

Bayesian Analysis of Response to Selection: A Case Study Using Litter Size in Danish Yorkshire Pigs

D. Sorensen,* A. Vernersen[†] and S. Andersen[†]

*Danish Institute of Agricultural Sciences, DK-8830 Tjele, Denmark and [†]National Committee for Pig Production, Axelborg, DK-1609 Copenhagen, Denmark

ABSTRACT

Implementation of a Bayesian analysis of a selection experiment is illustrated using litter size [total number of piglets born (TNB)] in Danish Yorkshire pigs. Other traits studied include average litter weight at birth (WTAB) and proportion of piglets born dead (PRBD). Response to selection for TNB was analyzed with a number of models, which differed in their level of hierarchy, in their prior distributions, and in the parametric form of the likelihoods. A model assessment study favored a particular form of an additive genetic model. With this model, the Monte Carlo estimate of the 95% probability interval of response to selection was (0.23; 0.60), with a posterior mean of 0.43 piglets. WTAB showed a correlated response of -7.2 g, with a 95% probability interval equal to (-33.1 ; 18.9). The posterior mean of the genetic correlation between TNB and WTAB was -0.23 with a 95% probability interval equal to (-0.46 ; -0.01). PRBD was studied informally; it increases with larger litters, when litter size is >7 piglets born. A number of methodological issues related to the Bayesian model assessment study are discussed, as well as the genetic consequences of inferring response to selection using additive genetic models.

OVER the last years, pig breeding programs have emphasized selection pressure on fertility traits. While it is recognized that fertility is a composite trait, most of the efforts to improve sow reproductive performance have been placed on selection for litter size. This is partly because litter size is easily recorded and partly because several studies have indicated that it is the most important economic component of sow reproductive performance (BICHARD *et al.* 1983; SMITH *et al.* 1983; TESS *et al.* 1983).

There is a substantial amount of information in the literature about genetic parameters for litter size in pigs; much of this has been reviewed by HALEY *et al.* (1988) and more recently by ROTHSCILD and BIDANEL (1998). ROTHSCILD and BIDANEL (1998) quote an average figure of heritability for total number of piglets born per litter, over 85 published results, of 0.11, with a range from 0.0 to 0.76. Using this average figure, and assuming a value for repeatability over several parities of 0.15, predicted rates of genetic improvement range from ~ 0.3 to 0.5 pigs per litter and generation, depending on the amount of information included to predict breeding values (ÁVALOS and SMITH 1987). These calculations are based on the infinitesimal model and assume infinite population size and in the late 1980s stimulated a reexamination of the possibilities for genetic improvement of litter size. At that time, the only experimental evidence for successful direct selection for increased litter

size was based on results of hyperprolific designs (BICHARD and SEIDEL 1983; TOMES and NIELSEN 1984; LEGAULT 1985). With the exception of results from Nebraska published in the 1990s, which focused on improving litter size by selecting on an index that included ovulation rate and embryo survival (CASEY *et al.* 1994), all other attempts to increase litter size in pigs by direct selection produced unconvincing results (see reviews by BLASCO *et al.* 1995 and ROTHSCILD and BIDANEL 1998).

Against such a background, it was decided in the late 1980s to generate convincing experimental evidence that litter size could be improved by selection. The criterion of selection should be inexpensive and simple to measure and the choice fell on total number of piglets born per litter. A large-scale selection experiment was conducted from 1989 to 1992. The experiment's results have been reported on occasion (BLASCO *et al.* 1995), but a formal investigation of the data, with the most up-to-date methods of inference, has not yet been published.

This article is an attempt to fill in this lack. The emphasis of this work is on the methodological aspects of the analysis. Over the last years, with the introduction of Markov chain Monte Carlo (MCMC) techniques, there has been a revival of Bayesian methods in the statistical literature, and geneticists have been receptive to the inferential possibilities that the Bayesian paradigm offers. This article illustrates the type of information that can be extracted from a selection experiment when MCMC is used in combination with the Bayesian approach.

Corresponding author: Daniel Sorensen, Danish Institute of Agricultural Sciences, Department of Animal Breeding and Genetics, P.O. Box 50, DK-8830 Tjele, Denmark. E-mail: snfds@genetics.agrsci.dk

TABLE 1

Distribution of the number of sows across parities among 261 female parents of animals in *selected* and *control* lines (top three lines) and raw means of total number of piglets born per litter, within parity and line (bottom three lines)

Parity	1	2	3	4	5	6	7	8
Selected line	40	27	34	7	8	2	7	6
Control line	36	45	27	8	6	6	2	—
Mothers to selected line ^a	10.7	12.0	12.1	12.0	12.9	12.7	13.5	13.0
Mothers to control line ^a	9.6	10.4	11.9	11.0	10.1	11.9	10.0	
Base individuals ^b	9.4	10.3	10.9	11.4	11.2	11.6	11.5	11.4

^a The 261 female parents of selected and control lines.

^b The 3534 females with records on 8988 litters, excluding data from selected and control lines.

EXPERIMENTAL DESIGN AND DATA

The data originate from the database of the national breeding program. Litter size (total number of piglets born per litter) had been recorded in the breeding herds, even though it had not been part of the breeding goal. In 1989, the large-scale selection experiment was started using data from the Yorkshire breed only. Information on litter size was gathered from all matings in the breeding herds, and predicted breeding values for each mating were computed using a model that is described below.

The data set used to compute predicted breeding values was produced as follows. On the basis of the matings available during an 8-month period, records on ancestors and collateral relatives were generated and updated each month. This data file consisted of records from 8988 litters and a total of 5796 animals, of which 3534 were females with records. These animals are referred to simply as the *base*. The pedigree of this file included animals born from 1981 until 1988.

During the period of 8 months, from the highest scoring 131 predicted breeding values of the registered matings, 1 gilt was chosen from each resulting litter, and from the highest scoring 36 predicted breeding values, 1 male piglet was chosen from each resulting litter. These 131 gilts and 36 male piglets constituted the *selected line*. A contemporaneous *control line* was created, choosing 130 gilts and 38 male piglets at random from the litters available. Selection took place on a monthly basis, during 8 months, and not just once at the end of the 8-month period. Table 1 shows the contribution of the different parities to the 261 sows from the selected and control lines, as well as the raw means of total number born per litter of the mothers of these 261 sows, within parity and line, and among the 3534 females that belong to the base. The weighted average of the difference across parities of records on females that contributed piglets to the contemporaneous selected and control lines is 1.6 piglets per litter.

Animals from both lines were purchased from the breeding herds and were transferred to a common farm

immediately after weaning at 17–23 days of age. On arrival at the common farm, the piglets underwent a period of quarantine in a nursery.

At the common farm, females of each line were randomly mated (mating took place after the second oestrus) with males of the corresponding line and produced two parities. These are records from *generation 1*. From the first of these two parities, one gilt was chosen from each litter. At sexual maturity, these females were mated twice to randomly chosen males from the appropriate line and produced, therefore, two litters. We refer to these as the records from *generation 2*. There was no directional selection to produce litters in generation 2, but the genotypic value of these females is expected to be a little higher than those from generation 1, because their 36 fathers represent a smaller proportion of the tail of the distribution than the fathers to females from generation 1.

In total, there were 1072 litters in the common farm, approximately equally distributed between the selected and the control lines. The total number of litters in the complete data set was 10,060 (8988 from the base plus 1072 from the selected and the control lines) and the total pedigree file consisted of 6437 individuals (5796 from the base and 641 from the selected and the control lines).

Animals from both lines were kept under identical conditions and it was not possible for the staff in the farm to identify to which line animals belonged. As customarily practiced in the breeding herds, litters in the common farm were also standardized at birth according to total number of piglets born. The traits recorded on the farm were total number of piglets born per litter (TNB), number of piglets born alive per litter (NBA), and average weight of liveborn piglets (WTAB). The latter was recorded within 24 hr after birth; the whole litter was weighed, rather than individual piglets.

Among the 6437 individuals in the pedigree file, 3537 had an inbreeding coefficient >0.0. The average inbreeding coefficient among inbred individuals was 2.4%; the modal value was a little >1%, and the maxi-

mum value was 32%. Only 88 inbreeding coefficients were $>10\%$.

