and the service sire effect. Using M1, the estimated heritability of insemination success decreased from 0.035 at day 50 to 0.032 at day 140 and then increased continuously with days in milk. The genetic correlations for insemination success at different time points ranged from 0.83 to 0.99. Their magnitude decreased with the increase of the interval between inseminations. A similar trend was observed for heritability and genetic correlations using M2. However, the average estimate of heritability was much higher (0.058) than those obtained using M1 or a repeatability model. In addition, the genetic correlations followed the same trend as M1, but were lower and with a higher rate of decrease when the interval between inseminations increased. Posterior mean of service sire variance was 0.01 for both models and permanent effect variance was 0.05 and 0.02 for models 1 and 2, respectively. Model comparison based on Bayes factor indicated that M1 was more plausible given the data.

Key words: Longitudinal binary data, random regression, fertility, dairy cattle

INTRODUCTION

In the last three decades, the dairy industry in the United States has experienced substantial changes. Royal et al. (2000) reported that in twenty years (1975-1997) pregnancy rates in Holstein cows decreased by 0.45% per year (Butler and Smith, 1989; Beam and Butler, 1999). The impaired fertility can be partially attributed to unfavorable genetic relationships between fertility and production traits. Although several traits are being used for genetic evaluation of reproductive performance, the outcome of an insemination event is becoming a trait of choice as it is measured early in the breeding season and hence has less environmental influences. Similar to other reproductive traits, success/failure of an insemination depends on both male and female fertility.

The current models and techniques used in fertility studies have concentrated on evaluating female and male fertility separately. Further, not all insemination events during a breeding season are being considered. In fact, traits such as success for first insemination, non-return rate (NR), and conception rate (CR) use only one record per cow. Thus, breeding information for cows having more than one insemination are not being considered leading to a loss of information. More importantly, with one record per cow there is no easy way to account properly for male fertility, as only one service bull will be considered. Consequently, the resulting service sire effects and female fertility could be biased. To account properly for service sire effect and female fertility in a joint analysis, all insemination events must be considered within the framework of a longitudinal threshold model.

In recent years, there has been an increased interest in using test-day models. Schaeffer (2004) reported that random regression models are being used to analyze conformation traits,

body condition scores, feed intake, heart girth measures in dairy cattle, weights and back fat thickness in swine and beef cattle, fork length and weights in rainbow trout, litter size in swine and production traits among others. In dairy cattle, random regression models were primarily used to analyze production traits (Marti and Funk 1994; Veerkamp and Goddard 1998; Swalve 2000; Jensen 2001; Heringstad et al. 2003) and health related traits such as mastitis (Kadarmideen et al. 2000, Heringstad et al. 2001, and Rekaya et al. 2003). Several random coefficient models including random regression, Legendre polynomials, and splines have been implemented and compared (White et al. 1999; Pool et al. 2000). The basic idea underlying all these models consists in modeling the additive genetic values (and/or other random effects in the model) as a function of an observed dependent variable (i.e. time, weight) through a set of random coefficients. Theoretical and biological advantages of random regression based models for longitudinal data compared to cross sectional models are numerous and have been repeatedly and extensively reported in the literature (Meyer 1998, Huisman et al. 2002).

Rekaya et al. (1998) were the first in proposing a longitudinal threshold model for analysis of a sequence of binary responses. Additionally to the advantages of the random regression model using continuous data, longitudinal threshold models offers the possibility of computing quantities of interest to animal breeders that could not be obtained using cross sectional analyses, such as the probability of observing a success/ failure event within period of time. Heringstad et al. (2001) applied the proposed model for the analysis of mastitis incidence in the Norwegian red and white dairy population. The estimated heritability ranged from 0.01 to 0.18 with the maximum being by the middle of the lactation. Furthermore, quantities including the expected number of mastitis episodes per lactation, the number of days without mastitis, and the probability of having at least one mastitis episode during a given period of time were

developed and used as an alternative to the single breeding value for sire selection. Kadarmideen et al. (2001) used a similar approach to analyze mastitis data in the UK Holstein population. Jakobsen et al. (2003) used a bivariate longitudinal threshold-continuous model for analysis of health and production traits in the Danish dairy population. Their model yielded interesting results such as the susceptibility to diseases and the shape of the lactation curve, the relationship between the level of production at the peak of lactation and its timing and the incidence of mastitis.

In this study, the repeatability threshold model proposed by Averill et al. (2004) for analysis of insemination outcome records in first lactation is extended via random coefficients based models with the specific objectives to 1) use all available breeding information for the joint evaluation of male and female fertility 2) implement and compare two random coefficient functions on time for modeling the additive breeding value using Bayesian information criteria, and 3) develop new selection criteria other than the single breeding value.

