Potential accuracy of genetic evaluation for calving difficulty with incomplete data on calving difficulty and/or birth weight using a bivariate threshold-linear animal model

Rodolfo RAMIREZ-VALVERDE, Ignacy MISZTAL, Keith BERTRAND

Animal and Dairy Science Department, University of Georgia, Athens, GA 30602, USA

Abstract. The purpose of this study was to evaluate the potential loss of accuracy in direct and maternal predicted breeding values (PBV) for calving difficulty (CD) with different levels of missing records of CD and/or birth weight (BW), using a bivariate threshold-linear animal model. Data obtained from the American Gelbvieh Association included 84,420 first-parity records with both CD and BW available. The final pedigree file included 178,858 animals. The model included fixed calf-sex–dam-age, random herd-year-season, and animal direct and maternal effects. Different levels of missing observations for CD and BW were obtained by randomly deleting 0, 25, 50, 75, and 100% of records for both traits in various combinations. Correlation estimates between PBV for CD obtained with complete and incomplete data were used to measure the changes in PBV for different levels of missing records. Reported correlations are means of three replicates. The results suggest that the information on direct and maternal PBV provided by CD records is more reliable than the information provided by BW records. The difference was especially large when a high proportion of CD records were missing. Correlations above 0.96 and 0.95 for direct and maternal PBV, respectively, when missing 25% or 0% of the CD or BW records suggest that small changes would be predicted with a low proportion incomplete data. For genetic prediction of popular sires (with > 100 pogeny), a higher proportion of missing records could be tolerated. The results suggest that the bivariate threshold-linear animal model is useful for routine genetic evaluation of CD with incomplete field data.

Key words: threshold-linear model, missing records, calving difficulty, beef cattle.

Introduction

Calving difficulty (CD) is an economically important trait in beef cattle. It is usually recorded in four or five subjective categories according to the assistance given

at birth to indicate the degree of difficulty associated with the birth. GIANOLA, FOULLEY (1983), HARVILLE, MEE (1984), and GILMOUR et al. (1985) developed the threshold model for genetic evaluation of categorical traits in order to manage the nonlinear structure of the data. The threshold methodology is recommended because it takes full account of the categorical nature of the trait, involves only a straightforward genetic extension of the linear mixed model methodology, and increases rates of genetic gain (MEUWISSEN et al. 1995). Even though a similar performance of the threshold and linear models has been reported for analysis of CD in beef cattle (RENAND et al. 1990, VARONA et al. 1999b) when birth weight (BW) is available, a bivariate threshold-linear animal model is superior to univariate or bivariate linear-linear animal models. Subsequently, that model was recommended for genetic evaluation of CD in beef cattle (VARONA et al. 1999b, RAMIREZ-VALVERDE et al. 2001).

In beef cattle field data, it is common to have missing observations on CD and/or BW. FOULLEY et al. (1983) proposed a joint evaluation method for one categorical and two continuous traits, assuming that all traits were recorded on all individuals, while JANSS and FOULLEY (1993) modified the method to handle data with missing patterns for analysis of one categorical and one continuous trait with unequal design matrices. HOESCHELE et al. (1995) extended the method to include a single categorical variable that has multiple categories with multiple continuous variables and any pattern of missing data. A system for routine joint sire genetic evaluation of CD and BW using a bivariate threshold-linear sire-maternal grandsire model accounting for missing data was implemented at Cornell University (WANG et al. 1997).

The contribution of CD and BW data to predicted breeding values (PBV) for CD depends on genetic correlations between these traits, completeness of the data, and the data structure (RAMIREZ-VALVERDE et al. 2001). Under practical conditions, when a large fraction of records of either trait is unavailable, the resulting loss of accuracy of PBV for CD would be dependent on the specific genetic correlations and data structure.

The objectives of this study were to determine the relative contribution of CD and BW records to the prediction of breeding values for CD, and to evaluate the potential loss of accuracy in direct and maternal PBV for CD with different levels of missing records of CD and/or BW, using a bivariate threshold-linear animal model.