Selection criterion—model for prediction of breeding values: The data available on litter size before the selection experiment was initiated consisted of records on TNB only. The criterion of selection was the average predicted breeding value of the sire and dam of the litter that contributed piglets to the selected and control lines. The model was a repeatability additive genetic model and the method of inference was best linear unbiased prediction (BLUP; *i.e.*, HENDERSON 1973). Thus, the variances of the random effects were assumed known and the values used for these parameters were those traditionally quoted in the literature at the time (HALEY *et al.* 1988): 10% for heritability and 15% for repeatability. The model included, as fixed effects, herd-year-type of insemination (artificial insemination or natural service), season, and parity; as random effects, additive genetic values and permanent environmental effects. The vector of additive genetic values was assumed to follow a multivariate normal distribution, with zero mean and with variance equal to $\mathbf{A}\sigma_a^2$, where \mathbf{A} is the additive genetic relationship matrix among all breeding values in the data, and σ_a^2 is the additive genetic variance. The vectors of permanent environmental effects and residual effects were assumed to follow normal distributions with zero mean, and with variances $\mathbf{I}\sigma_p^2$ and $\mathbf{I}\sigma_e^2$, respectively, where σ_p^2 denotes the variance due to permanent environmental effects and σ_e^2 is the residual variance. These three random vectors were assumed to be mutually independent.

The design used in the experiment was chosen on the basis of previous calculations using the above model. These calculations indicated a predicted deviation between selected and control line of 0.49 piglets per litter. With the experimental design implemented, the approximate probability of detecting such a deviation was estimated as 71%.

STATISTICAL METHODS

The main objective of this experiment has been to confirm that litter size (TNB) responds to selection in the Danish Yorkshire breed. To address this concern, a number of single-trait analyses of TNB were performed using a variety of models. The experiment also generates data to study correlated responses in proportion of piglets born dead (PRBD) and average piglet weight at birth, although it was not specifically designed to address these questions. Correlated responses in WTAB were analyzed using two-trait models only; PRBD was analyzed only informally. Most of the analyses were carried out within the Bayesian framework of inference.

In the analyses presented below, TNB and WTAB are both regarded as traits of the mother of the litter. In the case of WTAB, this assumption was dictated partly by

the availability of litter weights, rather than of individual piglet weights.

The analyses of TNB were done by fitting two types of models that differ fundamentally in the way data contribute to infer response to selection. The first type is a nonhierarchical model, where only data that belong to a generation in question contribute to infer the genetic mean of that generation. The second type is represented by hierarchical additive genetic models, where the complete data contribute information to the genetic mean of a particular generation, leading to sharper inferences. This is at the expense of a higher level of parameterization and stronger assumptions about the form of gene action.

Single-trait analyses of TNB: Nonhierarchical model: The simplest model entertained is one that ignores the family structure in the data. This is labeled the nonhierarchical (NH) model, in contrast with the additive genetic models described below. The NH model considered here assumes the following relationship between the dependent variable and the parameters,

$$y_{ijklm} = l_i + p_j + g_k + s_l + e_{ijklm},$$

$$i = 1, 2; j = 1, 2; k = 1, 2; l = 1, \dots, 4, \quad (1)$$

where y_{ijklm} is a record (TNB) from line i (l_i), parity j (p_j), generation k (g_k), and season l (s_l). Data from the common farm only (selected and control lines) are included in the analysis. Response to selection is defined as

$$R_1 = l_1 - l_2, \quad (2)$$

which is the difference between the effects of the selected and control lines and which is inferred from its marginal posterior distribution. In (2), no distinction is made between responses at generations 1 and 2. A model similar to (1), which included instead line-by-generation interaction effects, did not reveal a difference between the responses at generations 1 and 2, and the results of this analysis are therefore omitted.

Model (1) is written in matrix form as

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}, \quad (3)$$

where \mathbf{b} is the vector that has parameters associated with line, parity, generation, and season, and \mathbf{X} is the known incidence matrix. It is assumed that $\mathbf{y}|\mathbf{b}$ is normally distributed, with mean and variance equal to $\mathbf{X}\mathbf{b}$ and $\mathbf{I}\sigma_e^2$, respectively. *A priori*, the vector $\mathbf{b} = \{b_i\}$ is assigned a 10-dimensional uniform distribution, independent of the residual variance; the latter is assumed to be proportional to $1/\sigma_e^2$. Therefore,

$$p(\mathbf{b}, \sigma_e^2) \propto \frac{1}{\sigma_e^2}, \quad -\infty < b_i < \infty, \text{ for all } i; \quad 0 < \sigma_e^2 < \infty.$$

With the present choice of prior distributions, the marginal posterior distribution of selection response (define selection response as $R = \mathbf{k}'\mathbf{b}$) follows a univariate

t -distribution with mean equal to $E(R|y) = \mathbf{k}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, scale parameter equal to $\hat{\sigma}_e^2 \mathbf{k}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{k}$, and with degrees of freedom equal to $n - \text{rank}(\mathbf{X})$, where

$$\hat{\sigma}_e^2 = (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})'(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})/(n - \text{rank}(\mathbf{X}))$$

and n is the number of elements in \mathbf{y} (BOX and TIAO 1973). The marginal posterior distribution of σ_e^2 is a scaled inverted chi-square distribution with scale parameter $(n - \text{rank}(\mathbf{X})) \hat{\sigma}_e^2$ and $(n - \text{rank}(\mathbf{X}))$ d.f. Note that the mean of $[R|y]$ is numerically identical to the least-squares estimator.

Hierarchical models via Markov chain Monte Carlo: Hierarchical models include a number of additive genetic models that were fitted using the Gibbs sampler. With the additive genetic models, response to selection at generation 1 is defined as the difference in the average additive genetic values between individuals from the selected line, generation 1, and those from the base. This is labeled R_1 . Response to selection at generation 2 is correspondingly defined as the difference in the average additive genetic values between individuals from the selected line, generation 2, and those from the base. This is labeled R_2 .

An alternative expression of response to selection involves deviating the mean additive genetic value of the offspring of selected parents from the average additive genetic values of all matings available during the 8-month period (the group from which parents were chosen). However, since TNB shows no genetic trend over the period studied (not shown) and the mean of the posterior distribution of the average additive genetic values of the base is smaller than 10^{-4} , the above definition was chosen. Note that with the additive genetic model, response to selection is not defined as a deviation involving the control line.

The initial model choice was based partly on the work of ESTANY and SORENSEN (1995). In their analysis of litter size, using a larger data set that included data from 1978 to 1989, it was shown that neither group effects nor maternal effects contributed to variation in litter size in Danish Yorkshire. With this information as a background, three single-trait models were fitted to the complete data set. For the three models, it is assumed that the data vector containing records on TNB (\mathbf{y}) is conditionally normally distributed as

$$\mathbf{y}|\mathbf{b}, \mathbf{a}, \mathbf{p}, \sigma_e^2 \sim N(\mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{p}, \mathbf{I}\sigma_e^2), \quad (4)$$

where \mathbf{b} is a vector that contains location parameters whose *a priori* distributions are assumed to be uniform, with lower and upper bounds equal to \mathbf{b}_{\min} and \mathbf{b}_{\max} , respectively, and \mathbf{a} and \mathbf{p} are vectors that contain additive genetic values and permanent environmental effects, respectively. Matrices \mathbf{X} , \mathbf{Z} , and \mathbf{W} are known incidence matrices, and σ_e^2 is the residual variance. *A priori*, it is assumed that additive genetic values and permanent environmental effects are multivariate normally distributed:

$$\mathbf{a}|\mathbf{A}, \sigma_a^2 \sim N(\mathbf{0}, \mathbf{A}\sigma_a^2) \quad (5)$$

$$\mathbf{p}|\sigma_p^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_p^2). \quad (6)$$

In these expressions, σ_a^2 is the additive genetic variance in the population from which individuals with unknown parents were conceptually sampled, and σ_p^2 is the permanent environmental variance. The matrix \mathbf{A} has, as elements, additive genetic relationships among the genetic values.

The models differed in their prior distributions and in the specification of the herd-year-type of insemination effects. In model 1, vector \mathbf{b} has effects of herd-year-type of insemination, season, and parity. *A priori*, variance components are assumed to be independently and uniformly distributed, taking values in the following ranges: $0 < \sigma_a^2 \leq 4$, $0 < \sigma_p^2 \leq 2$, and $0 < \sigma_e^2 \leq 20$. These distributions for the variance components imply an *a priori* distribution of heritability and repeatability with respective modal values of 0.15 and 0.23 (SOERSEN 1996).