MATERIALS AND METHODS

The data consisted of insemination records from first parity dairy cows generated between 2002 and 2003 and provided by AgriTech Analytics. The trait of interest was the outcome of all insemination events 1's (success) and 0's (failure). Data editing consisted of keeping herd-year contemporary groups and technicians with at least five and ten records, respectively. All extreme case problem (ECP) classes for contemporary groups and technicians were removed. Further, cows with inconsistent identification, more than one successful insemination per lactation, or an unrealistic interval between consecutive inseminations were removed. After editing, the data consisted of a total of 210,373 cows with 369,353 records,

averaging 1.76 inseminations per cow. The data also included a total of 1582 technicians, 3210 service sires, and 967 herd-year contemporary groups.

The pedigree had a high proportion of missing or unknown sires, due to a high percentage of cows in California with missing sire identification. After matching the identification numbers in the data and pedigree files provided by AgriTech Analytics with the sire pedigrees from the National Association Animal Breeders, the international sire identification numbers were determined and used in building the pedigree. The pedigree file consisted of 218,706 animals.

Before describing the longitudinal setting, a basic latent variable model for a cross-sectional binary response is described. Assume the observed binary response y_i related to a continuous underlying variable l_i satisfying:

$$y_i = \begin{cases} 1 & \text{if } l_i > T \\ 0 & \text{if } l_i \leq T \end{cases} \quad (1)$$

where $l_i \sim N(\mu_i, \sigma_e^2)$ and T is a threshold value. The probability of observing a successful insemination (success) is:

$$\begin{aligned} P_i &= \text{pr}(l_i > T \mid \mu_i) = 1 - \text{pr}(l_i \leq T \mid \mu_i) \\ &= 1 - \Phi\left(\frac{T - \mu_i}{\sigma_e}\right) \end{aligned} \quad (2)$$

where Φ is the cumulative standard normal distribution function. As such, the model is not identifiable and μ_i , T and σ_e^2 could not be inferred separately. Thus, at least two restrictions are need. In this study, T and σ_e^2 were set to zero and one, respectively leading to:

$$P_i = \text{pr}(l_i > T \mid \mu_i) = 1 - \Phi(\mu_i) = \Phi(\mu_i) \quad (3)$$

Furthermore, μ_i can be linearly related to a set of systematic and random effects as:

$$\mu_i = \mathbf{x}_i \boldsymbol{\beta} + \mathbf{z}_i \mathbf{u}$$

where $\mathbf{x}_i$ and $\mathbf{z}_i$ are known incidence row vectors, and $\boldsymbol{\beta}$ and $\mathbf{u}$ are unknown location parameters corresponding to systematic and random effects, respectively.

The implementation of this model in a Bayesian analysis using data augmentation became feasible after Albert and Chib (1993) and Sorensen et al. (1995). All pertinent posterior distributions needed for Bayesian implementation via Markov Chain Monte Carlo methods are in closed form.

In a longitudinal situation, let $\mathbf{y}_i = (y_{it_1}, y_{it_2}, \dots, y_{it_{n_i}})'$ be a $n_i \times 1$ vector of binary responses for animal i ($i = 1, 2, \dots, q$) observed at times $t_1, t_2, \dots, t_{n_i}$. As in the case of cross-sectional analysis, the binary response observed at a time t_j related to a normally distributed underlying variable satisfying (1):

$$l_{ij} \sim N(\mu_{ij}, 1)$$

where μ_{ij} is now some function of time. In this study, two functional forms were used to model μ_{ij} .

Quadratic linear function (M1): A quadratic regression on the time elapsed between calving and insemination date was used to model the additive breeding value of the inseminated cow. Thus,

$$\mu_{ijkm} = \text{fixed} + a_{0i} + a_{1i} Z_{ij} + a_{2i} Z_{ij}^2 + s_k + p_i$$

where

μ_{ijkm} = conditional mean for cow i at time j ;

fixed = includes herd-year, technician, month of insemination, regression on age of service sire, and regression on early milk yield (first 100 days of lactation):

$\mathbf{u}_i = [a_{0i}, a_{1i}, a_{2i}]$ is a 3x1 vector to random regressions specific to cow i ;

s_k = effect of service sire k ;

p_i = environmental effect peculiar to all n_i records of cow i ;

$$z_{ij} = \frac{\text{days in milk } j \text{ for cow } i}{365}$$

Ali-Schaeffer function (M2): This is a five parameters function and was used to model the additive genetic values. Thus,

$$\mu_{ijkm} = \text{fixed} + a_{0i} + a_{1i}z_{ij} + a_{2i}z_{ij}^2 + a_{3i} \ln(z_{ij}^{-1}) + a_{4i} [\ln(z_{ij}^{-1})]^2 + s_k + p_i$$

where μ_{ijkm} , fixed, s_k , and p_i are as before and $\mathbf{u}_i = [a_{0i}, a_{1i}, a_{2i}, a_{3i}, a_{4i}]$ is a 5x1 vector to random regression specific to cow i .

