ISSN 0370-2731/2003 *Proceedings of the New Zealand Society of Animal Production 63:* 112-115

Comparison of multivariate statistical methods for the prediction of reproductive performance of Holstein Friesian cattle

B.L. HARRIS, J.E. PRYCE AND G. VERKERK[1]

Livestock Improvement Corporation, Private Bag 3016, Hamilton, New Zealand

ABSTRACT

The objectives of this study were to predict success or failure at mating (1/0) using information on 168 first lactation dairy cows from the Dexcel Strain Trial. The farmlets were established by allocating heifers from three genetic strains to eleven groups. The groups were managed under a range of feeding systems, ranging from moderate to generous allowances. Measurements were taken on milk, fat and protein yields, liveweight and condition score and variables derived from these used as predictors of success at mating in a series of models. Reproductive measures used for analysis were cows in calf at 42, 49 and 56 days after the planned start of mating. Three statistical methods were used to determine suitable predictive models of success at mating: logistic regression, discriminant analysis and partition analysis. Partition analysis was the best method of prediction. Kappa values, measuring the degree of agreement between observations and predictions ranged between 0.32 and 0.38 for discriminant analysis, 0.41 and 0.44 for logistic regression and 0.70 and 0.74 for partition analysis. The partition analysis produced a set of rules for predicting likely success at mating. Important rules were CS > 4.75 in month three of lactation and calving before week four after the start of calving for the herd.

Keywords: prediction models; fertility; dairy cattle; partition analysis.

INTRODUCTION

As milk production in New Zealand is predominantly seasonal and reliant on pasture growth, a high pregnancy rate within a short time after the planned start of mating is essential to match feed supply to production. New Zealand research has shown that cow fertility has declined as the proportion of overseas Holstein Friesian (HF) dairy cow genetics has increased (Harris & Kolver, 2001).The proportion of overseas HF has been increasing in the New Zealand population, either through the use of sires that are 100% overseas ancestry, or sires that are part-descendants of overseas bulls. The changing genetic background of New Zealand cattle presents new challenges for farm management and there is considerable interest in the development of management strategies for pastoral systems that also optimise reproductive performance. There are several measurements, already routinely recorded in some dairy herds prior to mating, that could be useful for predicting pregnancy of dairy cows. These measures include test-day yields of fat and protein, their percentages, volume of milk and measures of body condition score (CS) and live weight (LW). The objective of this study was to predict reproductive performance from information available prior to mating using three statistical methods and data from the Dexcel Holstein Friesian strain trial.

The Holstein-Friesian strain trial was established in Hamilton, New Zealand, to investigate the physical and financial performance of three strains of dairy cattle under a range of pasture-based systems that differed in stocking rate and supplementary feeding. The three strains of cows were: Overseas high-genetic merit (OS), New Zealand High genetic merit (NZH) and New Zealand Low genetic merit (1970s high-Breeding-Index strain; NZL).

MATERIALS AND METHODS

Genetic Strains

The strains were established from contract matings in herds within New Zealand. Potential dams were short-listed from the National database if they conformed to ancestry and genetic merit requirements of a strain. The OS strain animals were at least 7/8 overseas genetics and selected for high breeding values for production. The New Zealand High-strain animals were sired by bulls of less than 5/16 overseas ancestry and similar breeding values for production as the OS strain, and dams of at least 3/4 New Zealand ancestry. The 1970-strain sires were sourced through sires that were high Breeding Index in the 1970s and close to 100% New Zealand ancestry. The dams were selected from cows available in the 1998 season that were of low genetic merit. The data used in this study is obtained from the first year of lactation of animals born in 1999 (see Table 2 for the percent Holstein). There were 67 NZH, 67 OS and 36 NZL heifers calving from mid July 2001 to late September 2001.

Feeding Systems

Eleven farmlets were established by allocating heifers from OS and NZH strains to four groups per strain and heifers from NZL into three groups. The eleven groups were managed under a range of pasture-based feeding systems from moderate to generous annual feed allowances per cow. Feed allowance (including pasture silage plus supplements) ranged from 4.5 to 7.0 t DM/cow/yr (Table 1). Feed allowances were manipulated by altering both stocking rate and the availability of supplementary feed. The supplementary feed was predominantly maize silage with soya bean meal being added if required, to maintain an adequate crude protein concentration in the total diet (pasture and supplement). Five of the eleven farmlets had no supplementary feeding.

[1]Dexcel, Private Bag 3221, Hamilton, New Zealand

TABLE 1: Feed allowances, number of cows, stocking rate (cows per hectare) and supplementary feeding levels of the Dexcel Holstein Friesian strain trial (2001/2002).

