CUBIC SPLINES FOR ESTIMATING LACTATION CURVES AND GENETIC PARAMETERS OF FIRST LACTATION HOLSTEIN COWS TREATED WITH BOVINE SOMATOTROPIN

Bruce J. DeGroot

Jeffrey F. Keown

Stephen D. Kachman

Dale Van Vleck

See next page for additional authors

Recommended Citation

Author Information

Bruce J. DeGroot, Jeffrey F. Keown, Stephen D. Kachman, and Dale Van Vleck

CUBIC SPLINES FOR ESTIMATING LACTATION CURVES AND GENETIC PARAMETERS OF FIRST LACTATION HOLSTEIN COWS TREATED WITH BOVINE SOMATOTROPIN

Bruce J. DeGroot[1], Jeffrey F. Keown[1], Stephen D. Kachman[2], and L. Dale Van Vleck[3]

[1]Department of Animal Science, University of Nebraska-Lincoln
[2]Department of Biometry, University of Nebraska-Lincoln
[3]USDA, ARS, USMARC, Lincoln

Abstract

The objective was to estimate genetic parameters and fit lactation curves for cows treated or not treated with bovine somatotropin (bST) and fit specific lactation curves for each animal for both random genetic and permanent environmental components from individual test-day milk, fat, and protein yields with a cubic spline model. A total of 70,752 test-day observations for first lactation Holstein cows recorded as treated bST and 73,387 test-day observations for untreated cows that calved between 1994 and early 1999 were obtained from Dairy Records Management Systems in Raleigh, North Carolina. The model included herd test-day, age at first calving, bST treatment, and days in lactation when test-day yield was recorded as fixed effects. Cubic splines were fitted for the overall lactation curve, additive genetic effects, and permanent environmental effects. The cubic splines used five predetermined intervals between days 0, 50, 135, 220, and 305. Estimates of the (co)variances for the random components of cubic spline model with five knots were obtained with REML. Estimates of genetic parameters were calculated for the average test day model within each of the ten 30-d test day intervals. The estimates of heritability for milk, fat, and protein yields ranged from 0.09 to 0.15, 0.06 to 0.10, and 0.08 to 0.15 for test-day one to test-day ten. Estimates of genetic correlations between test-days ranged from 0.99 to 0.34 for milk yield, 0.99 to 0.49 for fat yield, and 0.99 to 0.36 for

protein yield. Estimates of phenotypic correlations between test-days ranged from 0.67 to 0.27 for milk yield, 0.52 to 0.16 for fat yield, and 0.60 to 0.19 for protein yield. Differences between bST treated and untreated cows of 2 to 4 kg and 0.10 to 0.16 kg for milk and fat yields (smaller for protein yield) at day 90 were maintained until about day 305 of lactation.

1. Introduction

There has been a growing interest of changing the data used for genetic evaluation of dairy cattle from combined 305-day mature equivalent lactation yields to test-day yields. The new analysis would use several daily measurements on an individual cow over the course of the lactation, usually taken once a month (test-day). The advantages of test-day models (TDM) compared with traditional models used with 305-day lactation yields are direct adjustments for fixed effects on the day the records are collected and that end-of-lactation yields do not need to be extended for culled cows or for cows with records in-progress. The disadvantages of TDM are that more records need to be analyzed and that models tend to be more complex which have more parameters to be estimated than with traditional models.

Various models have been proposed for analysis of test-day records. An earlier proposal was to use a multiple-trait model, with each test-day record on an individual cow treated as a separate trait. The major disadvantage was that this model was computationally unfeasible for national evaluations. Another model proposed was a two-step TDM (Wiggans and Goddard, 1997). In step one, the test-day records are adjusted for known environmental effects, such as age, season, year, stage of lactation, location, and milking frequency, using a fixed effects model. The residuals from this model for a cow during lactation are combined into one lactation record and analyzed as a single trait.