Model 2 differs from the above in that the vector of herd-year-type of insemination effects, \mathbf{h} , is assumed *a priori*, to follow a multivariate normal distribution:

$$\mathbf{h}|\sigma_h^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_h^2). \quad (7)$$

The vector \mathbf{b} now contains effects of parity, season, and type of insemination (artificial *vs.* natural service). All variance components (σ_a^2 , σ_h^2 , σ_p^2 , σ_e^2) are assumed to be *a priori*, independently and uniformly distributed, taking values in the ranges described above. For σ_h^2 , the range was $0 < \sigma_h^2 \leq 2$.

Finally, response was also evaluated using a model similar to model 1, except that heritability and repeatability were assumed to take values 0.10 and 0.15, respectively, and the posterior distribution of response was obtained, conditional on these values. This is denoted model 3 and is the same model used to carry out selection decisions and is numerically equivalent to a BLUP approach if heritability and repeatability are the true values. In a frequentist setting, the sampling variance of the "BLUP estimate of selection response" must be computed over the distribution of the selected data; this is not an analytically tractable task. The Bayesian analysis implemented here via the Gibbs sampler yields a measure of uncertainty associated with selection response in the form of a Monte Carlo estimate of the posterior variance.

Two other models were also fitted: one of these differed from model 1 in the *a priori* specification of the variance components. These were assumed to follow scaled inverted chi-square distributions, reflecting little information about the variance components. The other model differed from model 2 in the same way. The first of these two behaved almost identically to model 1, and the second one behaved almost identically to model 2. Therefore, results obtained from these two models are omitted.

The bounds \mathbf{b}_{\min} and \mathbf{b}_{\max} were set equal to 0 and 20, respectively, for all elements in either \mathbf{b}_{\min} or \mathbf{b}_{\max} for all models. Runs with different bounds yielded almost identical results and these are therefore not presented.

Two-stage regression of TNB on parental predicted additive genetic values: The data were analyzed also using a two-stage approach, designed to study whether TNB achieved after selection was consistent with predictions based on the additive genetic model 1 (ANDERSEN *et al.* 1998). In a first stage, data from the base only (excluding data from selection and control lines in the common farm) were used to obtain estimates of genetic variances and additive genetic values. In a second stage, using only the data from selection and control lines, generation 1, the following model was fitted to TNB,

$$y_{ijkl} = t_i + s_j + \beta \text{ped}_k + \varphi_{f,k} + \varphi_{m,k} + p_k + e_{ijkl} \quad (8)$$

where t_i is the effect of the i th parity, s_j is the effect of the j th season, $\text{ped}_k = (\hat{a}_{f,k} + \hat{a}_{m,k})/2$, the average predicted additive genetic values of the father and mother of sow k based on the first stage analysis, $\varphi_{f,k} = (a_{f,k} - \hat{a}_{f,k})/2$ and $\varphi_{m,k} = (a_{m,k} - \hat{a}_{m,k})/2$ are random errors of prediction associated with the father (f) and mother (m) of k , p_k is the sum of the effects of Mendelian sampling and permanent environmental effects, and e_{ijkl} is a residual effect peculiar to the l th record. Note that no distinction is made between records from selection or control lines. If TNB does not respond to selection, it is expected that $\beta = 0$; under the assumption that the model holds, $\beta = 1$. This analysis was carried out using a frequentist approach.

Single-trait analysis of PRBD: Data on PRBD were markedly skewed and it was not possible to find a transformation to remove this skewness. A formal analysis of the correlated response in PRBD, which requires inclusion of the trait selected for TNB, will be presented in the future. At this stage, we report results from a univariate analysis based on the approach in KERR and CAMERON (1995). TNB is considered as fixed, and NBA is modeled using a generalized linear mixed model with binomial errors and a probit link function of the binomial probability PRBD.

It was cursorily observed that PRBD increases when TNB is >7 . Therefore, the following phenotypic relationship was studied between PRBD and TNB,

$$\text{Probit}(\text{PRBD}_{ijklmnop}) = l_i + pg_j + s_k + t_i + \beta \text{TNB}_p I(\text{TNB}_p > 7) + S_n + D_o + A_p, \quad (9)$$

where l_i is the effect of line i ($i = 1, 2$); pg_j is the j th interaction effect of parity by generation; s_k is the effect of season k ; t_i is a classification variable that takes the value TNB if $\text{TNB} \leq 7$ and zero otherwise; $I(\text{TNB}_p > 7)$ is the indicator function that takes the value 1 if the argument is satisfied and zero otherwise; β is a regression coefficient; S_n is the n th sire effect assumed normally distributed, with mean zero and variance σ_s^2 ; D_o is

the o th dam effect, assumed normally distributed, with mean zero and variance σ_d^2 ; and A_p is the p th sow effect, assumed normally distributed, with mean zero and variance σ_a^2 . A similar model with a regression coefficient for each of the two lines was also fitted to the data, but no difference could be detected between the two regression coefficients.

Two-trait analysis of TNB and WTAB: As with the single-trait analyses, a number of two-trait models were investigated. The models differed in the form of the prior distributions and in the covariance structure of location parameters. Inferences about response to selection for TNB and correlated responses in WTAB were almost identical in all cases; we present, therefore, results from the most parsimonious among the models studied.

Let (\mathbf{y}, \mathbf{z}) denote the vector of length $10,060 \times 2$ representing the $n = 10,060$ records on TNB (vector \mathbf{y}), and the augmented WTAB data (\mathbf{z}). The latter was augmented with $10,060 - 1072 = 8988$ missing residuals. It is assumed that the i th record (y_i, z_i) is conditionally, bivariate normally distributed, given the parameters

$$\begin{pmatrix} y_i \\ z_i \end{pmatrix} \Big| \mathbf{b}_y, \mathbf{b}_z, \mathbf{p}_y, \mathbf{p}_z, \mathbf{a}_y, \mathbf{a}_z, \mathbf{R}_{e_i} \sim N \left(\begin{pmatrix} \mathbf{x}'_y \mathbf{b}_y + \mathbf{w}'_y \mathbf{p}_y + \mathbf{z}'_y \mathbf{a}_y \\ \mathbf{x}'_z \mathbf{b}_z + \mathbf{w}'_z \mathbf{p}_z + \mathbf{z}'_z \mathbf{a}_z \end{pmatrix}, \mathbf{R}_{e_i} \right), \quad (10)$$

$i = 1, \dots, n,$

where \mathbf{R}_{e_i} is the i th residual covariance matrix, which has the structure

$$\mathbf{R}_{e_i} = \eta_i \mathbf{R}_e = \begin{bmatrix} 1 & 0 \\ 0 & 1/n_i \end{bmatrix} \begin{bmatrix} \sigma_{e_y}^2 & \sigma_{e_{yz}} \\ \sigma_{e_{yz}} & \sigma_{e_z}^2 \end{bmatrix}. \quad (11)$$

In (11), n_i is the known number of records contributing to WTAB. Elements of the row vectors \mathbf{x}'_i , \mathbf{w}'_i , and \mathbf{z}'_i (subscripts y and z refer to TNB and to WTAB, respectively), associated with a missing residual, contain only zeroes. The vector \mathbf{b} has location parameters that are assumed to be uniformly distributed, *a priori*. For both TNB and for WTAB, these are herd-year-type of insemination effects, season, and parity. The vectors \mathbf{p}_y and \mathbf{p}_z represent permanent environmental effects and the vectors \mathbf{a}_y and \mathbf{a}_z represent additive genetic values. Additive genetic values and permanent environmental effects are assumed to be, *a priori*, multivariate normally distributed:

$$\begin{pmatrix} \mathbf{a}_y \\ \mathbf{a}_z \end{pmatrix} \Big| \mathbf{G}_o, \mathbf{A} \sim N \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, [\mathbf{G}_o \otimes \mathbf{A}] \right)$$

$$\begin{pmatrix} \mathbf{p}_y \\ \mathbf{p}_z \end{pmatrix} \Big| \sigma_{p_y}^2, \sigma_{p_z}^2 \sim N \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{I}_m \sigma_{p_y}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \sigma_{p_z}^2 \end{bmatrix} \right).$$

In these expressions, \mathbf{G}_o is the 2 by 2 matrix of additive genetic variances and covariances, and $\sigma_{p_y}^2$ and $\sigma_{p_z}^2$ are the variances due to permanent environmental effects associated with TNB and WTAB, respectively.