	Total, annual, feed allowance (pasture plus supplements) offered per cow (t DM/cow)					
	4.5	**5.0**	**5.5**	**6.0**	**6.5**	**7.0**
Strain: OS						
Cows			17	17	17	17
Cows per hectare			3.1	3.1	3.1	3.1
Supplementary feed (t DM/cow)			0	0.5	1.0	1.5
Strain: NZH						
Cows		17	17	17	17	
Cows per hectare		3.4	3.1	3.1	3.1	
Supplementary feed (t DM/cow)		0	0	0.5	1.0	
Strain: NZL						
Cows *	12		12	12		
Cows per hectare	3.8		3.1	3.1		
Supplementary feed (t DM/cow)	0		0	0.5		

(*The smaller numbers of cows and treatments for the NZL strain was due to the smaller number of available heifers).

Reproductive Measures

Reproductive measures were binary traits for success or failure to conceive at each of 42, 49 and 56 days after the planned start of mating (PSM). The planned start of mating was October 2nd 2001. The measurements were chosen as indicators of compact calving patterns in pasture based systems (Verkerk, 2003). There was a seven-week artificial-insemination (AI) period, followed by a seven-week natural-mating period. Cows that failed to ovulate based on progesterone samples in milk (McNaughton *et al.*, 2003), were treated with a CIDR device (InterAg, Hamilton) for 6 days. The treatment was applied to all cows that had not ovulated, but had calved for more than 28 days. At 24 hrs after CIDR removal, 1mg of oestradiol benzoate was administered. Treatment was carried out on groups of animals that were formed according to calving spread. Eleven OS heifers failed to become pregnant and were scored as failed to conceive. Early pregnancy detection was by transrectal ultrasonographic scanning. The first scan took place at approximately day 60 after PSM to detect pregnancies from the first round of mating. There was a further scan at 30-35 days after the bulls had been removed. The age of embryo was evaluated at each scanning to confirm the date of conception relative to known AI dates. A final pregnancy test was done towards the end of lactation (late May 2001). Scanning data were subsequently confirmed using calving dates from Spring, 2002.

Statistical analysis

Milk yield (MY) and fat and protein percentages (F%, P%) were measured weekly and live weight and condition score (CS) were assessed weekly until mating and then fortnightly. CS was scored on a scale of 1 to 10 (Harris, 2002). Monthly means were calculated from each of these, first according to the number of months in milk for each cow and then according to the number of months of the trial (from the first cow to calve). Cows that calve late in the season do not have information on their third month of lactation prior to the start of the mating period. Thus, the data on cows falling into this category were regarded as missing. Fat-to-protein ratios were calculated for each cow for the first three months of lactation and for the first three months of the trial (determined from first calving date). Changes in condition score were calculated between the first and second and between the second and third months of lactation, as well as from the first three months of the trial from the first calving. Preliminary analyses showed that weeks of lactation measures were more informative than weeks of trial, demonstrating that stage of lactation is more useful than measures affected by grass supply. It is worth noting that feed supply was not limiting for any of the treatments in the first season of the trial (data used in this study). Condition score changes, fat and protein yields and fat-to-protein ratios were less informative than CS, F% and P% for each month. Subsequent analyses were done using: average monthly CS, LW, MY, F%, P% for the first three months of lactation, treatment for non-cycling (0/1), feed allowance, % North American genetics (%Hol) and calving week (from first calving). Raw means of these data are presented in Table 2.

Data were analysed using JMP version 5 (SAS, 2002). Three methods of predicting P42, P49 and P56 were used: nominal logistic regression, discriminant analysis and partition analysis. Logistic nominal regression is multivariate regression for binomial or categorical data. It uses correlations between variables to determine the relationship between response and explanatory variables. Discriminant analysis seeks to find a way to predict a classification variable based on known continuous responses.

Partition analysis is often used as a data-mining technique because it is good for exploring relationships without having a good prior model. It is popular as a diagnostic tool for analysing symptoms and diagnoses of a given illness to provide a hierarchy of questions in order to provide a quick initial diagnosis. Partition analysis recursively sub-divides data creating a tree of partitions. Each split is chosen to maximise the difference in the responses between the two branches of the split; the most significant split is determined by the largest likelihood-ratio chi-square statistic. The data were split a maximum of 12 times. Predictions were compared to phenotypes

TABLE 2: Means of reproductive performance, North American Holstein percentage, milk, fat and protein yields and their components, live weight and condition score by genetic strain for first lactation heifers.