To complete the Bayesian implementation, prior distributions for all unknown parameters in the model have to be specified. It was assumed that:

$$\boldsymbol{\beta} \sim U[\beta_{\min}, \beta_{\max}];$$

$$\mathbf{u}|\mathbf{G}_0 \sim N(\mathbf{0}, \mathbf{A} \otimes \mathbf{G}_0);$$

where $\mathbf{G}_0$ is a 3x3 and 5x5 genetic (co)variances matrix for M1 and M2, respectively.

$$\mathbf{p}|\sigma_p^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_p^2)$$

$$\mathbf{s} \sim N(\mathbf{0}, \mathbf{I}\sigma_s^2)$$

$$\sigma_s^2 \sim U[0,1]$$

$$\sigma_p^2 \sim U[0,1]$$

where $U(\cdot)$ is the uniform distribution, $\mathbf{A}$ is the additive relationship matrix and $\beta_{\min}$ and $\beta_{\max}$ were set to -100 and 100 , respectively.

The vector of unknown parameters was augmented with the liabilities as suggested by Sorensen et al. (1995). The resulting joint posterior distribution and the conditional distributions needed for the implementation of Gibbs sampler were in closed form, being normal for the systematic and random effects, truncated normal for the liabilities and scaled inverted Whishart and scaled inverted chi square for $\mathbf{G}_0$, and σ_s^2 and σ_p^2 , respectively. For both models, a unique chain of 75,000 iterations was implemented with a burn-in period of 25,000 iterations.

The Bayes factor, as defined by Newton and Raftery (1994), was used to assess the plausibility of postulated models. The marginal density of the data under each of the models was estimated from the harmonic means of likelihood values evaluated at the posterior draws:

$$\hat{p}(\mathbf{y} \mid M_i) = \left\{ \frac{1}{N} \sum_{j=1}^N [p(\mathbf{y} \mid \boldsymbol{\theta}^{(j)}, M_i)]^{-1} \right\}^{-1}$$

where $\mathbf{y}$ is the vector of observed binary responses and $\boldsymbol{\theta}^{(j)}$ is the Gibbs sampling sample of j of parameters under model M_i . The estimated Bayes factor between models M_i and M_j is:

$$B_{M_i, M_j} = \frac{\hat{p}(\mathbf{y} \mid M_i)}{\hat{p}(\mathbf{y} \mid M_j)}$$

Genetic (co)variances and heritabilities

The genetic (co)variances and heritabilities of success/failure of insemination at the liability scale are a function of time and are easily obtained. For model M1, the genetic covariance between success/failure of inseminations at times, t_i and t_j is given by:

$$\text{cov}(t_i, t_j) = \mathbf{V}'(t_i) \mathbf{G}_0 \mathbf{V}(t_j)$$

where $\mathbf{G}_0$ is a 3x3 genetic (co)variances for the random regression parameters and

$$\mathbf{V}'(t_i) = [1 \ z_{ti} \ z_{ti}^2]$$

Similarly, for M2 the genetic covariance is given by:

$$\text{cov}(t_i, t_j) = \mathbf{W}'(t_i) \mathbf{G}_0 \mathbf{W}(t_j)$$

where $\mathbf{G}_0$ is a 5x5 genetic (co)variances matrix and

$$\mathbf{W}'(t_i) = [1 \ z_{ti} \ z_{ti}^2 \ \ln(z_{ti}^{-1}) \ \ln(z_{ti}^{-1})^2]$$

Heritabilities at the liability scale at time t are defined as:

$$h_t^2 = \frac{\mathbf{V}'(t) \mathbf{G}_0 \mathbf{V}(t)}{\mathbf{V}'(t) \mathbf{G}_0 \mathbf{V}(t) + \sigma_p^2 + \sigma_s^2 + 1} \quad (\text{M1})$$

$$h_t^2 = \frac{\mathbf{W}'(t) \mathbf{G}_0 \mathbf{W}(t)}{\mathbf{W}'(t) \mathbf{G}_0 \mathbf{W}(t) + \sigma_p^2 + \sigma_s^2 + 1} \quad (\text{M2})$$

Bayesian analysis via Markov Chain Monte Carlo techniques of longitudinal binary data allows, in a straightforward manner, the calculation of quantities of interest for selection decisions. In fact, criteria such as the probability of conception after first insemination, probability of conception within 365 days, probability of no conception within 365 days, and probability of no conception in the first three inseminations are computed as a by product of the

sampling process and can be used together with the estimated breeding values for making breeding decisions. The following selection criteria were computed in this study:

- a) Probability of conception after first insemination ($t_j=75$ days)

$$p(y_{i75} = 1) = 1 - p(y_{i75} = 0) = \Phi(\mu_{i75})$$

- b) Probability of conception within 365 days

$$p_i(1) = 1 - \prod_{j=1}^{n_i} [1 - \Phi(\mu_{ij})]$$

- c) Probability of no conception within 365 days

$$p_i(0) = 1 - p_i(1)$$

- d) Probability of no conception in the first three inseminations ($t_j=75$, $t_j=96$, and $t_j=117$)

$$p(y_{i75} = 0, y_{i96} = 0, y_{i117} = 0) = p(y_{i75} = 0)p(y_{i96} = 0)p(y_{i117} = 0)$$

RESULTS AND DISCUSSION

Table 4.1 presents the summary of the posterior distributions of the genetic (co)variances between the three parameters of model M1, and permanent and service sire variances. The intercept (first parameter of the quadratic function) explains over 87% of the total genetic variance for insemination success at the liability scale when $t = 110$ days. The third parameter explains only 1% of the total variance (at $t = 110$ days) indicating its limited effect on the ability of predicting a cow becoming pregnant after an insemination. The genetic correlations between the three parameters of model M1 (Table 4.2) indicates moderate to low negative correlations between the second parameter and the other two. However, the genetic correlation between first and third parameters was positive and high. These correlations indicate that for low/moderate values of t (days between calving and insemination), the second parameter of the function plays a

role in decreasing the genetic variance for success/failure at the liability scale as a result of its negative correlation with the intercept and the limited effect of the third parameter because the covariable, $(\frac{t}{365})^2$, is very small for low/moderate t. For large values to t, the contribution of the third parameter increases and offsets the decrease induced by the negative correlation between the first and second parameters. This trend is clearly evident in Table 4.3 (diagonal). The genetic variance decreases continuously from day 70 to approximately day 150 after calving and then increases with time. The change over time of the heritability of insemination success (Figure 4.1) shows the same trend observed for the genetic variance. It decreases at the beginning of the lactation to a minimum (0.032) between day 155 and 176 and then increases continuously to the end of lactation. However, the range of variation of heritability over time is small 0.032 to 0.041 with an average of 0.034. The latest, although higher than the estimate obtained using the repeatability model, is still within the range of estimates reported in the literature (Bar-Anan et al. 1985; Raheja et al. 1989). The genetic correlations for insemination success at different time points are positive and high and ranged from 0.787 to 0.998. The magnitude of these correlations decreases with the increase of the interval between inseminations. In fact, the correlation was 0.998 between t = 70 and t = 100 and dropped to 0.787 between t = 70 and t = 365. This result suggests that if a cow fails to get pregnant in a given insemination, her chances of becoming pregnant increases with the increase of the interval between the failed insemination and consecutive inseminations. Furthermore, these correlations have to be interpreted with caution given that cows getting pregnant in a given insemination have all their “hypothetical” future inseminations assumed by the model as if they were successful. In other words, for these pregnant cows, the genetic correlation between “hypothetical” future inseminations is equal to

one. Consequently, the true genetic correlation between an unsuccessful insemination and future insemination is smaller than the values presented in (Table 4.3).

The point estimate of the service sire variance (0.01) is slightly higher than the one obtained using the repeatability model (0.009) but in the range of reported estimates for male fertility (Hansen et al. 1983; Taylor et al. 1985; A-Ranberg et al 2003). The permanent effect variance was substantially smaller (0.05) than the estimate obtained by Averill et al. (2004) using the repeatability model (0.17). Furthermore, the new estimate is more realistic indicating perhaps a better modeling using a random regression model that accounts for the variation in the genetic variance over time.

Table 4.4 presents the summary of the posterior distributions of the genetic (co)variances between the five parameters of M2, and permanent and service sire variances. As with M1, the intercept explains a large portion of the total variance. The genetic correlations (Table 4.5) for the fourth and fifth parameters with the other three parameters were negative except for the correlation between the fourth and second parameters (0.11). All the genetic correlations between the first three parameters and between the fourth and fifth parameters were positive and ranged from 0.12 (first with second) to 0.60 (fourth and fifth). Although a detailed dissection of the effect of each parameter on the genetic variance of insemination success is more complex and less evident than with model M1 because the larger number of parameters in M2, in general it is seems that the fourth and fifth parameters have a more pronounce effect on the genetic variance at the beginning of the lactation as a result of the their negative correlation with the intercept and the magnitude of the covariables $\log(\frac{365}{t})$ and $\log(\frac{365}{t})^2$ for small values of t. As soon as t reaches a value around 90 days, the contribution of the second and third parameters

(positively correlated with the intercept) exceeds the decrease induced by the fourth and fifth parameters and the genetic variance of insemination success increases sharply. Change of heritability over time is presented in Figure 4.2. Similarly to the genetic variance, the heritability decreases to a minimum around day 90 and then increases continuously to reach estimates greater than 0.08 by day 300. The average estimate of heritability is much higher than those obtained using M1 (0.034) and the repeatability model (0.028) using the same data set with different editing criteria (Averill et al. 2004).