All variances and covariance matrices were assigned

improper uniform prior distributions, *a priori*. This is denoted *model 4*. In *model 5*, for TNB, herd-year-type of insemination effects were assumed to have, *a priori*, independent normal distributions, with zero means. Independent scaled inverted chi-square distributions were assigned to $\sigma_{p_y}^2$ and to $\sigma_{p_z}^2$ (and to $\sigma_{h_y}^2$), and scaled inverse Wishart distributions to \mathbf{G}_o and \mathbf{R}_c . The parameter “degrees of freedom” associated with all these distributions was set equal to 5. This is intended to reflect weak *a priori* information. Variants of the above models with respect to prior specifications were also fitted, with hardly any consequences on inferences about response and correlated response to selection. Therefore, results based on only models 4 and 5 are shown.

The first step in the implementation of the Gibbs sampler was the generation of the WTAB missing residuals from the fully conditional posterior distribution. From (10), this has the form

$$z_{m,i}|y_{o,i}, \mathbf{b}, \mathbf{p}, \mathbf{a}, \mathbf{R}_c \sim N(E(z_{m,i}|y_{o,i}, \mathbf{b}, \mathbf{p}, \mathbf{a}, \mathbf{R}_c), \text{Var}(z_{m,i}|y_{o,i}, \mathbf{b}, \mathbf{p}, \mathbf{a}, \mathbf{R}_c)),$$

where

$$E(z_{m,i}|y_{o,i}, \mathbf{b}, \mathbf{p}, \mathbf{a}, \mathbf{R}_c) = \frac{\sigma_{e,yz}}{\sigma_{e,y}^2} (y_{o,i} - \mathbf{x}'_{y_{o,i}} \mathbf{b}_y - \mathbf{w}'_{y_{o,i}} \mathbf{p}_y - \mathbf{z}'_{y_{o,i}} \mathbf{a}_y)$$

and

$$\text{Var}(z_{m,i}|y_{o,i}, \mathbf{b}, \mathbf{p}, \mathbf{a}, \mathbf{R}_c) = \sigma_{e,z}^2 (1 - r_{e,yz}^2).$$

In these expressions, $z_{m,i}$ represents the i th missing residual, $y_{o,i}$ is the i th TNB observation, $\mathbf{x}'_{y_{o,i}}$, $\mathbf{w}'_{y_{o,i}}$, and $\mathbf{z}'_{y_{o,i}}$ are the rows of \mathbf{X}_y , \mathbf{W}_y , and \mathbf{Z}_y associated with $y_{o,i}$, $\sigma_{e,yz}$ is the residual covariance between TNB and WTAB, $\sigma_{e,z}^2$ ($\sigma_{e,y}^2$) is the residual variance associated with WTAB (TNB), and $r_{e,yz} = \sigma_{e,yz} / \sqrt{\sigma_{e,z}^2 \sigma_{e,y}^2}$ is the residual correlation between traits. After generating the augmented data, location parameters were sampled from the appropriate multivariate normal distribution, variances were sampled from the appropriate fully conditional inverse chi-square posterior distributions, and covariance matrices from the appropriate scaled inverse Wishart distributions. Each iteration required generation of a new set of missing residuals and the construction of the mixed-model equations. Further details of the algorithm can be found, for example, in SORENSEN (1996).

Gibbs sampling implementation: The results presented below are based on Monte Carlo estimates of means of posterior distributions. A large number of experiments for each model were run with the Gibbs sampler. The burn-in periods and number of iterations of the various Gibbs chains were chosen partly pragmatically, looking at the results of various independent chains as time-series plots overlaid and seeing if these could be distinguished, and partly using more formal methods. These include the approaches of GELMAN and RUBIN (1992) and of RAFTERY and LEWIS (1993). Monte Carlo sampling errors of features of posterior distributions were also computed, using an estimator described

in GEYER (1992). This is used as a criterion to decide whether differences among estimates of posterior means from different chains can be attributed to sampling variation. Plots of histograms of samples from posterior distributions were also compared among runs, to confirm consistency of results. When the total length of the chain approached 100,000 and the burn-in period was 10,000, runs from different chains within models yielded practically identical results. We present numbers based on final runs using chains of length 500,000 with a burn-in period of 50,000. These numbers exceed, by a factor of 100, those suggested using the criterion of RAFTERY and LEWIS (1993). The value of the statistic proposed by GELMAN and RUBIN (1992) was never >1.18 (when this statistic is equal to 1.0, the chains are essentially overlapping). Among the estimates shown, the effective chain size (SORENSEN *et al.* 1995) for inferences about selection response was never <1900 . The estimates of the Monte Carlo standard error for the different expressions of selection response were of the order of 0.0025. The corresponding estimates for heritability, repeatability, and proportion of variance due to herd-year-type of insemination effects are <0.0004 .

It was estimated that the large proportion of missing data could cause very slow mixing. Therefore, after augmenting with the missing residuals, the sampler was implemented using a multiple-trait extension of the approach suggested by GARCIA-CORTES and SORENSEN (1996); rather than sampling each location parameter scalarly, all the location parameters are sampled jointly, in one pass.

Evaluating contributions to statistical information about response to selection: It is of interest to quantify the relative amount of information contributed by subsets of the data to the posterior distribution of selection response. To achieve this, response to selection is studied using two types of analyses. In the first one, phenotypic records from the selected and control lines (labeled vector \mathbf{y}_2) are excluded; only data from the base (labeled \mathbf{y}_1) contribute to inferences about response to selection. This is denoted the *predictive analysis*. Since the additive genetic models assume that parental additive genetic values combine additively to generate the additive genetic values in their offspring, and since the pedigree is known, selection response can be inferred using \mathbf{y}_1 only. In a second analysis, the complete data, $\mathbf{y}' = (\mathbf{y}'_1, \mathbf{y}'_2)$, are included to draw inferences; this is the *complete data analysis*.

Let θ represent the vector of parameters for a given model and let $p(\theta)$ denote the prior distribution of θ . Then the posterior distribution of θ , conditional on the complete data, is given by

$$\begin{aligned} p(\theta|\mathbf{y}_1, \mathbf{y}_2) &\propto p(\theta)p(\mathbf{y}_1, \mathbf{y}_2|\theta) \\ &= p(\theta)p(\mathbf{y}_1|\theta)p(\mathbf{y}_2|\theta), \end{aligned} \quad (12)$$

where the equality arises because in the cases of the

models considered here, \mathbf{y}_1 and \mathbf{y}_2 are conditionally independent, given θ . The predictive analysis is associated with the first two terms in the right-hand side of (12). Taking logarithms, and differentiating twice with respect to θ , yields

$$I(\theta|\mathbf{y}_1, \mathbf{y}_2) = I(\theta) + I(\mathbf{y}_1|\theta) + I(\mathbf{y}_2|\theta). \quad (13)$$

Equation 13 shows that the amount of information in the posterior distribution is equal to the sum of information contributed by the prior distribution, $I(\theta)$, by the sampling model of the base, $I(\mathbf{y}_1|\theta)$, and by the sampling model of data from selected and control lines only, $I(\mathbf{y}_2|\theta)$. Information measures are approximated by the inverse of the respective variances, and in this way it can be quantified how different sources have contributed information to posterior inferences about selection response. This approximation to the concept of information is based on asymptotic results. These analyses are implemented with the hierarchical models only, and are in similar spirit to the two-stage regression study described above.

Model assessment: Inferences about response to selection vary among the different models fitted. The problem of model comparison is addressed first, by comparing marginal likelihoods from different models (NEWTON and RAFTERY 1994), and second, by studying discrepancy measures involving the comparison between observed data and data generated from predictive distributions derived under the different models (GELFAND 1996). The former can be regarded as emphasizing the fidelity of the model to the observed data, whereas the latter places the emphasis on the predictive power of the model.

There is a voluminous literature on model determination, both from frequentist and Bayesian perspectives. A useful account of both schools of inference can be found in GEISSER (1993), while GELMAN *et al.* (1995) and CARLIN and LOUIS (1996) place emphasis on Bayesian approaches. In this section, results based on MCMC estimators of two related Bayesian quantities are presented, which are labeled $p(\mathbf{y}|M_i)$ and $p(\mathbf{z}_2|\mathbf{z}_1, M_i)$, where \mathbf{y} is the vector of the complete data, and vector \mathbf{z}_i is defined below.