		NZH			NZL			OS	
P42 (%)		67.2			75			49.2	
P49 (%)		70.1			77.7			61.5	
P56 (%)		71.6			83.3			69.2	
Holstein %		23			7.5			90.1	
Treated for non-cycling (%)		74.6			53			34.8	
Strain		**NZH**			**NZL**			**OS**	
Month in milk	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Milk yield (kg/d)	19.8	22.1	19.9	17.8	19.5	18.1	21.2	22.7	20.3
Fat %	4.78	4.40	4.35	4.86	4.32	4.28	4.57	4.01	3.90
Protein %	3.47	3.31	3.31	3.23	3.05	3.08	3.39	3.23	3.24
Fat yield (kg/d)	0.94	0.97	0.86	0.86	0.84	0.77	0.96	0.91	0.79
Protein yield (kg/d)	0.68	0.73	0.66	0.57	0.59	0.56	0.71	0.73	0.65
Live weight (kg)	436	413	409	418	403	407	454	424	426
Condition score	4.66	4.29	4.29	4.80	4.56	4.61	4.43	3.89	4.00

(NZH: New Zealand High; NZL; New Zealand Low; OS: overseas; P42: % of cows pregnant at 42 days; P49 % of cows pregnant at 49 days; P56 % of cows pregnant at 56 days)

TABLE 3: Kappa values measuring the degree of agreement between the model predictions and the data and the percentage of incorrectly classified animals for the statistical models

Method	**Logistic Regression**			**Discriminant Analysis**			**Partition Analysis**		
Trait	**P42**	**P49**	**P56**	**P42**	**P49**	**P56**	**P42**	**P49**	**P56**
Kappa	0.44	0.43	0.41	0.36	0.32	0.38	0.72	0.70	0.74
% Incorrectly classified	23	20	16	31	39	27	13	12	9

(P42: % pregnant at 42 days; P49 % pregnant at 49 days; P56 % pregnant at 56 days)

using contingency analysis for all three models.

RESULTS

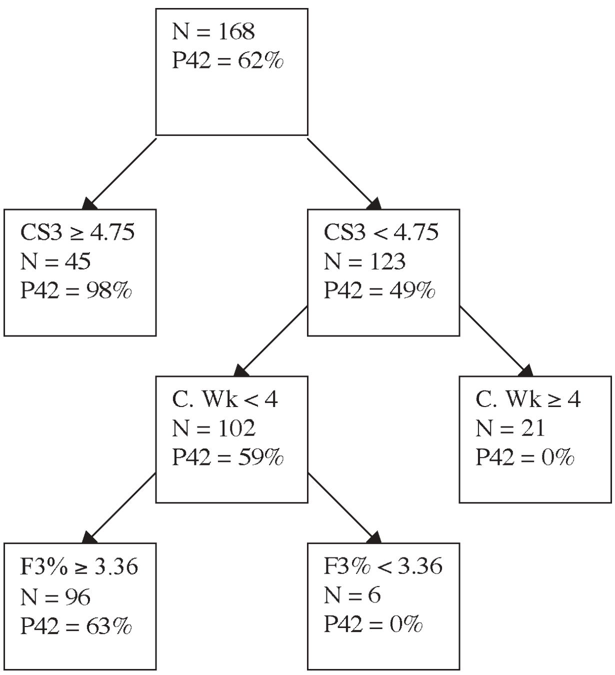
Partition analysis was the best method of prediction (Table 3). Kappa values, measuring the degree of agreement between observations and predictions ranged between 0.32 and 0.38 for discriminant analysis, 0.41 and 0.44 for logistic regression and 0.70 and 0.74 for partition analysis.

The 'best' prediction models for P42, P49 and P56 differed depending on the type of analysis used, although CS and calving week consistently emerged as important predictors in all three analyses. All the explanatory variables were used as predictors in both logistic regression and discriminant analysis. The contribution and importance of each of the predictors varied according to which pregnancy trait was used as the response variable. The three most important predictors of P42 using logistic regression were MY in month two of lactation, calving week and milk yield in month three. Using discriminant analysis the three most important predictors were: CS in month two, North American Holstein % and calving week.

The partition model predicted that all cows with CS greater than or equal to 4.75 in month three of lactation were pregnant by day 42 (data set mean of CS in month 3 was 4.29 and SD was 0.51). All cows with CS less than 4.75 in month three of lactation and calving after week four from start of calving all failed to conceive by day 42. These were also the primary rules in the analysis of P49 and P56, but after these splits different rules arose reflecting the relationship between strain composition and proportion pregnant. Rules to identify cows as pregnant or not pregnant are shown in Figure 1. There was a relatively small change between P42 and P56 in NZL and NZH, but for the OS cows P42 and P56 are 57% and 80%, respectively. High P% in months two and three and low Hol% were important predictors of P42, while high CS, LW and MY in month one were important predictors of success for P49 and P56.