Another set of TDM was based on incorporating the lactation curve into the analysis with some sort of curve fitting method. The first such TDM used fixed regressions to fit overall lactation curve using regression coefficients that are nested within classes of fixed effects. The traditional additive genetic and permanent environmental random effects are assumed to have constant variances throughout the lactation. Correlations between yields at different days in milk were assumed to be the same regardless of time elapsing between days of test-day measures. The assumption that the variances are homogeneous throughout the lactation is difficult to justify. Another type of TDM was developed to include both fixed and random regressions (Henderson Jr., 1982; Schaeffer and Dekkers, 1994). With such models, the shape of the lactation curve is modeled as a function of fixed effects and random genetic and permanent environmental effects associated with a cow are modeled as deviations from the fixed curve. This type of model can accommodate heterogeneous additive genetic and permanent environmental variances over the course of the lactation and correlations between yields at different days in the lactation can be less than one and can be different for different pairs of days.

The covariates used in the regression part of TDM are usually functions of days in lactation where the measurement was made. Ptak and Schaeffer (1993) used various combinations of linear, quadratic, and logarithmic functions of days as covariates. Other authors developed functions, for example Wilmink (1987), that model lactation curves based on the biology of lactation. Polynomials, such as orthogonal Legendre polynomials, have also been presented as general covariates for fitting the lactation curve (Jamrozik and Schaeffer, 1997; Meyer and Hill, 1997). White et al. (1999) described a method of using smoothing cubic splines to model the lactation curve using TD records. The method consists of fitting a series of cubic

polynomials that are continuous and centered through knots, which separate days in lactation into intervals. Spline models are flexible and only four (co)variance parameters need to be estimated compared to polynomials, which require 0.5q(q+1) (co)variances parameters to be estimated for each random function where q is the order of the polynomial.

The objective of this study was to measure the response in lactation yield for first lactation Holstein cows treated with bovine somatotropin (bST), growth hormone used to stimulate milk production compared with untreated cows by fitting the mean lactation curves with a cubic spline model and to estimate genetic parameters for test-day yields within the lactation.

2. Materials and Methods

First lactation test-day yields of Holstein cows were obtained from Dairy Records Management Systems of Raleigh, North Carolina. The original data consisted of 2.5 million lactation records from Dairy Herd Improvement (DHI) herds throughout the United States. In 1994 DHI began recording of whether cows were treated with bST, POSILAC produced by Monsanto Co. Each test-day record was coded whether or not the cow was or was not treated with bST. The bST treatment is administrated subcutaneously every two weeks and usually begins by the ninth week of lactation. To compare cows with and without bST treatment, only herds in which at least half of the cows received bST treatment were included in the analysis. Each cow was required to have a lactation of at least 305-day with two times a day milking and with at least eight test-day records during the first lactation. The data set after edits comprised 215 herds and 144,139 test-day yields of 17,168 cows that calved between 1994 and 1999 for

milk, fat, and protein, with 70,752 test-day records for untreated cows and 73,387 test-day records for cows treated with bST .

Cubic spline Model

A single trait cubic spline model was used to fit fixed effects lactation curves and deviations for each animal for both random genetic and permanent environmental components. The cubic spline model consists of a series of piecewise cubic polynomials that are defined on each of the intervals for days in lactation. The model is constrained so that cubic spline function and its first two derivatives are continuous at the knots (breakpoints) over the course of the interval (White et al., 1999). For q knots, the piecewise cubic spline function on an interval $x_j \leq t \leq x_{j+1}$, $(j = 1, \ldots, q-1)$, is represented by:

$$g(t) = \frac{(t - x_j)g_{j+1} + (x_{j+1} - t)g_j}{x_{j+1} - x_j} - \frac{1}{6}(t - x_j)(x_{j+1} - t)\left[\left(1 + \frac{t - x_j}{x_{j+1} - x_j}\right)v_{j+1} + \left(1 + \frac{x_{j+1} - t}{x_{j+1} - x_j}\right)v_j\right] \quad (1)$$

where t is the actual day of measurement within an interval $x_j \leq t \leq x_{j+1}$, g_j and g_{j+1} are linear coefficients, and v_j and v_{j+1} are cubic spline coefficients. The first term produces a linear interpolation between the knots. The second term represents a cubic deviation, which vanishes at the knots. The spline function (1) can be written in matrix notation as

$$g(t) = \xi(t)^T g + \lambda(t)^T v \quad (2)$$

The coefficients (g_j, v_j) of cubic spline satisfy the q-2 constraints

$$\frac{g_{j+1} - g_j}{x_{j+1} - x_j} - \frac{g_j - g_{j-1}}{x_j - x_{j-1}} = \frac{1}{6}(x_j - x_{j-1})v_{j-1} + \frac{1}{3}(x_{j+1} - x_{j-1})v_j + \frac{1}{6}(x_{j+1} - x_j)v_{j+1} \quad (3)$$