The first of these quantities, $p(\mathbf{y}|M_i)$, is the marginal density of the data, or marginal likelihood, and is the basis for the computation of Bayes factors (*i.e.*, O'HAGAN 1994). These are the ratios of the marginal probabilities of the data under models M_i and M_j , (with associated parameters θ_i and θ_j):

$$B_{ij} = \frac{p(\mathbf{y}|M_i)}{p(\mathbf{y}|M_j)} = \frac{\int p(\mathbf{y}|\theta_i, M_i) p(\theta_i|M_i) d\theta_i}{\int p(\mathbf{y}|\theta_j, M_j) p(\theta_j|M_j) d\theta_j}. \quad (14)$$

Here an MCMC estimator of (14) is used, proposed by NEWTON and RAFTERY (1994). It is based on the ratios of the Monte Carlo estimates of marginal likelihoods; these have the form

$$\hat{p}(\mathbf{y}|M_i) = \left[\frac{1}{m} \sum_{j=1}^m p^{-1}(\mathbf{y}|\theta_i^{(j)}, M_i) \right]^{-1}. \quad (15)$$

In (15), $\theta_i^{(j)}$ represents the j th Gibbs sample from $p(\theta_i|\mathbf{y}, M_i)$ and m is the total number of samples drawn. The marginal likelihood can be interpreted as the probability of obtaining the data actually observed under model i , calculated before any data became available (KASS and RAFTERY 1995). It is thus a global measure of the model's ability to explain the observed data.

We are also interested in studying how the models predict specific features of the data. Since the scientific purpose of the present experiment was to generate a difference between the selected and the control lines, using the data from the base, the models were tested by comparing discrepancy measures involving observed data and simulated data \mathbf{z}_2 from the predictive density $p(\mathbf{z}_2|\mathbf{z}_1, M_i)$. The vector \mathbf{z}_1 contains data from the base plus parity 1 records from the selected and the control lines. The latter were added to the base data to guarantee that all the parameters needed to sample from $p(\mathbf{z}_2|\mathbf{z}_1, M_i)$ had been inferred from $p(\theta_i|\mathbf{z}_1, M_i)$. The vector \mathbf{z}_2 is a simulated value of all the parity 2 records of individuals in the selected and the control lines.

Simulation of data from $p(\mathbf{z}_2|\mathbf{z}_1, M_i)$ can be easily accomplished and is based on the identity

$$p(\mathbf{z}_2|\mathbf{z}_1, M_i) = \int p(\mathbf{z}_2|\theta, M_i) p(\theta|\mathbf{z}_1, M_i) d\theta. \quad (16)$$

Using the Gibbs sampler, draws $\theta_i^{(j)}$ ($j = 1, \dots, m$) were obtained from $p(\theta_i|\mathbf{z}_1, M_i)$ and draws $\mathbf{z}_2^{(j)}$ were obtained from $p(\mathbf{z}_2|\theta_i^{(j)}, M_i)$. The $\mathbf{z}_2^{(j)}$'s are approximate draws from $p(\mathbf{z}_2|\mathbf{z}_1, M_i)$.

Having obtained draws from $p(\mathbf{z}_2|\mathbf{z}_1, M_i)$, the posterior distributions of a number of discrepancy measures were chosen. The first one was the average difference between the simulated data $\mathbf{z}_2^{(j)}$ and the observed data \mathbf{z}_2 . This discrepancy measure is D_T ; the Monte Carlo estimator of D_T is

$$\hat{D}_T = \frac{1}{m} \sum_{j=1}^m \frac{1}{n_2} \mathbf{1}' [\mathbf{z}_2^{(j)} - \mathbf{z}_2], \quad (17)$$

where m is the length of the Gibbs chain, $\mathbf{1}$ is a vector of ones (of length n_2), n_2 is the number of elements in \mathbf{z}_2 , and $\mathbf{z}_2^{(j)}$ is the j th draw from the predictive density $p(\mathbf{z}_2|\mathbf{z}_1, M_i)$, ($j = 1, \dots, m$).

The second discrepancy measure chosen was similar to (17), except that only the most extreme observations from \mathbf{z}_2 were kept (4 or less piglets born per litter and 16 or more piglets born per litter). We are interested in comparing the ability of the different models to predict extreme records. The Monte Carlo estimate of this discrepancy is labeled \hat{D}_{EXTH} for the extreme high and \hat{D}_{EXTL} for the extreme low observations.

The third and final discrepancy measure chosen involves the difference between the raw averages of parity 2 records from the selected and control lines. The ob-

TABLE 2
Predictive analysis

	\hat{R}_1	\hat{R}_2	h_2	r	hy^2
Model 1	0.58(0.15)	0.69(0.17)	0.16(0.019)	0.24(0.011)	
Model 2	0.52(0.12)	0.60(0.16)	0.15(0.018)	0.22(0.010)	0.07(0.008)
Model 3	0.49(0.10)	0.58(0.13)			

Estimates of posterior means of response to selection at generations 1 (\hat{R}_1) and 2 (\hat{R}_2) and estimates of posterior means of heritability (h^2), repeatability (r), and variance due to herd-year-type of insemination as proportion of total variance (hy^2), fitting three models (see text for explanation). Posterior standard deviations are in parentheses.

served average difference is $D_{SC} = \bar{z}_{s,2} - \bar{z}_{c,2} = 0.40$ piglets per litter, where $\bar{z}_{s,2}$ is the observed raw average of the observed parity 2 records from the selected line, and $\bar{z}_{c,2}$ is the observed raw average of the parity 2 records from the control line. From the simulated data $\mathbf{z}_2^{(j)}$ generated from the predictive density (16), values for $\bar{z}_{s,2}$ and $\bar{z}_{c,2}$ were computed and the Monte Carlo estimator of the difference between the averages of parity 2 records from the selected and control lines is labeled \hat{D}_{SC} .

Monte Carlo estimates of the average squared differences of the above discrepancies were also computed. The results do not contribute further insight and are therefore not presented.

RESULTS

Single-trait analysis based on the nonhierarchical model: The posterior mean of response to selection for TNB computed using the NH model is 0.40 piglets per litter. The scale parameter $\hat{\sigma}_k^2 \mathbf{X}'\mathbf{X}^{-1} \mathbf{k}$ of the t -distribution is equal to 0.0609 and the 95% probability interval of the response to selection is $(-0.07, 0.89)$. Thus, 91% of the posterior distribution includes positive values of response to selection. This model understates uncertainty because no account is taken of the correlated structure in the data.

Single-trait analyses based on hierarchical models (additive genetic models): Results from the predictive analyses are shown first. Estimates of posterior means of response to selection, heritability, repeatability, and proportion of variance due to herd-year-type of insemination effects, using models 1, 2, and 3, are shown in Table 2.

Differences among the predictions from model 1 and the other two models vary substantially. All models predict a higher response at generation 2, due to the fact that generation 2 females are offspring of the 36 highly selected sires.

Estimates of posterior means and standard deviations of response to selection for the three models, based on the complete data analyses, are shown in Table 3. Comparison of results in Tables 2 and 3 shows that the complete data analyses produce less differences among models than the predictive analyses in the pattern of

selection response. The posterior mean based on the complete data analyses for models 2 and 3 is smaller than for model 1 and is very similar to the one from the NH model. Inferences about response to selection on the basis of the hierarchical models are sharper than those on the basis of the NH model: the posterior variance is approximately six times smaller.

The estimates of posterior means in Table 3 produce a consistent picture of the pattern of response to selection across the three models. A comparison between model 1 and model 2 reveals that a smaller estimate is obtained when it is assumed that herd-year-type of insemination effects are *a priori* normally distributed (model 2). It is interesting to note this difference, especially since herd-year-type of insemination effects account for only 6% of the total variance in TNB, as we show below. Inferences under models 2 and 3 are similar, despite the fact that the posterior mean of heritability retrieved from model 2 (0.15) is 50% larger than the value assumed in the conditional model 3 (0.10).

All posterior distributions of response are symmetric (not shown). Estimates of the median and of the mode of the distributions agree (up to two decimal places) with estimates of the mean.