Material and methods

Data

General information about the data provided by the American Gelbvieh Association is shown in Table 1. The edited data file included records from 84,820 calves

Table 1. Description of the Gelbvieh data set

	No. of observations	Percentage, %
Final data file	84,820	
Final pedigree file	179,857	
Sires with progeny in data file:		
Total	6,954	
Sires with more than 100 progeny	94	
Calving difficulty:		
1. No assistance	62,243	73.4
2. Minor assistance	15,882	18.7
3. Major assistance	5,348	6.3
4. Caesarean	1,347	1.6
Calving by year:		
1973 to 1980, 1998	< 1,000	5.2
1981 to 1985	1,000-3,000	10.6
1986 to 1990	3,001-6,000	25.4
1991 to 1997	> 6,000	58.8
Calving by season:		
Jan. to Mar.	55,890	66.0
Apr. to Jun.	19,724	23.2
Jul. to Sep.	5,283	6.2
Oct. to Dec.	3,923	4.6

born from first-parity cows (age of the dam from 550 to 930 d) with both BW and CD score recorded. Birth weight records were distributed with a mean of 40.6 kg and a standard deviation of 4.9 kg. Calving difficulty was recorded in four ordered categories according to the required amount of assistance. The scores were 1, 2, 3, and 4; for no assistance, minor assistance, major assistance, and Caesarean, respectively. The typical distribution of CD was that most of the records fell in category 1 (73.4%), and few records in category 4 (1.6%).

Data included information from 2,351 herds that submitted data from 1973 to 1998, but the majority of records came from 1991 to 1997 (58.8%). Seasonal calving was observed, with births concentrated from January to March (66.0%), and from April to June (23.2%). The final pedigree file included 95,037 additional animals.

Model of analysis

Analyses were performed using a bivariate threshold for CD and linear for BW animal model. The model included eight fixed classes of calf sex combined with dam age (< 700 d, 700-724 d, 725-750 d, and > 750 d) and 16,650 random herd-year-season levels. The bivariate model used was:

$$\mathbf{L}_{cd} = \mathbf{X}_{cd}\boldsymbol{\beta}_{cd} + \mathbf{Z}_1\mathbf{h}_{cd} + \mathbf{Z}_2\mathbf{u}_{dcd} + \mathbf{Z}_3\mathbf{u}_{mcd} + \mathbf{e}_{cd}, \text{ and}$$

$$\mathbf{Y}_{bw} = \mathbf{X}_{bw}\boldsymbol{\beta}_{bw} + \mathbf{Z}_1\mathbf{h}_{bw} + \mathbf{Z}_2\mathbf{u}_{dbw} + \mathbf{Z}_3\mathbf{u}_{mbw} + \mathbf{e}_{bw},$$

where $\mathbf{b}$ are the vectors of effects associated with calf sex and dam age; $\mathbf{h}$ are the vectors of effects of random herd-year-season; $\mathbf{u}_d$ are the vectors of genetic direct additive effects; $\mathbf{u}_m$ are the vectors of genetic maternal breeding; $\mathbf{e}$ are the vectors of residuals; $\mathbf{X}$, $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Z}_3$ are incidence matrices that linked data with fixed calf-sex–dam-age effects, random herd-year-season effects, and direct and maternal genetic effects, respectively. The model was assumed as a continuous trait for BW ($\mathbf{Y}_{bw}$), but CD was assumed to have an underlying distribution ($\mathbf{L}_{cd}$), with the response of CD modelled with the following threshold distribution:

$$f(\mathbf{y}|\mathbf{L}) = \prod_{i=1}^{n} f(y_i|L_i) = \prod_{i=1}^{n} I(L_i < \tau_1)I(y_i = 1) + I(\tau_1 < L_i < \tau_2)I(y_i = 2) +$$

$$I(\tau_2 < L_i < \tau_3)I(y_i = 3) + I(\tau_3 < L_i)I(y_i = 4),$$

where τ_1, τ_2, and τ_3 are thresholds that define the four categories of response; $I(.)$ is an indicator function that takes value 1 if the condition specified is true, otherwise the value is 0; CD and BW were assumed to be correlated.

