Estimating Daughter Yield Deviation and Validation of Genetic Trend for Somatic Cell Score in Holstein Cattle Using Random Regression Test Day Model

H. Khanzadeh\textsuperscript{1}, and N. Ghavi Hossein-Zadeh\textsuperscript{1*}

ABSTRACT

The objective of this study was to estimate Daughter Yield Deviations (DYDs) of bulls and Yield Deviations (YDs) for cows using a random regression model and validation of genetic trend using estimated DYDs and Method II of Interbull for test-day records of Somatic Cell Score (SCS) in the first lactation of Iranian Holsteins. Data set included the 108995 test day records collected by the Animal Breeding Center of Iran from 2001 to 2010. Results of the present study indicated that variation in YDs of cows at different stages of lactation corresponds closely with their Estimated Breeding Values (EBVs). Because YDs and DYDs are considered as an additional measure of an animal’s genetic merit, their correlation with EBVs is very important. The correlation between DYDs and EBVs of bulls for SCS was 0.88. High correlation estimates between DYDs and EBVs indicated that, in addition to EBV, the DYD can be an appropriate measure for dairy cattle breeding programs. The correlation increased with increase in the number of bull daughters and the average number of test-days of daughters. Estimated DYDs for each production year were used to validate the genetic trend obtained from the model which was used for genetic evaluation. Results indicated that genetic trend for SCS in the first lactation of Iranian Holsteins was slightly overestimated.

Keywords: Dairy cow, Estimated breeding value, Genetic progress, Mastitis, Validation of genetic trend.

INTRODUCTION

Mastitis, or inflammation of the mammary gland, is one of the most complex and costly diseases affecting dairy cattle. Costs due to clinical mastitis include lower milk production, poor milk quality, discarded milk, veterinary costs, and premature culling of cows (Kadarmideen and Pryce, 2001; Koivula \textit{et al.}, 2004). Selection against mastitis, in countries where incidences are not recorded, is carried out indirectly by selecting against Somatic Cell Count (SCC) (Mrode and Swanson, 1996). Daily Somatic Cell Score (SCS) has usually been analyzed as repeated measurements of the same trait. However, the genetic correlations between SCS at different stages of lactation are less than unity, which violates the assumptions of the repeatability model (Reents \textit{et al.}, 1994; Mrode \textit{et al.}, 1996; Haile Mariam \textit{et al.}, 2001; Ødegard \textit{et al.}, 2003). The recent trend in dairy cattle genetic evaluations is towards the application of Random Regression Models (RRM) using Test Day (TD) records (Mrode and Swanson, 2004) that eliminated the deficiency of repeatability model.

Besides the Estimated Breeding Value (EBV), Yield Deviation (YD) of cows and Daughter Yield Deviation (DYD) of bulls are important quantities used in dairy cattle selection (Szyda \textit{et al.}, 2008). The YD is a weighted average of the cows’ yields adjusted...
for all effects of the model other than genetic merit and error. The DYD of bulls is the average performance of their daughters that are adjusted for fixed and non-genetic random effects of the daughters and genetic effect of their mates (Van Raden and Wiggans, 1991; Liu et al., 2004; Freyer et al., 2002). DYD is not regressed on breeding value of bulls and is the most independent and accurate measure of phenotypic performance of a bull’s daughters (Van Raden and Wiggans, 1991; Liu et al., 2004). Van Raden and Wiggans (1991) showed the calculation method of YD and DYD for repeatability animal model. Mrode and Swanson (2002) presented this calculation for a random regression model. Liu et al. (2003) developed a method for calculation of DYD under general multiple trait models.

For SCS records, Mrode and Swanson (2004) reported DYD for Holstein-Friesian heifers. Also Liu et al. (2004) calculated DYD for SCS in Holstein, Red and Jersey dairy cattle from Austria, Germany and Luxembourg. Calculation of YD and DYD in Iranian Holsteins was performed by Sheikhloo et al. (2009) for milk and fat traits using repeatability animal model. Khanzadeh et al. (2013) calculated YD and DYD for production traits of Iranian Holsteins using both repeatability animal and Random Regression Test Day Models (RRTDMs). However, the YD and DYD have not been calculated for SCS records until now in Iranian dairy cows. Hence, the objective of this study was to estimate YD and DYD for SCS using RRM and its application for the genetic evaluation of Iranian Holsteins.