$(j = 2,\ldots,q-1)$, which ensure the continuity of the first derivate (White et al., 1999). The cubic spline is also defined to be linear outside the two end knots and then have $v_1 = v_q = 0$. The

constraints in (2) in matrix notation are $\mathbf{Q}^\mathsf{T}\mathbf{g} = \mathbf{Rv}$, which satisfies the condition for natural cubic splines. Let $h_j = x_{j+1} - x_j$, then elements of $\mathbf{Q}$ with dimensions of q x (q-2) are $q_{j-1,j} = h_{j-1}^{-1}$, $q_{j,j} = -h_{j-1}^{-1} - h_j^{-1}$, and $q_{j+1,j} = h_j^{-1}$ all for j = 2, …, q-1. The elements of $\mathbf{R}$ with dimensions (q-2) x (q-2) are $r_{j,j} = (1/3)(h_{j-1} + h_j)$ for j = 2, …, q-1 and $r_{j,j+1} = r_{j+1,j} = (1/6)(h_j)$ for j = 2, …, q-2. All other elements of $\mathbf{Q}$ and $\mathbf{R}$ are zero.

Verbyla et al. (1999) and White et al. (1999) demonstrated how to incorporate the natural cubic spline into the standard mixed model when the knots are determined before the analysis. White et al. (1999) showed that (2) could be written for the constraints as

$$g = \mathbf{Xb} + \mathbf{Q}(\mathbf{Q}^\mathsf{T}\mathbf{Q})^{-1}\mathbf{Rv} \tag{4}$$

where the intercept and slope are in $\mathbf{b}$ and the spline coefficients are in $\mathbf{v}$ and $\mathbf{X}$ is a (q x 2) matrix with two columns, the first vector of ones and the second vector of the actual days of the measurement. The equation (4) can be written in the standard mixed model equation

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z_s u_s} + \mathbf{e} \tag{5}$$

where $\mathbf{X\beta}$ fits the linear regression and $\mathbf{Z_s u_s}$ fits a cubic spline based on number of knots and the random vectors $\mathbf{u_s} \sim \mathbf{N(0, R}\sigma_s^2)$ and $\mathbf{e} \sim \mathbf{N(0, I}\sigma^2)$. The model (5) can be further written as a random regression animal model:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z_s s} + \mathbf{W_a a_l} + \mathbf{Z_a a_s} + \mathbf{W_{pe} pe_l} + \mathbf{Z_{pe} pe_s} + \mathbf{e} \tag{6}$$

where $\mathbf{y}$ is a vector of test-day milk, fat, or protein yields. The component β is a vector of fixed effects including fixed regressions coefficients and $\mathbf{X}$ is the incidence matrix for the fixed effects, which are cows treated or not treated with bST, herd x test-day, covariate of age at the beginning of lactation, and covariate of day in milk for each test-day record. The random effects

of the model are the following: **s** is a vector of overall spline coefficients, $\mathbf{a_l}$ is a vector of animal genetic intercept and slope coefficients, $\mathbf{a_s}$ is a vector of animal genetic spline coefficients, $\mathbf{pe_l}$ is a vector of permanent environmental intercept and slope coefficients, $\mathbf{pe_s}$ is a vector of permanent environmental spline coefficients, **e** is vector of residual effects, $\mathbf{W_a}$ and $\mathbf{W_{pe}}$ are incidence matrices for animal genetic and permanent environmental linear effects, and $\mathbf{Z_s}$, $\mathbf{Z_a}$, and $\mathbf{Z_{pe}}$ are covariate matrices for overall spline, animal genetic, and permanent environmental spline effects based on the number of predetermined knots. From equation (4) with **b** and **v** independent, the random effects are defined as