The data were also analysed within strain. The first two splits for NZH in the partition analysis were on CS greater than or equal to 4.75 and then on calving week less than four after planned start of calving. These two

FIGURE 1: Partition analysis example of data-splitting rules for percentage of cows pregnant at 42 days after planned start of mating (only partially shown), where N = number of animals;CS3 = CS at month 3 of lactation; C. Wk = calving week and F3% = Fat percentage in month 3 of lactation



splits were the same as splits of the entire data set across strain. In contrast, for NZL strain, the first two splits were for animals with milk yield in month three after calving of greater than or equal to 19.2kg/day that were all in calf by day 42 (16 out of 36 cows). Of the remaining cows, 7 animals with milk yield in month three after calving of greater than 17.2kg/day all failed to get in calf by 42 days. The average milk yield of NZL in month 3 after calving was 18.1kg/d. In OS animals, the probability of conceiving within 42 days was improved if fat percentage was greater than or equal to 3.46% in month three of lactation and within this group, protein percentage in month one of lactation was less than 3.69.

DISCUSSION

The main purpose of the present study was to investigate the usefulness of three statistical models in determining reproductive performance, using information available prior to mating. The results presented here are from first lactation animals only and therefore should be treated with caution, as results may differ from data that includes older animals. We plan to repeat the analysis with data from subsequent years of the trial.

Partition analysis is an effective way of identifying characteristics that relate to traits of importance in animal production. The method works by identifying thresholds in explanatory variables in such a way that the resultant partitions correspond well to the explanatory variable rather than estimating a continuous relationship between dependent and independent variables (which is the approach used in regression-based analyses). In our analyses, partition analysis was better than both logistic regression and discriminant analysis in prediction of P42, P49 and P56, by having the fewest wrongly classified animals.

An important predictor of reproductive performance in the partition analysis was CS greater than 4.75. Condition score was also an important predictor in the other methods of analysis. Research elsewhere has shown that CS is a useful predictor of reproductive success and that there is a genetic basis to the relationship (Berry *et al.*, 2002, Dechow *et al.*, 2001; Pryce *et al.*, 2001).

It is interesting to note that Holstein percent did not emerge as an important predictor of the pregnancy measures, even though P42 of NZL was 18% higher than OS. Neither was strain a predictor of the pregnancy measures in the analysis across strain. Yet, within strain, there were differences between partition models. For OS cows, it could be hypothesised that the relationship is connected to energy balance, as high fat percentage and low protein percentage were important. Fat is energy demanding to produce, suggesting that cows with high milk fat percentages do not have the same degree of energy deficit as those with low percentages.

The limitation with the present study is that the data set consisting of only first lactation animals is small. However, there are advantages in using data from research herds, as many more explanatory variables are generally available than in 'national data' and the environments and feeding systems are usually more adequately described.

Partition analysis is ideal for defining "rules of thumb" and can then be applied to data sets in which there are several explanatory and dependent variables. The results from this study should be viewed as providing useful guidelines for herd managers and extension workers. Management of first-lactation heifers to achieve high CS in the second or third month of lactation, and to ensure that a high proportion calve early in the herd's calving period, will increase the probability that these animals will become pregnant in the first six to eight weeks of the mating period.

ACKNOWLEDGEMENTS

This work was funded by NZDB Global Programmes, Dairy InSight and Livestock Improvement Corporation.

REFERENCES

Dechow, C.D.; Rogers, G.W.; Clay, J.S. 2001: Heritabilities and correlations among body condition scores, production traits and reproductive performance. *Journal of dairy science 84:* 266-275.

Berry D. P.; Buckley, F.; Dillon, P.; Evans R. D.; Rath M.; Veerkamp R. F. 2002: Genetic Parameters for Level and Change of Body Condition Score and Body Weight in Dairy Cows. *Journal of dairy science 85:* 2030-2039.

Harris, B. L. 2002: Genetics of body condition score in New Zealand dairy cattle. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production. CD-ROM communication no. 01-03.*

Harris, B.L.; Kolver, E.S. 2001: Review of Holsteinization on intensive pastoral dairy farming in New Zealand. *Journal of dairy science 84 (E. Suppl.):* E56-E61.

McNaughton, L.A.; Verkerk, G.A.; Parkinson, T.J.; Macdonald, K.A.; Holmes, C.W. 2003: Postpartum anoestrous intervals and reproductive performance of three genotypes of Holstein-Friesian dairy cattle. *Proceeding New Zealand Animal Production Society 63:* 77-79

Pryce, J.E.; Coffey, M.P.; Simm, G. 2001: The relationship between body condition score and reproductive performance. *Journal of dairy science 84*: 1508-1515.

SAS. 2002. JMP Version 5.0, SAS Institute, Cary, NC, USA.

Verkerk, G.A. 2003. Pasture-based dairying: challenges and rewards for New Zealand producers. *Theriogenology 59(2):* 553 - 561.