MATERIALS AND METHODS

Data and Model

A total of 108,995 daily SCS records from 2001 to 2010 for the first lactation of the Iranian Holsteins were obtained from the Animal Breeding Center of Iran and analyzed by the following RRM using the AIREML algorithm of the WOMBAT program (Meyer, 2006):

\[
Y_{\text{unprev}} = YS_i + HTD_m + \sum_{f=1}^{2} C_f (age_n)^f + \\
\sum_{r=0}^{k-1} \beta_r \varnothing_r (dim_i) + \sum_{r=0}^{k_p} \alpha_r \varnothing_r (dim_i) + \\
\sum_{r=0}^{k_r} \gamma_r \varnothing_r (dim_i) + e_{\text{unprev}}
\]

Where, \(Y_{\text{unprev}}\) is test day SCS record \(i\) obtained at \(DIM_i\) of cow \(p\) calved at the \(n^{th}\) age in year-season of calving \(s\) and herd-test date \(m\); \(YS_i\) is fixed effect of the \(s^{th}\) year-season of calving; \(HTD_m\) is fixed effect of the \(m^{th}\) herd-test date; \(C_f\) is the \(f^{th}\) fixed regression coefficient for calving age; \(age_n\) is the \(n^{th}\) calving age; \(k\) is the order of fit for fixed regression coefficients (\(k=3\) or \(4\)), \(\beta\) is the \(r^{th}\) fixed regression coefficient; \(k_p\) is the order of fit for additive genetic random regression coefficients; \(k_r\) is the order of fit for permanent environmental random regression coefficients; \(\alpha_r\) is the \(r^{th}\) random regression coefficient of additive genetic value of \(p^{th}\) cow; \(\gamma_r\) is the \(r^{th}\) random regression coefficient of permanent environmental effect of \(p^{th}\) cow; \(\varnothing_r (dim_i)\) is the \(r^{th}\) coefficient of Legendre polynomials evaluated at days in milk \(t\); and \(e_{\text{unprev}}\) is the random residual error.

In general, 16 different models were fitted for the analysis of the data set (Table 1). These models differed in terms of the Legendre polynomials used to fit the covariance functions for additive genetic and permanent environmental effects and in the number of classes for the residual variances. Selection of models was based on Akaike’s Information Criterion (AIC) (Akaike, 1973). Model 15 had the lowest AIC value, therefore, it was chosen for the analysis of SCS records.

Calculating Yield Deviations

Equations to calculate the contribution of information from different sources of random regression coefficients in the random regression model for any animal is
Table 1. Different orders of fit for random regression coefficients in this study.

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed regression order of fit</th>
<th>ka</th>
<th>kp</th>
<th>Residual variance class</th>
<th>Np</th>
<th>Log l</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>16</td>
<td>-101666.74</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>10</td>
<td>22</td>
<td>-101659.68</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>10</td>
<td>24</td>
<td>-101582.75</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>10</td>
<td>26</td>
<td>24</td>
<td>-101575.09</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>10</td>
<td>24</td>
<td>-101581.36</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>10</td>
<td>30</td>
<td>29</td>
<td>-101573.64</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>10</td>
<td>35</td>
<td>-101541.436</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>10</td>
<td>35</td>
<td>35</td>
<td>-101544.567</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>10</td>
<td>16</td>
<td>-101583.32</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>10</td>
<td>22</td>
<td>20</td>
<td>-101577.75</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>10</td>
<td>26</td>
<td>-101495.36</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>10</td>
<td>26</td>
<td>24</td>
<td>-101500.81</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>10</td>
<td>30</td>
<td>-101494.24</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>10</td>
<td>35</td>
<td>-101466.114</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>29</td>
<td>35</td>
<td>-101462.042</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>10</td>
<td>35</td>
<td>35</td>
<td>-101462.042</td>
</tr>
</tbody>
</table>

* Orders of fit for additive genetic and permanent environmental effects, respectively; b Number of parameters for estimated variance function; c Maximum log likelihood; d Akaike’s Information Criterion; Bold values correspond to the best model.