The BLUP90THR program (MISZTAL 1999) was used to obtain the PBV. Variance-covariance components were obtained from a smaller Gelbvieh data set by performing a Bayesian analysis of the bivariate model using a Gibbs sampling algorithm (VARONA et al. 1999a). The posterior means of direct and maternal heritabilities were 0.21 and 0.09, respectively for CD, and 0.26 and 0.05, respectively for BW. Genetic correlation estimates between genetic direct effects, and between genetic maternal effects in both traits were 0.81 and 0.45, respectively. The genetic correlation between direct and maternal effects was –0.33 and –0.34, for CD and BW, respectively. Moreover, genetic correlation between the genetic direct effects for BW and the genetic maternal effects for CD was –0.39, and –0.33 for genetic direct effects on CD and maternal genetic effects for BW. The estimate of the residual correlation between CD and BW was 0.50.

Description and comparison with the incomplete data

Several combinations of incomplete data were included to evaluate different patterns of missing records for CD and BW. Records of CD (0, 25, 50, 75, and 100%) were deleted randomly in all combinations with BW observations also deleted randomly these levels.

Genetic direct and maternal PBV obtained with the complete data set for all animals included in the pedigree file or for sires with more than 100 progeny were correlated with each of the correspondent PBV obtained with the different levels of CD and BW missing records. The correlation estimates between PBV for CD obtained with complete data and data sets with differing percentages of missing observations on CD and BW were used to measure the potential loss of accuracy of PBV for different levels of missing records. Reported correlations were means of three replicates from samples with random missing records.

Results and discussion

The correlation estimates between PBV for CD obtained with complete data and different percentages of missing CD or BW records are presented in Table 2. For direct PBV considering all animals in the pedigree file, the correlation obtained with all CD records only was slightly higher than considering all BW records only (0.88 vs 0.85). This difference was larger when incomplete records were considered of CD and no records of BW, as opposed to no records of CD and incomplete records of BW (0.84 vs 0.67, 0.78 vs 0.59, and 0.67 vs 0.47; for 25, 50, and 75% of missing records in the incomplete trait, respectively). However, when incomplete records were used in two traits or incomplete records in one trait and complete records in the other, a similar correlation was found in the opposite missing trait situation. These results suggest a similar amount of information contained in the discrete and continuous observations, especially when the discrete trait is recorded and when genetic and residual correlations between the two traits are fully accounted for. In univariate analysis of CD or BW, JANSS and FOULLEY (1993) found that the accuracy for the continuous trait was higher than for the discrete trait, indicating that information in a discrete observation is less reliable than information in a continuous observation.

For direct PBV of sires with more than 100 progeny, the correlations obtained with more CD than BW records were in all cases higher than with more BW than CD records (Table 2). For example, the correlation with all CD records and no BW records was 15% higher than the opposite situation with no CD records and all BW records (0.98 vs 0.85). This suggests that the relative contribution of the categorical trait is more important than the contribution of the continuous trait for the genetic evaluation of the categorical trait of more accurately evaluated animals. When the amount of information for a particular animal for CD is high, the correlated BW information does not increase substantially the accuracy of PBV for CD. When the amount of CD information is low, the contribution of BW is higher but it is limited by the value of the genetic correlation between CD and BW. The accuracy of PBV based on correlated traits is limited by the genetic correlation between these traits, irrespective of the number of correlated records. Thus, the accuracy of breeding values for CD based on BW records is ≤ 0.85.

Table 2. Correlation between predicted breeding values (PBV) for calving difficulty (CD) obtained with complete data and data sets with different percentages of missing CD or birth weight (BW) records in a bivariate threshold-linear animal model

Missing CD records, %	Missing BW records, %	Direct PBV		Maternal PBV	
		All animals in pedigree	Sires with > 100 progeny	All animals in pedigree	Sires with > 100 progeny
	75	0.47	0.44	0.29	0.29
100	50	0.59	0.57	0.39	0.38
	25	0.67	0.65	0.42	0.39
	0	0.85	0.85	0.52	0.52
	100	0.67	0.85	0.75	0.80
	75	0.81	0.91	0.77	0.81
75	50	0.87	0.94	0.79	0.81
	25	0.92	0.94	0.81	0.83
	0	0.94	0.95	0.87	0.87
	100	0.78	0.94	0.85	0.92
	75	0.87	0.95	0.87	0.92
50	50	0.92	0.96	0.90	0.95
	25	0.94	0.96	0.92	0.95
	0	0.97	0.97	0.93	0.94
	100	0.84	0.96	0.89	0.95
	75	0.92	0.98	0.93	0.97
25	50	0.94	0.98	0.94	0.98
	25	0.96	0.98	0.95	0.97
	0	0.99	0.99	0.97	0.98
	100	0.88	0.98	0.93	0.98
	75	0.93	0.99	0.95	0.99
0	50	0.96	0.99	0.97	0.99
	25	0.98	1.00	0.98	0.99