The genetic correlations for insemination success at different time points are positive but of smaller magnitude than those obtained using model M1 (Table 4.6). As with model M1, the magnitudes of these correlations decrease with the increase of the interval between inseminations. For two inseminations realized at an interval of less than 100 days, the genetic correlation was always greater than 0.90, except for between day 70 and 150 (Table 4.5). However, the decrease in the magnitude of the genetic correlation with the increase of the interval between inseminations is more pronounced than with M1 (Table 4.3). In fact, the correlation was 0.90 between $t = 70$ and $t = 100$ and dropped to 0.44 between $t = 70$ and $t = 365$.

The point estimate of the service sire variance was similar to the one obtained using M1 (0.01) and a repeatability model (0.009). However, the permanent effect variance was smaller (0.02) than the estimate obtained using model M1 (0.05) and a repeatability model (0.17).

Based on Bayes factor, model M1 was more plausible than M2 given data. This result suggests that M2 has too many parameters to fit a data with limited number of records per cows. In fact, the average number of records per cow was 1.76 with 37% of cows have only one record. Furthermore, this result helps explain the non-expected estimates of heritabilities, genetic correlation between insemination events and permanent effect variance. We stress that holding

the permanent effect variance constant over time could be part of the problem as it has been the case in analyzing milk yield data using test-day models. However, increasing the number of parameters by including random regression coefficients to model the permanent effect could be counter productive given the limited number of records per cow in the data used in this study.

The ranking of animals based on the three new selection criteria provides an extra tool for selection. Table 4.7 presents the top and bottom five cows based on the defined three probabilities using model 1. The five best cows have a very high probability of conception after first insemination and consequently low probabilities for not being pregnant after 3 inseminations or after 365 days. In fact, all five cows got pregnant on their first insemination. The worst five cows have an almost null probability of conception after first insemination and very high p_2 and p_3 . Based on their phenotypic data, all five cows were empty after more than 300 days since calving.

CONCLUSION

The use of longitudinal models allows for the inclusion of all breeding information over a period of time. It avoids the elimination of useful information and makes possible the joint analysis of male and female fertility. Longitudinal models account for all service sires, as well as the order in which they were used, leading to potentially more precise estimates. In this study, two random regression models with different parameter functions, a quadratic linear model (model 1) and Ali-Schaeffer model (model 2) were implemented and compared. The two model estimates of genetic parameters indicate the possibility of genetic improvement of reproductive performances through selection. Furthermore, random regression models account for the variation in the genetic variance for insemination success over time leading potentially to more realistic modeling compared with repeatability model. Estimates of genetic parameters were

quite different between the two random regression models. The estimates of heritability using model 2 were higher than those expected for fertility traits, especially by the end of the lactation. Comparison between the two models based on Bayes factor indicated that model 1 was more plausible given the data

REFERENCES

- A.-Ranberg, I. M., B. Heringstad, G. Klemetsdal, M. Svendsen, and T. Steine. 2003. Heifer fertility in Norwegian dairy cattle: Variance components and genetic change. *J. Dairy Sci.* 86:2706-2714.
- Albert, J. H and S. Chib. 1993. Bayesian analysis of binary and polychotomous response data. *J. American Statist. Association.* 88:669-679.
- Averill, T., R. Rekaya, and K. Weigel. 2004. Genetic analysis of male and female fertility using longitudinal binary data. *J. Dairy Sci.* (accepted on minor revision).
- Bar-Anan, R., M. Ron, and G. R. Wiggans. 1985. Associations among milk yield, yield persistency, conception, and culling of Israeli Holstein dairy cattle. *J. Dairy Sci.* 68:382-386.
- Beam, S. W. and W. R. Butler. 1999. Energy balance effects on follicular development and first ovulation in post-partum cows. *J. Reprod. Fertil.* 54:411-424.
- Hansen, L. B., A. E. Freeman, and P. J. Berger. 1983. Variances, repeatabilities, and age adjustments of yield and fertility in dairy cattle. *J. Dairy Sci.* 66:281-292.
- Heringstad, B., R. Rekaya, D. Gianola, G. Klemetsdal, and K. A. Weigel. 2001. Bayesian analysis of liability to clinical mastitis in Norwegian cattle with a threshold model: effects of data sampling method and model specification. *J. Dairy Sci.* 84:2337-2346.