Presented by Mrode and Swanson (2004). The equation for YD calculation in RRM can be written as follows:

\[
YD = (Q R^{-1}Q)^{-1}(Q R^{-1}Y_C)
\]

Where, YD is a vector of weighted regressions of the animal’s TD yields adjusted for all effects other than additive genetic effect on orthogonal polynomials for DIM, Q is a matrix of orthogonal polynomials of days in milk of order 4 for random animal effect, R is a diagonal matrix for residual variances and YC is a vector of test day records of cows that is adjusted for all effects in the model, except the additive genetic and residual effects. YD was estimated for 12,142 Holstein cows.

Calculating Daughter Yield Deviations

To calculate DYD of a bull, only records of his own daughters should be considered, male progeny must be excluded, because they do not have own performance records for production traits (Liu et al., 2003). In this study, DYD was calculated as follows (Mrode and Swanson, 2004):

\[
DYD = \sum G^{-1}W_{prog} q_{prog} (2YD_{prog} - u_{mate})
\]

\[
G^{-1}W_{prog} q_{prog}
\]

Where, DYD is a vector of DYD of bulls expressed as random regression coefficients; G is genetic covariance matrix; qprog equals 1 if other parent of the progeny is known and \( \frac{2}{3} \) if unknown; YDprog and umate are estimated YD for daughters of bull and breeding value for mates of bull, respectively; and

\[
W_{prog} = (G^{-1}a_{anim} + DIAG)^{-1} DIAG
\]

where, \( a_{par} = 1, 2/3 \) or 1/2 if both, one, or neither parents are known, respectively, and \( a_{prog} = 1 \) if animal’s mate is known and \( 2/3 \) if unknown. Note that \( a_{anim} = 2a_{par} + 0.5a_{prog} \) and \( DIAG = Q R^{-1}Q \). Computation of DYD was performed using the computing strategy of DYD illustrated by Mrode and Swanson (2004). In the present study, bulls with number of daughters less than 10 were removed from the data set.
Validation of Genetic Trend

Boichard et al. (1995) described a method to validate the estimation of genetic trend using \( \text{DYD} \) (Method II of Interbull). Genetic trend validation comprises the estimation of a regression coefficient of \( \text{DYD} \) on the production year (Szyda et al., 2008). After estimating \( \text{DYD} \), regression coefficient of \( \text{DYDs} \) on the bulls' birth years was estimated using the regression procedure of the SAS software package (SAS, 2002). The model is validated by Interbull when the absolute value of the regression coefficient is less than \( 0.01 \times \text{SD} \), where \( \text{SD} \) is the genetic Standard Deviation for the trait (van Steenbergen et al., 2005).

RESULTS AND DISCUSSION

Simple descriptive statistics calculated for \( \text{YDs} \) and \( \text{EBVs} \) of cows for 305 days are presented in Table 2. Because \( \text{YD} \) and \( \text{DYD} \) in RRM expresses in the form of regression coefficients, any linear function of the regression coefficient estimates can be derived for individual cows and bulls, respectively. In the present study, \( \text{EBVs} \) and \( \text{YDs} \) were calculated for individual days of lactation for three cows (with positive \( \text{EBV} \), with negative \( \text{EBV} \), and with \( \text{EBV} \) close to zero). These \( \text{EBVs} \) and \( \text{YDs} \) are presented in Figure 1. As shown in Figure 1, in these cows, the trend in daily \( \text{YDs} \) at different stages of lactation corresponds closely to their \( \text{EBVs} \). As indicated in Table 2, variation of \( \text{YDs} \) is greater than the variation of \( \text{EBVs} \) which is sensible for the three cows presented in Figure 1. The \( \text{DYD} \) statistics for SCS records and correlation of \( \text{DYD} \) with \( \text{EBV} \) by number of daughters are presented in Table 3. As average number of daughters increased, correlation between \( \text{DYD} \) and \( \text{EBV} \) increased.