$$\mathbf{s} \sim \mathbf{N}(0, \mathbf{R}\sigma^2_s), \; \mathbf{a_S} \sim \mathbf{N}(0, \mathbf{A} \otimes \mathbf{R}\sigma^2_{as}), \; \mathbf{pe_S} \sim \mathbf{N}(0, \mathbf{I} \otimes \mathbf{R}\sigma^2_{pes})$$

$$\mathbf{a_l} \sim \mathbf{N}(0, \mathbf{A} \otimes \Phi_\mathbf{a}), \; \mathbf{pe_l} \sim \mathbf{N}(0, \mathbf{I} \otimes \Phi_\mathbf{pe}), \; \mathbf{e} \sim \mathbf{N}(0, \mathbf{I}\sigma^2),$$

$$\Phi_\mathbf{a} = \begin{pmatrix} \sigma^2_{a_i} & \sigma_{a_i a_{sl}} \\ \sigma_{a_i a_{sl}} & \sigma^2_{a_{sl}} \end{pmatrix}, \text{ and } \Phi_\mathbf{pe} = \begin{pmatrix} \sigma^2_{pe_i} & \sigma_{pe_i pe_{sl}} \\ \sigma_{pe_i pe_{sl}} & \sigma^2_{pe_{sl}} \end{pmatrix}.$$

The predetermined knots were at days 0, 50, 135, 220, and 305. The ASREML program was used to estimate effects of bST for milk, fat, and protein yields after convergence of estimates of the (co)variance components (Gilmour et al., 1997).

3. Results and Discussion

Estimates of genetic, permanent environmental, and phenotypic variances and heritabilities for milk, fat, and protein yield are in Table 1 as calculated for the midpoints of the 10 intervals. Milk yield had estimates of heritability that ranged from 0.09 to 0.15. Heritability dropped slightly from Test 1 to Test 2 and then increased steadily to 0.15 at Test 10. The estimates of heritability were low compared to estimates reported in a review of different types of TDM used to estimate genetic parameters, but were comparable to estimates reported in other

studies reviewed (Misztal et al., 2000). Estimates of heritability for fat yield ranged from 0.06 to 0.10 and protein yield ranged from 0.08 to 0.15 over the course of the lactation. The general pattern of the heritability estimates was similar to those reported for milk yield.

Estimates of genetic and phenotypic correlations among test-days for midpoints of the intervals for milk, fat, and protein yields are in Tables 2, 3, and 4, respectively. Estimates of genetic correlations ranged from 0.99 to 0.34 for milk yield, 0.99 to 0.49 for fat yield, and 0.99 to 0.36 for protein yield. The patterns of estimates among test-day yields were similar for milk, fat, and protein yields. Estimates of correlations between yields decreased rapidly for days closer together and slowly decreased as days measured were further apart. Estimates of phenotypic correlations ranged from 0.67 to 0.27 for milk yield, 0.52 to 0.16 for fat yield, and 0.60 to 0.19 for protein yield. The estimates of phenotypic correlations followed the same pattern as estimates of genetic correlations.

Figures 1, 2, and 3 show the mean lactation curves for treated and untreated bST cows for milk, fat, and protein yields. The cubic spline model provided considerable flexibility in fitting the lactation curve for the three yield traits. The lactation curve for milk yield showed a rapid increase in production to about day 60 and followed by a steady decline. The difference between treated and untreated bST cows was 2 to 4 kg, which was maintained from day 90 to day 305. The lactation curves for fat yield were little differences for untreated cows and treated cows. The 0.10 to 0.16 kg difference between treated and untreated cows was maintained from day 90 to day 280. Differences between treated and untreated bST cows were small for protein yield.

5. Conclusions

The cubic spline model provided flexibility for estimated genetic parameters and fitting lactation curves from test-day yields. Estimates of heritability increased over with days in lactation. Estimates of genetic and phenotypic correlations decreased as days between when yields were measured. Lactation curves calculated from cubic spline were different between treated and untreated bST cows. Cubic spline models could also be used for other types of longitudinal data that have non-linear curves, such as growth from birth to maturity.

6. References

Gilmour, A. R., R. Thompson, B. R. Cullis, and S. J. Welham. 1997. ASREML Manual. New South Wales Dep. Agric., Orange, 2800, Australia.