For maternal PBV considering either all animals of the pedigree file or only sires with more than 100 progeny, the correlation estimates followed a similar pattern, with higher values when the categorical trait had more records than the continuous trait, as compared with the opposite cases with more continuous than discrete records. The difference was larger with few or no records of CD compared with few or no records of BW, such as 0.29 vs 0.75 for estimates from all an-

imals using only 25% of the BW records compared with using only 25% of the CD records, respectively. However, the difference decreased as the proportion of missing observations decreased. For example, with 25% of missing CD records and all BW records, the correlation considering all animals was 0.97, and it was 0.98 with all CD records and 25% of missing observations on BW. These results indicate that for maternal genetic evaluation, the loss of CD records is more critical than the loss of BW records because the used genetic correlation between CD and BW is lower for the maternal effect (0.45 and 0.81 for maternal and direct genetic correlations, respectively).

The high values, in general, of correlation coefficients in this study are not surprising given the high genetic correlations used to compute the PBV. In general, due to the high genetic association between CD and BW (PHILIPSSON 1976, MEIJERING 1985, VARONA et al. 1999a), it is expected that the continuous trait supplies information to the genetic evaluation of the discrete trait. The addition of BW records using a bivariate threshold-linear animal model to evaluate CD for genetic direct and maternal effects results in an increase of accuracy for the categorical trait. JANSS and FOULLEY (1993) concluded that advantages of a joint analysis of CD and BW include an increase in the accuracy of evaluation of the categorical trait, and adjustment for selective deletion of observations for the continuous trait. In the situations described in this study, it seems that the relative contribution of CD observations is slightly higher than the information provided by BW records, and especially large differences were observed when a high proportion or all CD records were missing.

The trend for correlations between PBV for CD obtained with complete data and for different percentages of missing CD or BW records is shown in Table 2. In general, correlations for both direct and maternal effects increased as the proportion of missing observations decreased. A similar pattern of increases was observed with 100, 75, 50, and 25% of missing CD records and different levels of missing BW records. For example, when direct PBV were obtained using 100% of the CD records, the correlation estimates increased from 0.88 to 0.98 with decreasing percentage of missing BW records from 100 to 25%. However, when using only BW records, the increase in correlation was higher with the addition of BW observations but from a much lower level. For instance, for PBV considering all animals with 100% missing CD records, the correlations increased from 0.47 to 0.85 when the proportion of missing BW records changed from 75 to 0%.

Correlation estimates above 0.95 were found for 50% of missing CD records and 0% of missing BW records, for 25% of missing CD records and 25 or 0% of missing BW records, and for 0% of missing CD records and 50 or 25% of missing BW records. These results suggest that PBV would not change much for direct genetic evaluation even when 50% of the CD or BW records were missing, as long as 100% of one trait is available. JANSS and FOULLEY (1993) mentioned that the increase in accuracy was particularly substantial for the discrete trait when a high proportion of the data are included. The changes in direct PBV were even smaller

for sires with more than 100 progeny. Correlations > 0.94 were observed when using 50% or more of the CD data, even without BW records.

Table 2 also shows trends for the correlations between maternal PBV for CD. The trends for maternal PBV considering all animals were similar to those obtained for direct PBV, especially when less than 50% of the CD records were missing. Correlations ≥ 0.94 were found for 25% of missing observations on CD and 50, 25, or 0% of missing BW records, and for 0% of missing CD records and 75, 50, or 25% of missing BW records. These results suggest small changes in the prediction of maternal breeding values when less than 50% of the CD or BW records were missing. The changes in prediction of maternal breeding values were even smaller for sires with more than 100 progeny, where correlations > 0.92 were observed when missing less than 50% of the CD records.

The small changes for direct or maternal PBV on CD when using incomplete data show the potential for using the bivariate threshold-linear animal model even with incomplete data on CD or BW.