As shown in Table 4, correlation between \( \text{EBVs} \) and \( \text{DYDs} \) are much more variable when the average number of test-days per bull’s daughters is considered. As average number of test-days increased, correlation also increased. High correlations (> 0.90) between \( \text{DYD} \) and \( \text{EBV} \) were observed, on average, with the minimum of 9 test-days per bull’s daughter. As indicated in Tables 2 and 3, \( \text{DYD} \) are less variable than \( \text{YD} \). \( \text{EBVs} \) and \( \text{DYDs} \) means of 305 days and

### Table 2. Descriptive statistics for \( \text{YD} \).

<table>
<thead>
<tr>
<th></th>
<th>Yield Deviations (( \text{YD} ))</th>
<th>Estimated Breeding Values (( \text{EBV} ))</th>
<th>Correlation of ( \text{YD} ) with ( \text{EBV} )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>-372.25</td>
<td>-0.31</td>
<td>0.50</td>
</tr>
<tr>
<td><strong>STD</strong></td>
<td>140.92</td>
<td>34.83</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1.** (A) Yield deviations, and (B) \( \text{EBVs} \) for SCS at different stages of lactations for a cow with a positive \( \text{EBV} \) (cow A), a cow with \( \text{EBV} \) very close to zero (cow B) and a cow with a negative \( \text{EBV} \) (cow C).
Table 3. Descriptive statistics for DYD and correlation of DYD with EBV by number of daughters.

<table>
<thead>
<tr>
<th>Number of daughters</th>
<th>Number of bulls</th>
<th>DYD Mean</th>
<th>DYD STD</th>
<th>EBV Mean</th>
<th>EBV STD</th>
<th>Correlation of DYD with EBV</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 - 19</td>
<td>172</td>
<td>-367.84</td>
<td>41.55</td>
<td>-2.25</td>
<td>38.65</td>
<td>0.87</td>
</tr>
<tr>
<td>20 - 49</td>
<td>166</td>
<td>-362.76</td>
<td>36.44</td>
<td>5.76</td>
<td>47.50</td>
<td>0.89</td>
</tr>
<tr>
<td>50 - 99</td>
<td>59</td>
<td>-369.16</td>
<td>35.82</td>
<td>-1.88</td>
<td>57.55</td>
<td>0.95</td>
</tr>
<tr>
<td>≥ 100</td>
<td>6</td>
<td>-403.84</td>
<td>46.49</td>
<td>-49.26</td>
<td>58.53</td>
<td>0.97</td>
</tr>
<tr>
<td>Total</td>
<td>403</td>
<td>-366.48</td>
<td>38.98</td>
<td>0.40</td>
<td>46.71</td>
<td>0.88</td>
</tr>
</tbody>
</table>

Table 4. Correlation of DYD with EBV by average number of test day records of daughters.

<table>
<thead>
<tr>
<th>Average TD records of daughters</th>
<th>Number of bulls</th>
<th>Average no of daughters</th>
<th>DYD Mean</th>
<th>DYD STD</th>
<th>EBV Mean</th>
<th>EBV STD</th>
<th>Correlation of DYD with EBV</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 – 8.5</td>
<td>85</td>
<td>23</td>
<td>-357.68</td>
<td>38.33</td>
<td>3.96</td>
<td>46.79</td>
<td>0.89</td>
</tr>
<tr>
<td>8.51 - 9</td>
<td>283</td>
<td>33</td>
<td>-368.44</td>
<td>37.80</td>
<td>0.39</td>
<td>46.33</td>
<td>0.87</td>
</tr>
</tbody>
</table>
| > 9                             | 35             | 21                      | -371.93  | 47.24   | -8.16    | 45.81   | 0.92                      

correlations between EBVs and DYDs stratified by bulls’ birth year are shown in Figure 2. Correlation in all birth years was high and variation of EBV and DYD means were symmetric.

In the present study, validation of genetic trend was performed based on Interbull Method II for SCS records of Iranian Holsteins. Additive genetic standard deviation was 4.76 and regression coefficient of DYD on production years (SE) was 1.12 (0.37). Regression coefficients of DYDs on bull birth years, calculated for SCS, was positive and greater than 0.01×SD (0.23×SD), indicating that genetic trend was slightly overestimated. Bonaiti et al. (1994) indicated that when the estimate of genetic trend is unbiased, the year effect has a zero expectation, and should not significantly differ from zero. Alternatively, the year effect shows a decreasing or increasing trend when the estimate of genetic trend is underestimated or overestimated, respectively. Similar trends in daily YDs and EBVs at different stages of lactation are also reported by Mrode and Swanson (2004). YD provides a good indication of contributions from yield records of the cow to her PTA (Predicted Transmitting Ability), thus, YD could be useful in understanding cow evaluations (Mrode and Swanson, 2004).