Shown in the last three columns of Tables 2 and 3 are estimates of posterior means of heritability, repeatability, and of proportion of the variance due to herd-year-type of insemination effects. The numbers are very similar in both tables because a vast proportion of the information about these genetic parameters is contributed by the base. The values obtained for heritability and repeatability are a little higher for the model where it is assumed that herd-year-type of insemination effects are, *a priori*, uniformly distributed (model 1). The difference is due to a larger estimate of the posterior mean of the additive genetic variance (1.46 *vs.* 1.34) and to a smaller estimate of the posterior mean of the phenotypic variance (8.81 *vs.* 9.28). The figures of 0.17 for heritability and 0.24 for repeatability retrieved from model 1 are a little larger than the usual average value of 0.10 reported in the literature. Posterior distributions were highly symmetric (not shown; posterior means, modes, and medians agreed to two decimal places).

Quantifying statistical information about response to

selection from different sources: The breakdown of the relative contributions to information about selection response based on Equation 13, expressed as the percentage contributions of the base plus prior, and the data from selected and control lines, relative to the information in the posterior distribution, yield 69.4 and 30.6%, respectively. These numbers are very similar for all models, except for model 3, which assumes variances known. Here, the relative contributions are 81.0 and 19.0%. Clearly, for the hierarchical models, there is a very substantial part of the statistical information arising from the base and the prior distribution. However, within the set of prior distributions studied, the amount of statistical information contributed by these two sources is very similar. Notwithstanding, because this amount is very substantial, it is important to compare inferences drawn from the hierarchical models and from the less-parameterized NH model.

For heritability, the prior distribution contributes with 2.7% of the total information contained in the posterior distribution. The (complete) data are overwhelmingly informative relative to the prior.

Two-stage regression analysis of TNB on parental predicted additive genetic values: The estimate of the regression coefficient in model 8 was 0.76, with a standard error of 0.26. This leads to a rejection of the hypothesis of no effect of pedigree index on TNB, with a *P* value of 0.2%. On the other hand, the analysis fails to reject the hypothesis that $\beta = 1$. In other words, even though the estimated response is 76% of the response predicted by the model, the analysis does not generate statistical evidence to question the predictive ability of the model.

This is based on the analysis fitting model 1 to data set y_1 . Fitting models that treat herd-year-type of insemination as normally distributed yields estimates of the regression coefficient closer to 1. However, the conclusion is the same as above.

Analysis of PRBD: On the probit scale, the estimated line difference (selected – control) was 0.132 with a standard error of 0.049, and the estimate of β was 0.017 with a standard error of 0.007. On the probability scale, the line difference is 2.1% if the level of PRBD in the control is 8% and 2.5% if the level of PRBD in the control is 10%. If the level of PRBD is 8%, an increase

of one piglet born increases PRBD to 8.25%, and if the level of PRBD is 10%, an increase of one piglet born increases PRBD to 10.3%.

Two-trait analysis of TNB and WTAB: Only complete data analyses are discussed. Shown in Table 4 are estimates of means of posterior distributions of heritabilities, repeatabilities, and genetic correlations from the bivariate analyses. The parameters associated with TNB do not differ from the single-trait estimates. Both models produce almost identical results. The posterior standard deviation of the genetic correlation reflects a high degree of uncertainty. However, the posterior distribution assigns most of the probability mass to negative values of the genetic correlation; the 95% probability interval is (–0.46, –0.01).

As was the case with the single-trait analysis, the posterior mean of heritability is a little smaller for the model that assumes that herd-year-type of insemination effects for TNB are, *a priori*, normally distributed, and the posterior mean of the proportion of variance due to herd-year-type of insemination effects is ~6%. A comparison of the posterior standard deviations in Table 5 with those of Table 4 reveals that the amount of information contributed by WTAB to inferences about TNB is undetectable.

Estimates of posterior means of selection response, using models 4 and 5, are shown in Table 5.

Estimates of response for TNB are almost identical to the results obtained from the analogous models in the single-trait analysis. As before, model 5, which assumes that for TNB, herd-year-type of insemination effects are, *a priori*, normally distributed, leads to slightly smaller estimates of posterior means.

Estimates of posterior means of correlated response in WTAB are negative. Posterior standard deviations are relatively large, indicating a considerable degree of uncertainty. Thus, for model 4, the Monte Carlo estimate of the 95% probability intervals for WTAB, associated with \hat{R}_1 and with \hat{R}_2 , are (–33.3, 18.1) and (–44.8, 14.7), respectively. The probability that the correlated response (based on \hat{R}_1) is negative is 76%. The posterior distributions associated with TNB are sharper; the corresponding intervals for TNB are (0.24, 0.63) and (0.30, 0.77).

TABLE 3
Complete data analysis

	\hat{R}_1	\hat{R}_2	h^2	r	hy^2
Model 1	0.48(0.12)	0.60(0.14)	0.17(0.017)	0.24(0.010)	
Model 2	0.43(0.10)	0.52(0.12)	0.15(0.018)	0.23(0.012)	0.07(0.007)
Model 3	0.43(0.09)	0.53(0.11)	—	—	—

Estimates of posterior means of response to selection at generations 1 (\hat{R}_1) and 2 (\hat{R}_2) and estimates of posterior means of heritability (h^2), repeatability (r), and variance due to herd-year-type of insemination as proportion of total variance (hy^2), fitting three models (see text for explanation). Posterior standard deviations are in parentheses.

TABLE 4

Estimates of posterior means and standard deviations (SD), obtained from models 4 and 5, for heritability (h^2), repeatability (r), environmental correlation (r_E), genetic correlation (r_A), and proportion of variance due to herd-year-type of insemination effects (hy^2)

	h_y^2	h_z^2	r_y	r_z	r_A	r_E	hy_y^2
Model 4	0.17	0.25	0.24	0.29	-0.23	-0.39	
SD	0.016	0.038	0.010	0.034	0.112	0.032	
Model 5	0.15	0.26	0.22	0.29	-0.21	-0.37	0.06
SD	0.016	0.040	0.011	0.037	0.108	0.034	0.007

Subscripts y and z refer to TNB and to WTAB, respectively.

Model assessment: Results from the model assessment study are shown in Table 6. Models 1 and 4 [those that considered that herd-year-type of insemination effects were, *a priori*, uniformly distributed (HYT-Un) in Table 6] produced almost indistinguishable results. So did models 2 and 5 [those that considered that herd-year-type of insemination effects were, *a priori*, normally distributed (HYT-Nor)]. Therefore, to economize on the presentation of results, these two sets of models were grouped. For TNB, the fact that there is no difference in the behavior of single-trait and two-trait models confirms that WTAB does not contribute detectable information to inferences about TNB.

The first two rows of Table 6 compare the two groups of hierarchical models. The difference between the natural logarithms of marginal likelihood (the logarithm of the Bayes factor) provides *very strong* evidence (using the standard in KASS and RAFTERY 1995) in favor of models that assume herd-year-type of insemination effects are, *a priori*, normally distributed, relative to the alternative HYT-Un group of models. In fact, if equal probabilities are assigned for the two models *a priori*, the results obtained indicate that model HYT-Nor is $\exp(30.8)$ times more likely *a posteriori* than model HYT-Un. A similar conclusion favoring the HYT-Nor model based on a likelihood analysis, a different data set, and a different criterion of model assessment was arrived at by ESTANY and SORENSEN (1995).

Despite the rather marked difference in the marginal likelihoods, the difference between the two groups of models to predict specific features of the data via the

discrepancy measures chosen is small. However, over all the criteria explored, the HYT-Nor group of models performed better than the HYT-Un group of models.

Conditionally on \mathbf{z}_1 , both groups of hierarchical models tend to overpredict the difference between the means of parity 2 records from selected and control lines (the observed difference is 0.40 piglets per litter), with HYT-Nor behaving slightly better than HYT-Un. However, this observed difference is well within the range of posterior uncertainty: the Monte Carlo estimate of the 95% posterior interval of the marginal posterior distribution of D_{SC} under HYT-Nor is (0.14, 1.12) and under HYT-Un is (0.20, 1.15).

The estimates of marginal likelihoods provide very strong evidence against the NH model. This is consistent with the observation that the posterior distribution of heritability assigns most of the probability mass away from zero. The NH model predicts the average observed difference of parity 2 records between selected and control lines very accurately. However, the posterior uncertainty of this predictive distribution is much larger than that obtained fitting the hierarchical models.

All the models implemented in this study underpredict extreme observations (last two columns of Table 6). A closer look at the extreme observations did not produce insight as the underlying cause for this result; errors in recording cannot be dismissed. The NH model performs worse than the hierarchical models; it performs poorly also, relative to the hierarchical models, in predicting individual records, on average.