- Heringstad, B., R. Rekaya, D. Gianola, G. Klemetsdal, and K. A. Weigel. 2003. Genetic change for clinical mastitis in Norwegian cattle: a threshold model analysis. *J. Dairy Sci.* 86:369-375.
- Huisman, A. E., E. Kanis, J. A. M. van Arendonk. 2002. Application of random regression models in pig breeding to select on growth and feed intake patterns. 53rd Annual Meeting of the European Association for Animal Production. p.1-8.
- Jakobsen, J. H., R. Rekaya, J. Jensen, D. A. Sorensen, P. Madsen, D. Gianola, L. G. Christensen, J. Pederson. 20003. Bayesian estimates of covariance components between lactation curve parameters and disease liability in Danish Holstein cows. *J. Dairy Sci.* 86:3000-3007.
- Jensen, J. 2001. Genetic evaluation of dairy cattle using test-day models. *J. Dairy Sci.* 84:2803-2812.
- Kadarmideen, H. N., R. Thompson, and G. Simm. 2000. Linear and threshold model genetic parameters for disease, fertility and milk production in dairy cattle. *Anim. Sci.* 71:411-419.
- Kadarmideen, H., R. Rekaya, and D. Gianola. 2001. Bayesian approach to cross-sectional genetic analysis of mastitis in dairy cattle. *J. Anim. Sci.* 73:229-240.
- Meyer, K. 1998. Estimating covariance functions for longitudinal data using a random regression model. *Genet. Sel. Evol.* 30:221-240.
- Newton, M. A. and A. E. Raftery. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. *J. R. Statist. Soc. B.* 56:3-48.

- Pool, M. H., L. L. G. Janss, and T. H. E. Meusissen. 2000. Genetic parameters of Legendre polynomials for first parity lactation curves. *J. Dairy Sci.* 83:2640-2649.
- Raheja, K. L., E. B. Burnside, and L. R. Schaeffer. 1989. Heifer fertility and its relationship with cow fertility and production traits in Holstein dairy cattle. *J. Dairy Sci.* 72:2665-2669.
- Rekaya, R., D. Gianola, K. Weigel, and G. Shook. 2003. Longitudinal random effects models for genetic analysis of binary data with application to mastitis in dairy cattle. *Genet. Sel. Evol.* 35:1-12.
- Rekaya, R., S. L. Rodriguez-Zas, D. Gianola, and G. E. Shook. 1998. Test-day models for longitudinal binary responses: an application to mastitis in Holsteins. Page 44 in book of abstracts 49th Annual meeting of EAAP.
- Schaeffer, L. R. 2004. Application of random regression models in animal breeding. *Livest. Prod. Sci.* 86:35-45.
- Sorensen, D. A., S. Anderson, D. Gianola, and I. Korsgaard. 1995. Bayesian inference in threshold models using Gibbs sampling. *Genet. Sel. Evol.* 27:229-249.
- Swalve, H. H. 2000. Theoretical basis and computational methods for different test-day genetic evaluation methods. *J. Dairy Sci.* 83:1115-1124.
- Taylor, J. F., R. W. Everett, and B. Bean. 1985. Systematic environmental, direct, and service sire effects on conception rate in artificially inseminated Holstein cows. *J. Dairy Sci.* 68:3004-3022.
- Veerkamp, R. F., and M. E. Goddard. 1998. Covariance functions across herd production levels for test day records on milk, fat, and protein yields. *J. Dairy Sci.* 81:1690-1701.

White, I. M. S., R. Thompson, and S. Brotherstone. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *J. Dairy Sci.* 82:632-638.

Table 4.1: Summary of posterior distributions of the genetic (co)variances between the three random coefficients, and permanent and service sire variances: *Model 1*

Component ¹	Mean	SD	HPD (95%)	
g_{00}	0.0422	0.0049	0.0329	0.0506
g_{11}	0.0294	0.0065	0.0187	0.0455
g_{22}	0.0010	0.0002	0.0008	0.0013
g_{01}	-0.0169	0.0009	-0.0031	-0.0004
g_{02}	0.0056	0.0006	0.0044	0.0067
g_{12}	-0.0013	0.0001	-0.0003	8.5E-6
σ^2_p	0.0519	0.0294	0.0122	0.1009
σ^2_s	0.0101	0.0012	0.0078	0.0125

¹ g_{ij} is the genetic (co)variance between random coefficients i and j

Table 4.2: Posterior means of genetic correlations between random coefficients (above diagonal) and associated standard errors (below diagonal): *Model 1*

Random coefficient	a_0	a_1	a_2
a_0		-0.4592	0.8479
a_1	0.2097		-0.2357
a_2	0.0671	0.1722	

Table 4.3: Posterior means of genetic variances (SD) of insemination success (diagonal), genetic correlations between specific time points (above diagonal), and associated standard deviations (below diagonal): *Model 1*