Green, P. J. and B. W. Silverman. 1994. Nonparametric regression and generalized linear models. London: Chapman and Hall.

Henderson, C. R., Jr. 1982 Analysis of covariance in the mixed model: Higher level, nonhomogenous, and random regressions. Biometrics 38:623-640.

Jamrozik, J., and L. R. Schaeffer. 1997 Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. J. Dairy Sci. 80:762-770.

Meyer, K. and W. G. Hill. 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or 'repeated' records by restricted maximum likelihood. Livest. Prod. Sci. 47:185-200.

Misztal, I., T. Strabel, J. Jamrozik, E. A. Mäntysaari, and T. H. E. Meuwissen. 2000. Strategies for estimating the parameters needed for different test-day models. J. Dairy Sci. 83:1125-1134.

Ptak, E. and L. R. Schaeffer. 1993. Use of test day yields for genetic evaluation of dairy sires and cows. Livest. Prod. Sci. 34:23-34.

Schaeffer, L. R., and J. C. M. Dekkers. 1994. Random regressions in animal models for test-day production in dairy cattle. Proc. 5th World Congr. Genet. Appl. Livest Prod. Guelph, Ontario, Canada XVIII:433-446.

Verbyla, A. P., B. R. Cullis, M. G. Kenward, and S. J. Welham. 1999. The analysis of designed experiments and longitudinal data by using smoothing splines. Appl. Statist. 48, Part 3:269-311.

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.

Wiggans, G. R. and M. E. Goddard. 1997. A computationally feasible test day model for genetic evaluation of yield traits in the United States. J. Dairy Sci. 80:1795-1800.

Wilmink, J. B. M. 1987 Adjustment of test-day milk, fat, and protein yields for age season and stage of lactation. Livest. Prod. Sci. 16:335-348.

Table 1. Estimates of genetic variance (σ^2_a), permanent environmental variance (σ^2_{pe}), phenotypic variance (σ^2_p), and heritability (h^2) for ten representative days in milk (DIM) for milk, fat, and protein yields (kg)

Test	DIM	Milk Yield				Fat Yield				Protein Yield			
		σ^2_a	σ^2_{pe}	σ^2_p	h^2	σ^2_a	σ^2_{pe}	σ^2_p	h^2	σ^2_a	σ^2_{pe}	σ^2_p	h^2
1	18	3.43	20.58	36.01	0.095	0.0049	0.0395	0.0795	0.062	0.0027	0.0156	0.0316	0.085
2	46	3.01	17.68	32.70	0.092	0.0045	0.0320	0.0716	0.063	0.0024	0.0129	0.0287	0.085
3	76	2.98	17.16	32.14	0.093	0.0049	0.0296	0.0696	0.071	0.0024	0.0126	0.0283	0.084
4	106	3.20	18.04	33.24	0.096	0.0057	0.0302	0.0710	0.080	0.0025	0.0137	0.0295	0.084
5	136	3.57	19.45	35.03	0.102	0.0065	0.0320	0.0735	0.088	0.0027	0.0153	0.0314	0.087
6	167	3.96	20.37	36.33	0.109	0.0070	0.0325	0.0745	0.094	0.0031	0.0164	0.0328	0.093
7	196	4.31	20.64	36.95	0.117	0.0072	0.0315	0.0738	0.098	0.0035	0.0168	0.0336	0.103
8	227	4.74	20.70	37.44	0.126	0.0073	0.0298	0.0722	0.101	0.0040	0.0169	0.0342	0.116
9	256	5.18	20.64	37.82	0.137	0.0072	0.0279	0.0701	0.102	0.0046	0.0169	0.0348	0.131
10	288	5.91	21.72	39.63	0.149	0.0073	0.0283	0.0707	0.104	0.0054	0.0179	0.0367	0.146