A little digression is in place to describe the computa-

TABLE 5

Estimates of posterior means of response to selection for TNB at generations 1 (\hat{R}_1) and 2 (\hat{R}_2) and of posterior means of the correlated responses for WTAB (in grams) at generations 1 (\widehat{CR}_1) and 2 (\widehat{CR}_2), fitting Models 4 and 5

	TNB		WTAB	
	\hat{R}_1	\hat{R}_2	\widehat{CR}_1	\widehat{CR}_2
Model 4	0.48(0.12)	0.60(0.14)	-8.4(13.8)	-17.4(15.6)
Model 5	0.43(0.10)	0.52(0.12)	-7.2(13.3)	-15.0(15.03)

Standard deviations of posterior distributions are in parentheses.

TABLE 6
Model comparison via marginal likelihoods and
discrepancy measures

Model	$\ln \hat{p}(\mathbf{y} M)^f$	\hat{D}_{SC}	\hat{D}_T	\hat{D}_{EXTH}	\hat{D}_{EXTL}
HYT-Un ^a	-22914.1	0.69	0.68	-2.95	3.61
HYT-Nor ^b	-22883.3	0.65	0.64	-2.54	3.19
SD ^c		0.28	0.19	0.92	0.90
NH ^d	-24025.7	0.39	0.72	-7.01	8.97
SD ^e		0.37	0.22	0.96	0.95

Monte Carlo estimates of log marginal likelihood, $\ln \hat{p}(\mathbf{y}|M)$ (computed using Equation 15), and Monte Carlo estimates of posterior means of discrepancy measures (\hat{D}_{SC} , difference between average of parity of two simulated records in selected and control lines; the observed difference is 0.40 piglets per litter); \hat{D}_T , average difference between observed and simulated data; \hat{D}_{EXTH} , average difference between highest extreme observed and simulated records; \hat{D}_{EXTL} , average difference between lowest extreme observed and simulated records.

^a Models 1 and 4 (herd-year-type of insemination uniformly distributed, *a priori*).

^b Models 2 and 5 (herd-year-type of insemination normally distributed, *a priori*).

^d Nonhierarchical NH model, slightly modified for the computation of $\hat{p}(\mathbf{y}|M)$ —see text for explanation.

^e Approximate standard deviation of posterior distribution associated with the hierarchical/NH models.

^f Empirical Monte Carlo standard error of $\ln \hat{p}(\mathbf{y}|M)$: 6.2.

tion of the marginal likelihood associated with the NH model. Expression (14) shows that the Bayes factor is undetermined when $p(\theta_i|M_i)$ is improper, as defined in the NH model. To circumvent this problem, $\hat{p}(\mathbf{y}|M = \text{NH})$, where $M = \text{NH}$ refers to the NH model, was computed with a model similar to the NH model, except that $[\mathbf{b}]$ was assigned a uniform distribution with support defined by upper and lower limits $\mathbf{b}_{\min} = \mathbf{0}$ and $\mathbf{b}_{\max} = \mathbf{20}$ ($\mathbf{0}$ and $\mathbf{20}$ are vectors of order equal to the number of elements in \mathbf{b}). Further, σ_c^2 was assigned a uniform prior with lower limit equal to 0 and upper limit equal to 20.0. Several other runs with varying limits for $[\mathbf{b}]$ and $[\sigma_c^2]$ gave virtually identical results.

Discrepancy measures derived from the predictive density $p(\mathbf{z}_2|\mathbf{z}_1, M_i)$ were computed using the unmodified NH model, because $p(\mathbf{z}_2|\mathbf{z}_1, M_i)$ is not affected by the impropriety of prior distributions. In fact, it can be shown that $p(\mathbf{z}_2|\mathbf{z}_1, M_i)$ has the form of a multivariate *t*-distribution.

DISCUSSION

The majority of the analyses presented in this work were based on the Bayesian approach to inference. An important objective of this work is to illustrate how this approach allows the analyst to quantify probabilistically the uncertainty involved in the description of direct and correlated selection responses. This is done conditionally on the data at hand, without invoking the use of problematic mental constructs involving the distribu-

tion of an estimator in conceptual repetitions of the experiment under identical conditions. Bayesian methods used in conjunction with MCMC provide considerable freedom in building hierarchical models that describe the perceived underlying structures in the data. We have also illustrated how a Bayesian analysis used in conjunction with MCMC allows models to be compared using a variety of criteria and how a sensitivity analysis—to the prior distribution or to the likelihood—can be implemented.

Concerning TNB, the most parsimonious model, the NH model yielded an estimate of selection response of 0.40 piglets per litter. Point inferences using this simple model are appealing, because no assumptions are made about the underlying genetic model (SORENSEN and KENNEDY 1986). The mean of the posterior distribution of selection response, $E(R_1|\mathbf{y})$, is approximately equal to $\bar{y}_{1...} - \bar{y}_{2...}$, the difference between phenotypic means of the selected and control lines. Thus, using the NH model response is evaluated using data from these lines only.

This is in contrast to the more highly parameterized additive genetic models, which invoke an infinitesimal additive model as the genetic mechanism through which response to selection is generated. These models were fitted here using the Gibbs sampler. The properties of this approach to the analysis of selection experiments are described in SORENSEN *et al.* (1994). Very succinctly, we mention here that, in general, the Bayesian approach provides a unified way of drawing inferences about variances and selection response, without recourse to analytic approximations or to *ad hoc* arguments. The posterior variance of a parameter in question (or a function of it), obtained here empirically via the Gibbs sampler, provides a probabilistic description of uncertainty, which accounts for the loss of information incurred in the estimation of other parameters of the model.

Posterior variances account for genetic drift; in a Bayesian setting, genetic drift can be interpreted as the relative loss of information about the genetic mean that evolves with time, due to the increased correlated structure in the data that builds as the experiment progresses. As a result of this, say, when the number of additive genetic values contributing to the (additive) genetic means of the different generations is the same, the posterior variance of the genetic mean increases with generations.

In the analyses based on hierarchical models, the whole data on which selection decisions were taken are included in the analysis. This holds for the analysis of direct response to selection for TNB and for the analysis of correlated response for WTAB. Therefore, joint and marginal posterior distributions are not affected by selection (GIANOLA and FERNANDO 1986). This implies that it is not required to model the selection process to draw inferences.

Expressions of response to selection obtained via the additive genetic model contain information arising

from the prior distribution, from what we have called the base and from animals that belonged to the selected and control lines in the common farm. The latter group comprised 1072 litter records, whereas the former contributed 8988 litter records. An attempt was made to quantify the information coming from these sources using the concept of observed statistical information and appealing to asymptotic results to estimate it. The amount of information [as defined in (13)] contributed by the prior distribution and by the 8988 litter records, relative to that from the selected and control lines, was almost 70%, and this number was consistent for all the prior distributions studied. In other words, had the lines not differed phenotypically in TNB, the additive genetic model-based inferences would have yielded a positive response. This is why it is important, especially from a genetic point of view, to contrast results obtained by fitting the additive genetic models and the NH model, which infers response to selection using the difference between the means of selected and control lines only. In our experiment, both types of models yielded similar inferences about selection response. This agreement is indicative that an additive genetic model is operationally adequate to describe the results of this experiment. Once this point is confirmed, there is little doubt that the hierarchical model yields a much sharper picture of the outcome of the selection experiment. The Monte Carlo estimate of the 95% probability interval of selection response for TNB based on the hierarchical model HYT-Nor is (0.23, 0.60); the corresponding interval from the NH model is (-0.07, 0.89).

The model assessment study shows that the most appropriate model among the ones fitted was the hierarchical model, which assumes that herd-year-type of insemination effects are, *a priori*, normally distributed. This conclusion is based on the marginal likelihood, which is a global test of the model, and on the ability of the models to predict specific features of the data. Due to the present experimental protocol, it was judged meaningful to study the latter, generating simulated data from posterior predictive distributions under a variety of models, conditional on data from base individuals. Replicated data could have easily been simulated from posterior predictive distributions under the various models, rather than conditioning on the base records. The general idea here is to study the ability of the models to generate replicated data that look similar to the data actually observed (GELMAN *et al.* 1995).