Time points	70	100	150	250	300	365
70	0.0371 (0.0039)	0.9977	0.9824	0.9097	0.8582	0.7867
100	0.0005	0.0359 (0.0042)	0.9929	0.9358	0.8909	0.8264
150	0.0048	0.0022	0.0350 (0.0054)	0.9710	0.9383	0.8868
250	0.0317	0.0246	0.0126	0.0374 (0.0086)	0.9937	0.9713
300	0.0530	0.0441	0.0279	0.0032	0.0408 (0.0104)	0.9918
365	0.0828	0.0724	0.0521	0.0148	0.0043	0.0474 (0.0128)

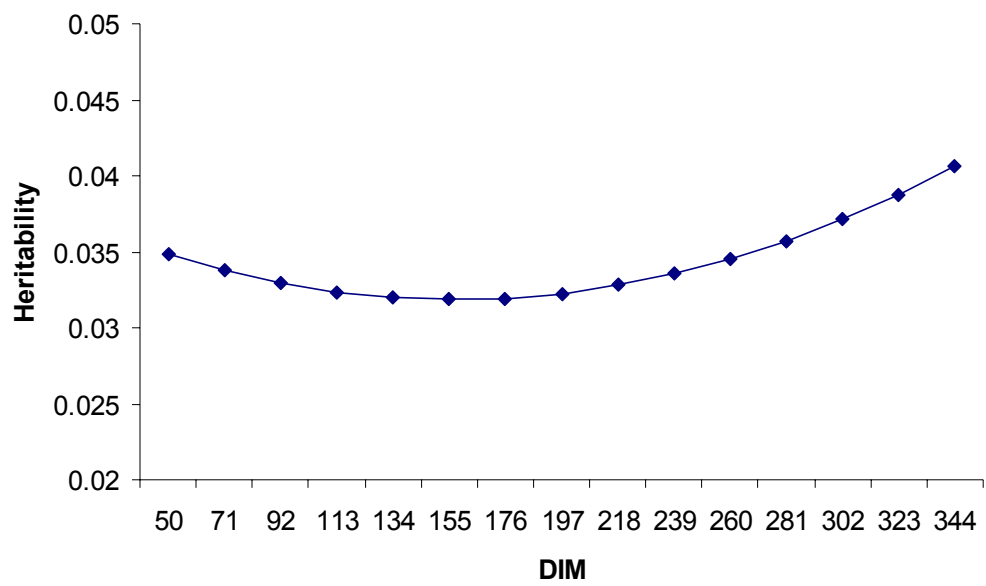


Figure 4.1: Heritability of insemination success as a function of time using model 1.

Table 4.4: Summary of posterior distributions of the genetic (co)variances between the five random coefficients, and permanent and service sire variances:
Model 2

Component ¹	Mean	SD	HPD (95%)	
g_{00}	0.0970	0.0175	0.0627	0.1174
g_{11}	0.0011	0.0002	0.0007	0.0013
g_{22}	0.0007	4.5E-5	0.0006	0.0008
g_{33}	0.0207	0.0061	0.0097	0.0283
g_{44}	0.0008	9.2E-5	0.0007	0.0010
g_{01}	0.0009	0.0037	-0.0048	0.0048
g_{02}	0.0041	0.0012	0.0023	0.0055
g_{03}	-0.0342	0.0162	-0.0507	-0.0017
g_{04}	-0.0070	0.0007	-0.0081	-0.0055
g_{12}	0.0003	0.0002	-0.0002	0.0007
g_{13}	0.0007	0.0016	-0.0013	0.0031
g_{14}	-7.1E-5	0.0002	-0.0004	0.0004
g_{23}	-0.0012	0.0008	-0.0023	0.0002
g_{24}	-3.3E-5	6.6E-5	-0.0001	0.0001
g_{34}	0.0025	0.0012	-9.0E-6	0.0038
σ_p^2	0.0212	0.0086	0.0080	0.0376
σ_s^2	0.0101	0.0012	0.0077	0.0125

¹ g_{ij} is the genetic (co)variance between random coefficients i and j

Table 4.5: Posterior means of genetic correlations between random coefficients (above diagonal) and associated standard errors (below diagonal): *Model 2*

Random coefficient	a_0	a_1	a_2	a_3	a_4
a_0		0.1176	0.4993	-0.7073	-0.7947
a_1	0.3662		0.3013	0.1145	-0.0715
a_2	0.1042	0.2841		-0.3047	-0.0474
a_3	0.2646	0.3119	0.1871		0.5969
a_4	0.022	0.2532	0.0909	0.2581	

Table 4.6: Posterior means of genetic variances (SD) of insemination success (diagonal), genetic correlations between specific time points (above diagonal), and associated standard deviations (below diagonal): *Model 2*