Applied Statistics in Agriculture

Table 2. Estimates[1] of genetic and phenotypic correlations for milk yield of representative days in milk (DIM) among test-days

| DIM | DIM | | | | | | | | | |
	18	46	76	106	136	167	196	227	256	288
18	...	0.96	0.86	0.73	0.61	0.52	0.45	0.40	0.37	0.34
46	0.62	...	0.96	0.88	0.78	0.69	0.62	0.56	0.51	0.46
76	0.55	0.61	...	0.97	0.91	0.84	0.78	0.71	0.65	0.57
106	0.47	0.55	0.61	...	0.98	0.94	0.88	0.81	0.74	0.66
136	0.40	0.50	0.58	0.64	...	0.99	0.95	0.89	0.82	0.73
167	0.34	0.45	0.54	0.61	0.65	...	0.99	0.95	0.89	0.81
196	0.31	0.41	0.51	0.58	0.63	0.66	...	0.99	0.95	0.87
227	0.29	0.38	0.46	0.53	0.58	0.63	0.67	...	0.99	0.94
256	0.28	0.35	0.42	0.48	0.53	0.59	0.63	0.67	...	0.98
288	0.27	0.33	0.38	0.42	0.46	0.51	0.57	0.62	0.67	...

[1]Genetic above the diagonal and phenotypic below the diagonal

Table 3. Estimates[1] of genetic and phenotypic correlations for fat yield of representative days in milk (DIM) among test-days

| DIM | DIM | | | | | | | | | |
	18	46	76	106	136	167	196	227	256	288
18	...	0.95	0.82	0.69	0.60	0.56	0.55	0.54	0.52	0.49
46	0.51	...	0.96	0.88	0.80	0.75	0.73	0.72	0.68	0.63
76	0.43	0.48	...	0.97	0.93	0.88	0.85	0.82	0.80	0.77
106	0.35	0.43	0.48	...	0.99	0.96	0.92	0.88	0.83	0.76
136	0.28	0.37	0.45	0.50	...	0.99	0.96	0.91	0.86	0.77
167	0.23	0.32	0.41	0.48	0.52	...	0.99	0.95	0.90	0.80
196	0.19	0.28	0.37	0.44	0.49	0.52	...	0.99	0.95	0.85
227	0.17	0.25	0.32	0.39	0.44	0.48	0.51	...	0.98	0.92
256	0.16	0.22	0.28	0.33	0.37	0.42	0.46	0.49	...	0.97
288	0.16	0.19	0.23	0.26	0.29	0.33	0.38	0.43	0.48	...

[1]Genetic above the diagonal and phenotypic below the diagonal

Table 4. Estimates[1] of genetic and phenotypic correlations for protein yield of representative days in milk (DIM) among test-days

DIM	DIM									
	18	46	76	106	136	167	196	227	256	288
18	...	0.98	0.91	0.82	0.72	0.63	0.55	0.48	0.42	0.36
46	0.53	...	0.98	0.92	0.84	0.76	0.69	0.61	0.55	0.49
76	0.45	0.51	...	0.98	0.93	0.87	0.81	0.74	0.68	0.62
106	0.37	0.46	0.52	...	0.98	0.95	0.90	0.84	0.79	0.72
136	0.30	0.40	0.49	0.55	...	0.99	0.96	0.91	0.86	0.81
167	0.24	0.35	0.45	0.53	0.57	...	0.99	0.96	0.93	0.88
196	0.21	0.32	0.42	0.50	0.55	0.59	...	0.99	0.97	0.93
227	0.19	0.29	0.38	0.45	0.51	0.56	0.59	...	0.99	0.97
256	0.19	0.26	0.34	0.40	0.46	0.51	0.56	0.60	...	0.99
288	0.19	0.24	0.29	0.34	0.38	0.44	0.50	0.56	0.60	...