Among the criteria used to compare models, the NH model performed worse than the hierarchical models in all cases, except in the ability to predict the difference between the raw means of selected and control lines among parity 2 records. This statement is based on the mean of the relevant posterior distribution as a criterion to decide what is "best." However, despite the fact that the hierarchical models yield a mean of the posterior predictive difference larger than the observed differ-

ence (0.65 *vs.* 0.40), the observed value of 0.40 piglets per litter is well within the range of posterior uncertainty. We note also that the numbers in Table 6 show that the variance of the posterior predictive difference is 1.75 times larger in the case of the NH model.

Litter size is known to be affected by inbreeding depression—a phenomenon that the additive genetic models implemented in this study are unable to explain. To account for inbreeding depression, some form of interaction effects within or between loci, or both, is required as part of the mechanism responsible for genetic variation. Another way of testing the operational adequacy of the additive genetic model is to compare predictions of crossbred performance under such a model, using data from purebred performance. ANDERSEN *et al.* (1998) carried out such comparisons using purebred data from Danish Yorkshire and Landrace and the crossbred data from both breeds. The results of their analysis were consistent with what one would expect, assuming an additive genetic model. Under the low levels of inbreeding characteristic in the data from these studies, the added computational burden involved to account for nonadditive gene action does not seem to be justified.

The analysis of TNB and WTAB with model 5 indicates that the posterior probability that the correlated response in WTAB is negative is $\sim 76\%$. The 95% posterior probability interval for the genetic correlation is (-0.46; -0.01), and the Monte Carlo estimate of the mean of the posterior distribution is -0.21. The conclusions based on these numbers are qualitatively in agreement with those of KERR and CAMERON (1995) and ROEHE (1999). ROEHE (1999) used TNB as a covariate in a model for individual birth weights and he found that these decrease by 44 g with each additional piglet born. KERR and CAMERON (1995) found that individual birth weight decreases by 30 g with each additional piglet born. On the basis of the numbers in Table 4, a genetic regression of 55 g per born piglet is obtained, and a phenotypic regression of 23 g per born piglet.

Litter size has been part of the Danish breeding objective since 1992. The decision is fully justified on the grounds of the results of this experiment. However, it seems prudent to monitor birth weight and piglet viability, together with other components of reproduction, to avoid undesirable genetic changes.

LITERATURE CITED

- ANDERSEN, S., B. PEDERSEN and A. VERNERSEN, 1998 Impact of nucleus selection at production level. Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Armidale, Australia. Vol. XXIII, pp. 515-518.
- ÁVALOS, E., and C. SMITH, 1987 Genetic improvement of litter size in pigs. *Anim. Prod.* **44**: 153-164.
- BICHARD, M., and C. M. SEIDEL, 1983 Selection for reproductive performance in maternal lines of pigs. Proceedings of the 2nd World Congress on Genetics Applied to Livestock Production, Madrid, Spain. Vol. VIII, pp. 565-569.

- BICHARD, M., M. BOVEY, L. SEIDEL, P. DAVID and C. TOMKINS, 1983 *New Developments in Scientific Pig Breeding*, No. 3. Pig Improvement Company, Fyfield Wick, Abingdon, Oxfordshire, UK.
- BLASCO, A., J. P. BIDANEL and C. HALEY, 1995 Genetics and neonatal survival, pp. 17–38 *The Neonatal Pig. Development and Survival*, edited by M. A. VARLEY. CAB International, Wallingford, Oxon, UK.
- BOX, G. E. P., and G. C. TIAO, 1973 *Bayesian Inference in Statistical Analysis*. John Wiley & Sons, New York.
- CARLIN, B. P., and T. A. LOUIS, 1996 *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall, London.
- CASEY, D., T. A. RATHJE and R. K. JOHNSON, 1994 Second thoughts on selection for components of reproduction in swine. Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, Guelph, ON, Canada. Vol. XVII, pp. 315–318.
- ESTANY, J., and D. SORENSEN, 1995 Estimation of genetic parameters for litter size in Danish Landrace and Yorkshire pigs. *Anim. Sci.* **60**: 315–324.
- GARCIA-CORTES, L. A., and D. SORENSEN, 1996 On a multivariate implementation of the Gibbs sampler. *Genet. Sel. Evol.* **28**: 121–126.
- GEISSER, S., 1993 *Predictive Inference: An Introduction*. Chapman & Hall, London/New York.
- GELFAND, A. E., 1996 Model determination using sampling-based methods, pp. 145–161 in *Markov Chain Monte Carlo in Practice*, edited by W. R. WILKS, S. RICHARDSON and D. J. SPIEGELHALTER. Chapman & Hall, London/New York.
- GELMAN, A., and D. B. RUBIN, 1992 Inference from iterative simulation using multiple sequences (with discussion). *Stat. Sci.* **7**: 467–511.
- GELMAN, A., J. B. CARLIN, H. S. STERN and D. B. RUBIN, 1995 *Bayesian Data Analysis*. Chapman & Hall, London/New York.
- GEYER, C. J., 1992 Practical Markov chain Monte-Carlo (with discussion). *Stat. Sci.* **7**: 473–511.
- GIANOLA, D., and R. FERNANDO, 1986 Bayesian methods in animal breeding. *J. Anim. Sci.* **63**: 217–244.
- HALEY, C. S., E. ÁVALOS and C. SMITH, 1988 Selection for litter size in pigs. *Anim. Breed. Abstr.* **56**: 317–332.
- HENDERSON, C. R., 1973 Sire evaluation and genetic trends. Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush, Champaign, IL, pp. 10–41.
- KASS, R. E., and A. E. RAFTERY, 1995 Bayes factors. *J. Am. Stat. Assoc.* **90**: 773–795.
- KERR, J. C., and N. D. CAMERON, 1995 Reproductive performance of pigs selected for components of efficient lean growth. *Anim. Sci.* **60**: 281–290.
- LEGAULT, C., 1985 Selection of breeds, strains and individual pigs for prolificacy. *J. Reprod. Fertil.* **33** (Suppl.): 151–166.
- NEWTON, M. A., and A. E. RAFTERY, 1994 Approximate Bayesian inference by the weighted likelihood bootstrap. *J. R. Stat. Soc. Ser. B* **56**: 3–48.
- O'HAGAN, A., 1994 *Kendall's Advanced Theory of Statistics, Vol. 2B: Bayesian Inference*. Edward Arnold, London.
- RAFTERY, A. E., and S. M. LEWIS, 1993 How many iterations in the Gibbs sampler? pp. 763–773 in *Bayesian Statistics*, Vol. 4, edited by J. M. BERNARDO, J. O. BERGER, A. P. DAWID and A. F. M. SMITH. Oxford University Press, Oxford.
- ROEHE, R., 1999 Genetic determination of individual birthweight and its association with sows' productivity traits using Bayesian analysis. *J. Anim. Sci.* **77**: 330–343.
- ROTHSCHILD, M., and J. P. BIDANEL, 1998 Biology and genetics of reproduction, pp. 313–343 in *The Genetics of the Pig*, edited by M. F. ROTHSCCHILD and A. RUVISNKY. CAB International, Wallingford, Oxon, UK.
- SMITH, C., G. E. DICKERSON, M. W. TESS and G. L. BENETT, 1983 Expected relative responses to selection for alternative measures of life cycle economic efficiency of pork production. *J. Anim. Sci.* **56**: 1306–1314.
- SORENSEN, D., 1996 *Gibbs Sampling in Quantitative Genetics*, Internal Report No. 82, National Institute of Animal Science, Tjele, Denmark.
- SORENSEN, D., and B. W. KENNEDY, 1986 Analysis of selection experiments using mixed model methodology. *J. Anim. Sci.* **63**: 245–258.
- SORENSEN, D., C. S. WANG, J. JENSEN and D. GIANOLA, 1994 Bayesian analysis of genetic change due to selection using Gibbs sampling. *Genet. Sel. Evol.* **26**: 333–360.
- SORENSEN, D., S. ANDERSEN, D. GIANOLA and I. R. KORSGAARD, 1995 Bayesian inference in threshold models using Gibbs sampling. *Genet. Sel. Evol.* **27**: 229–249.
- TESS, M. W., G. L. BENETT and G. DICKERSON, 1983 Simulation of genetic changes in life cycle efficiency of pork production. II. Effects of components on efficiency. *J. Anim. Sci.* **56**: 354–368.
- TOMES, G. J., and H. E. NIELSEN, 1984 Some aspects of selection for increased reproductive performance in sow herds. Proceedings of the 8th International Pig Veterinary Society Congress, Ghent, Belgium, p. 203.

Communicating editor: C. HALEY