Time points	70	100	150	250	300	365
70	0.0321 (0.0128)	0.9028	0.7043	0.5211	0.4758	0.4366
100	0.04662	0.0343 (0.0122)	0.9397	0.8318	0.8001	0.7707
150	0.1239	0.0224	0.0481 (0.0079)	0.9704	0.9549	0.9382
250	0.1612	0.0440	0.0045	0.0782 (0.0056)	0.9982	0.9933
300	0.1626	0.0454	0.0065	0.0006	0.0921 (0.0097)	0.9984
365	0.1597	0.0440	0.0099	0.0026	0.0008	0.1093 (0.0152)

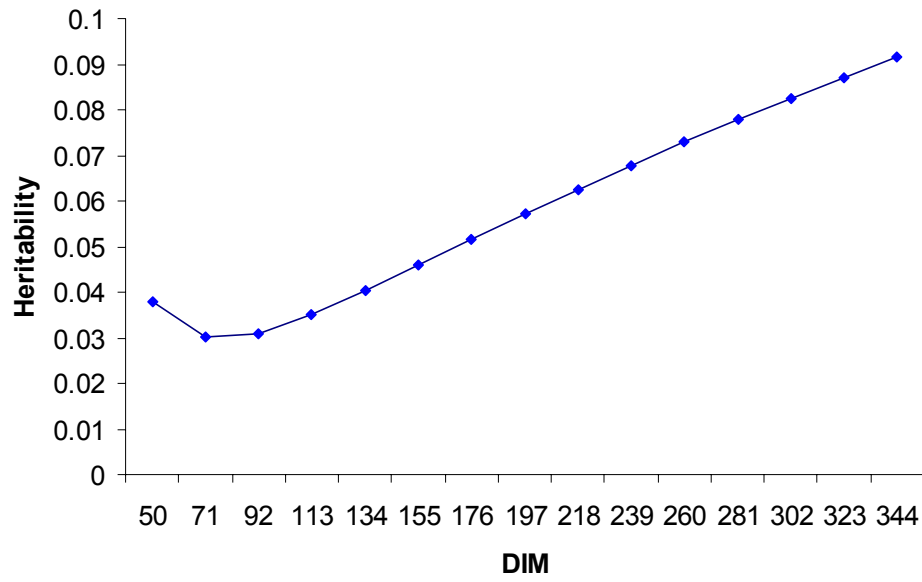


Figure 4.2: Heritability of insemination success as a function of time using model 2.

Table 4.7: Probability of conception after first insemination, probability of no conception after 3 inseminations and probability of no conception after 365 days for the best and worst five cows:
Model 1

Cow	p_1^1	p_2^2	p_3^3	Successful Insemination
Best 5 cows				
1	0.993	0.04	0.01	1
2	0.991	0.04	0.01	1
3	0.991	0.04	0.01	1
4	0.990	0.05	0.01	1
5	0.990	0.05	0.01	1
Worst 5 cows				
1	0.03	0.995	0.992	Open
2	0.03	0.994	0.992	Open
3	0.04	0.994	0.992	Open
4	0.06	0.993	0.991	Open
5	0.06	0.993	0.991	Open

¹ Probability of conception after first insemination ($t_j=75$ days)

² Probability of no conception in the first three inseminations ($t_j=75$, $t_j=96$, and $t_j=117$)

³ Probability of no conception within 365 days

CHAPTER 5

CONCLUSIONS

Long time selection for production responses has caused deterioration of some secondary or non-production traits such as fertility or health conditions as a result of their antagonist relationships. Thus, improvement in or at least the maintenance of herd fertility has become a major objective in genetic improvement programs worldwide. Although the heritability of fertility traits is generally low, ranging from one to ten percent depending on definition of the trait and the methodology used for analysis, there is a consensus that sufficient genetic variability exists and can be exploited to improve reproductive performance. Further, current models and techniques used in fertility studies have concentrated on evaluating female and male fertility separately.

Longitudinal data analysis offers an opportunity for joint genetic evaluation of male and female fertility. It accounts for all services sires used to inseminate a cow during a breeding season as well as the order in which they were used. Furthermore, it avoids eliminating useful information, leading potentially to more precise estimates.

Outcome of insemination events scored as binary traits were analyzed longitudinally using a repeatability model and two random regression models. In both studies, the estimated genetic parameters indicate the existence of sufficient genetic variability, which can be used for genetic improvement of male and female fertility. Based on the results of both studies, it seems that a random regression model with limited number of random coefficients is more appropriate to analyze longitudinal fertility data given the small number of records available per cow.

Allowing for the permanent effects to change over time could lead potentially to a more realistic model. However, it is not clear that it is going to lead to more accurate results with fertility data given the large number of cows with only one record.