Calculation of DYD requires estimates of all fixed effects and non-genetic random effects and EBV of bulls’ mates obtained from a genetic evaluation (Liu et al., 2003). For bulls with granddaughters, DYD does not include all information from descendants, because information from granddaughters and sons is excluded (Van Raden and Wiggans, 1991). In DYD calculation, only the path of cow to sire is considered and other paths such as son to sire are ignored. Because DYD of one bull does not affect DYD of other bulls, DYD calculation can be done on a within-bull basis. Parental contribution to bull is irrelevant for the calculation of DYD of bulls (Liu et al., 2003). For routine genetic evaluations, 305-day lactation DYD values and DYD lactation curves are published for bulls satisfying the requirement for official DYD mentioned above, in addition to lactation EBV and genetic lactation curves for bulls (Liu et al., 2004).
Increase in correlation between $DYD$ and $EBV$ with increase in the number of daughters per bull in this study is in agreement with results obtained by Liu et al. (2003, 2004), Mrode and Swanson (2004), Szyda et al. (2008) and Khanzadeh et al. (2013). This is expected as

$$DYD = \frac{1}{2} a_{\text{bull}} + \sum_{m} e / m,$$

where $m$ is the number of daughters. Therefore, as $m$ increases, $\sum_{m} e / m$ tends towards zero and correlation of $DYD$ and $EBV$ increases (Mrode and Swanson 2004). Currently, $DYD$ is provided to the dairy industry for bulls with 10 or more daughters, but not for cows (Van Raden and Wiggans, 1991). Lactation $DYDs$ estimated from shorter lactations are more influenced by extrapolation than lactation $DYDs$ from longer lactations. To minimize the impact of extrapolation, at least 10 daughters are required to pass 120 DIM in lactation in order to make $DYD$ of this lactation official for a bull (Liu et al., 2003).

Liu et al. (2003), Szyda et al. (2008) and Khanzadeh et al. (2013) reported that correlations between bulls’ $EBVs$ and $DYDs$ increased by increasing average number of test-days per daughter. Szyda et al. (2008) suggested that a large number of test-days (minimum 6) is required to obtain a good projection of $DYDs$ over a 305-day lactation, while with increasing number of daughters per bull (over 40), the residual variance component of the total $DYD$ variance strongly decreases. Since $DYDs$ are considered as an additional measure of animal’s genetic merit, their correlation with $EBVs$ is of primary importance (Szyda et al., 2008).

The large variation of $YD$ is partially caused by short lactations of cows (Szyda et al., 2008). On the other hand, $YD$ is at cow level (based on one or few records) while $DYD$ is at bull level (based on records from a large number of daughters). Therefore, variation between $YD$ is larger than $DYD$.

Expected $DYD$ values should only depend on the bull and are theoretically independent of any environmental effect, particularly the birth year of the daughters. This property of the residuals may be used to validate the estimation of genetic trend, which is in this case the combined sire trends (Theron et al., 2002). A possible reason for the observed overestimation of a trend could arise from the fact that modeling of time-related effects was not very accurate in the corresponding genetic evaluation model, e.g. age of calving or year of calving (Szyda et al., 2008).
CONCLUSIONS

YDs and DYDs were calculated for Iranian Holstein SCS records and, then, genetic trends were validated using calculated DYDs. Trend in daily YDs of cows at different stages of lactation corresponded closely with their daily EBVs. Calculated DYDs were highly correlated with EBVs; therefore, it can be indicated that, in addition to EBV, the DYD can be an appropriate measure for dairy cattle breeding programs. Results of the validation of genetic trend with the Interbull Method II indicated that estimate of genetic trend for SCS was slightly overestimated.

ACKNOWLEDGEMENTS

Authors wish to thank the Animal Breeding Center of Iran for providing the data used in this study.

REFERENCES