In beef cattle, observations on CD and BW for all animals are frequently incomplete. In a genetic evaluation of French beef bulls, approximately 20% of BW records are missing (JANSS, FOULLEY 1993); for American Simmental, about 33% of the records have either CD or BW missing (WANG et al. 1997). The results obtained in this study suggest a low potential loss of accuracy for genetic and maternal genetic effects when using incomplete data on CD and BW. Therefore, the bivariate threshold-linear animal model is useful for routine genetic evaluation of calving difficulty for incomplete field data.

Conclusions

For routine genetic evaluation of calving difficulty in beef cattle, the bivariate threshold-linear animal model with the correlated trait birth weight seems to be appropriate for the prediction of direct and maternal genetic effects when less than 50% of the animals have missing records of calving difficulty or birth weight. For genetic prediction of popular sires, a higher proportion of missing records could be tolerated. The relative contribution of calving difficulty records was slightly higher than the information provided by birth weight records, and larger differences were observed when a high proportion or all of the calving difficulty records were missing. The joint analysis of categorical and continuous traits using incomplete data would enable prediction of calving difficulty for animals without records in both traits and/or animals with records in only one of the traits. The obtained results strongly depend on the high genetic correlation between birth weight and calving difficulty.

REFERENCES

FOULLEY J.L., GIANOLA D., THOMPSON R. (1983). Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth weight and pelvic opening. Genet. Sel. Evol. 15: 401-424.

GIANOLA D., FOULLEY J.L. (1983). Sire evaluation for ordered categorical data with a threshold model. Genet. Sel. Evol. 15: 201-244.

GILMOUR A.R., ANDERSON R.D., RAE A.C. (1985). The analysis of binomial data by a generalized linear mixed model. Biometrika 72: 593-599.

HARVILLE D.A., MEE R.W. (1984). A mixed-model procedure for analyzing ordered categorical data. Biometrics 40: 393-408.

HOESCHELE I., TIER B., GRASER H.U. (1995). Multiple-trait genetic evaluation for one polychotomous trait and several continuous traits with missing data and unequal models. J. Anim. Sci. 73: 1609-1627.

JANSS L.L.G., FOULLEY J.L. (1993). Bivariate analysis for one continuous and one threshold dichotomous trait with unequal design matrices and an application to birth weight and calving difficulty. Livest. Prod. Sci. 33: 183-198.

MEIJERING A. (1984). Dystocia and stillbirth in cattle: a review of causes, relations and implications. Livest. Prod. Sci. 11: 143-177.

MEUWISSEN T.H.E., ENGEL B., Van Der WERF J.H.J. (1995). Maximizing selection efficiency for categorical traits. J. Anim. Sci. 73: 1933-1939.

MISZTAL I. (1999). Complex models, more data: Simpler programming? Interbull Proc. Computational Cattle Breeding '99. Bulletin No. 20: 33-42.

PHILIPSSON J., FOULLEY J.L, LEDERER J., LIBORIUSSEN J., OSINGA A. (1979). Sire evaluation standards and breeding strategies for limiting dystocia and stillbirth. Livest. Prod. Sci. 6: 111-127.

RAMIREZ-VALVERDE R., BERTRAND J.K., MISZTAL I. (2001). Comparison of threshold vs linear and animal vs sire models for predicting direct and maternal genetic effects on calving difficulty in beef cattle. J. Anim. Sci. 79: 333-338.

RENAND G., JANSS L.L.G, GAILLARD J. (1990). Sire evaluation for direct effects on dystocia by linear and threshold models. Proc. 4th World Congr. Genet. Appl. Livest. Prod.: 465-467.

VARONA L., MISZTAL I., BERTRAND J.K. (1999a). Threshold-linear versus linear-linear analysis of birth weight and calving ease using an animal model. I. Variance components estimation. J. Anim. Sci. 77: 1994-2002.

VARONA L., MISZTAL I., BERTRAND J.K. (1999b). Threshold-linear versus linear-linear analysis of birth weight and calving ease using an animal model. II. Comparison of models J. Anim. Sci. 77: 2003-2007.

WANG C.S., QUAAS R.L, POLLAK E.J. (1997). Bayesian analysis of calving ease scores and birth weights. Genet. Sel. Evol. 20: 117-143.