[1] Genetic above the diagonal and phenotypic below the diagonal

Figure 1. Fitted mean spline lactation curve for milk yield for treated and untreated bST cows.
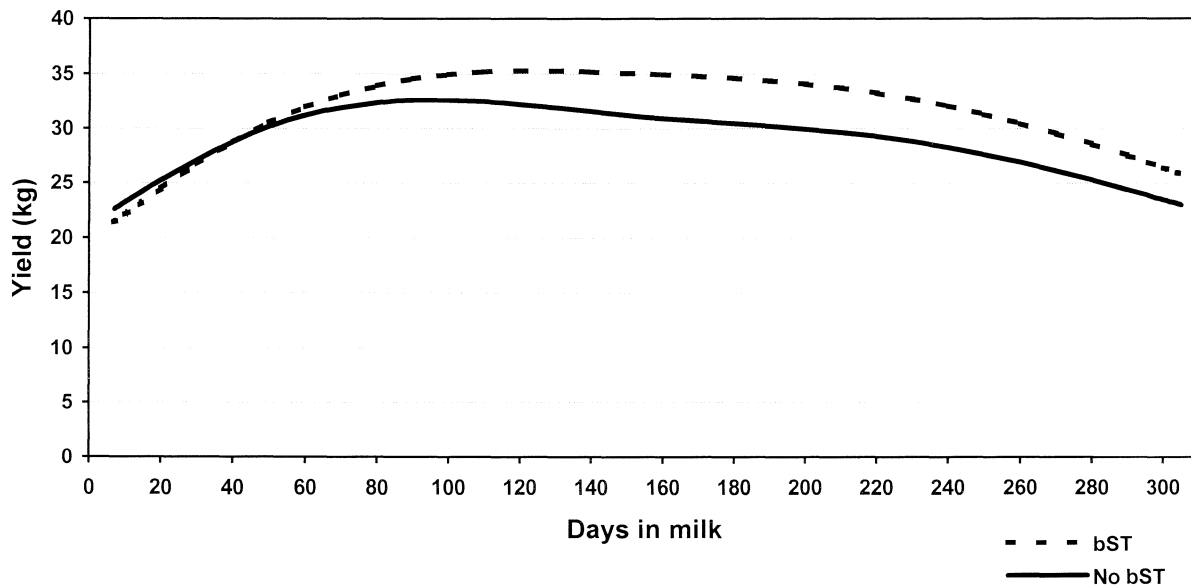
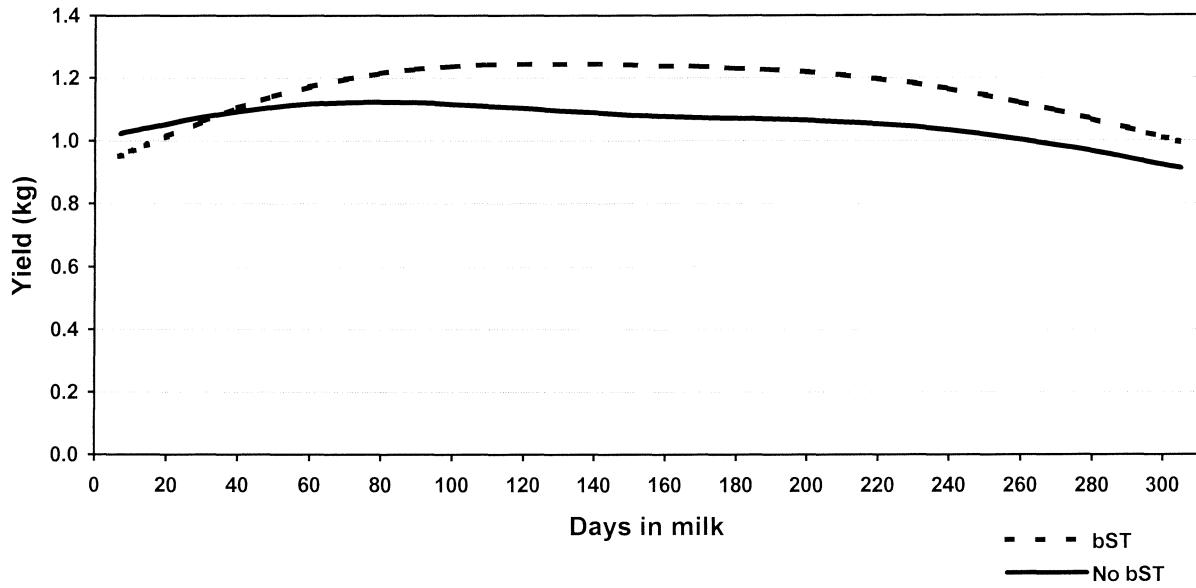**Figure 2.** Fitted mean spline lactation curve for fat yield for treated and untreated bST cows.

Figure 3. Fitted mean spline lactation curve for protein yield for treated and untreated bST cows.