

Single-step genomic evaluation for linear type traits of Holstein cows in Czech Republic*

Ludmila Zavadilová**, Josef Příbyl, Luboš Vostrý, Jiří Bauer

Institute of Animal Science, Přátelství 815, 104 01 Praha - Uhřetěves, Czech Republic

(Accepted June 3, 2014)

The single-step method of genomic evaluation using both a uni-trait and a multi-trait animal model was applied for 20 linear type traits of Holstein cows in the Czech Republic. Phenotypic data on linear scoring were available for 143,208 Holstein cows first calved between years 2005 to 2010. Single nucleotide polymorphism (SNP) markers from the Illumina BovineSNP50 BeadChipV2 (Illumina Inc., San Diego, USA) were available for 631 sires. Breeding values (traditional breeding value, EBV) were estimated by uni-trait and multi-trait BLUP animal models using the pedigree-based relationship matrix. Genomic breeding value (GEBV) was estimated by single step genomic BLUP (ssGBLUP) using the pedigree-based relationship matrix augmented by the genomic relationship matrix. The model included fixed effects of herd-date of classification, classifier and season of calving, linear and quadratic regressions on age at calving and on days in milk and the random additive genetic effect of animal. Correlations between parent average, predicted GEBV and post-progeny test EBV were slightly higher for uni-trait analyses than for multi-trait analyses. Excluding udder traits, the average uni-trait correlations with the post-progeny test EBV were 0.44 for parent average and 0.46 for predicted GEBV, whereas multi-trait correlations were 0.42 and 0.44, respectively. When udder traits were included, average uni-trait correlations with the post-progeny test EBV were 0.09 for parent average and 0.14 for predicted GEBV, while multi-trait correlations were 0.29 and 0.33, respectively.

KEYWORDS: Holstein cattle / linear type traits / single-step genomic evaluation

Generally, results of genomic evaluation show a superiority of genomic evaluations over pedigree-only based ones. Misztal *et al.* [2009], Aguilar *et al.* [2010] and

*Supported by the Ministry of Agriculture of Czech Republic, Projects No. Mze 0002701404 and Q1111A167

** Corresponding author: zavadilova.ludmila@vuzv.cz

Christensen and Lund [2010] described a single-step method of genomic evaluation. That simultaneously uses phenotypic, genomic and relationship information. This method was successfully applied for final score of over 6 million Holsteins [Aguilar *et al.* 2010]. Although the multi-trait, single-step approach is computationally demanding for large datasets [Su *et al.* 2012], the Czech HF population is small. In practice, the multi-trait single-step approach was feasible for 18 linear traits, as reported by Tsuruta *et al.* [2011] in US Holsteins and for conception rate as described by Aguilar *et al.* [2010].

A primary reason for collecting and utilizing information on type traits is to aid breeders in selecting profitable, functional cows and to reduce early culling for causes unrelated to yield (involuntary culling) – Misztal *et al.* [1992]. The Czech Holstein cattle population has been monitored for linear type traits, at least in part, for almost 20 years. Genetic evaluation is practiced and breeding values for linear type traits are used for selecting Holstein sires. Research has focused on the relationship of linear type traits with longevity, reproduction and milk production [Zavadilová *et al.* 2011, Zink *et al.* 2011, Zavadilová and Štípková 2012]. Genetic correlations among type traits have also been estimated [Němcová *et al.* 2011]. In the Czech Republic, intensive research has been carried out for the implementation of genomic evaluation using the single-step method [Příbyl *et al.* 2012] primarily for production traits of dairy cattle.

The aim of the present study was to assess the single-step method of genomic evaluation in uni-trait and multi-trait animal models for genetic evaluation of 20 linear type traits of Holstein cattle in the Czech Republic.

Material and methods

For the most part of Holstein cattle population, linear scoring was applied on first-parity Holstein cows only after year 2000. In addition, there have been methodological changes in linear scoring concerning small but significant changes in the definition of type traits, and some additional exterior traits (body condition score, locomotion, hock quality) have been added. Therefore only that portion of the database that includes the newest traits was used for analysis. Phenotypic data on linear scoring were available for 143,208 Holstein cows that delivered their first calf between year 2005 and 2010. Type classification records were extracted from the Czech-Moravian Breeder's Corporation official database. All cows were scored for conformation traits between 30 and 210 days in milk (DIM) of their first lactation. The following type traits scored on a 9-point scale were analysed: body condition score (BCS), eight udder traits – fore udder attachment, rear udder height, udder depth, rear udder width, central ligament, teat length, front teat placement and rear teat position; six body traits – stature, angularity, chest width, body depth, rump angle, and rump width; five foot and leg traits – rear legs rear view, rear leg set (side view), foot angle, bone quality and locomotion. The analysed traits were scored on all cows, except of locomotion (129,131 records). The basic statistics for analysed linear traits are given in Table 1.

Table 1. Means, standard deviations and heritability of linear type traits, mean and standard deviations (SD) for traditional breeding value after progeny test (EBV_i) of the 101 young genotyped bulls (testing set) obtained by multi-trait analysis

Trait	h ²	Phenotypic	EBV _i
		mean (SD)	multi-trait mean (SD)
Angularity	0.30	5.46 (1.18)	0.66(0.41)
BCS	0.30	4.90 (1.20)	-0.57(0.44)
Chest width	0.18	5.65 (1.30)	0.09(0.41)
Body depth	0.24	5.62 (1.35)	0.31(0.46)
Stature	0.45	5.84 (1.32)	0.81(0.55)
Rump angle	0.34	4.81 (1.20)	-0.21(0.53)
Rump width	0.40	5.63 (1.31)	0.49(0.55)
Rear legs rear view	0.10	5.32 (1.56)	0.09(0.31)
Rear legs set (side view)	0.16	4.96 (1.26)	0.04(0.28)
Foot angle	0.10	4.98 (1.13)	0.06(0.21)
Bone quality	0.28	5.70 (1.36)	0.35(0.53)
Locomotion	0.05	5.19 (1.60)	0.08(0.26)
Fore udder attachment	0.24	5.11 (1.47)	0.31(0.52)
Rear udder height	0.24	5.44 (1.40)	0.61(0.37)
Udder depth	0.32	5.74 (1.41)	0.27(0.49)
Rear udder width	0.17	5.35 (1.45)	0.57(0.31)
Central ligament	0.20	5.62 (1.49)	0.30(0.46)
Teat length	0.28	4.63 (1.16)	0.00(0.47)
Front teat placement	0.26	5.02 (1.22)	0.48(0.45)
Rear teat position	0.28	5.68 (1.42)	0.53(0.55)

The genotypic data were edited both by animal and *by loci*. Bulls with genotype call rates <95% were excluded from further evaluation (4 bulls). SNPs were discarded if more than 5% of SNP calls were missing. The procedure used for selecting SNPs was described by Přibyl *et al.* [2012]. The 631 bulls that were genotyped using the Illumina BovineSNP50 BeadChip V2 (Illumina Inc., San Diego, USA) were used for preparation of a genomic relationship matrix (**G**).

For the prediction of parent averages of test bulls, a data subset was created that included 66,285 cows calved between year 2005 and 2007. A training set consisted of 221 sires that had at least 20 first crop daughters that calved between year 2005 and 2007. The testing set included 101 sires with SNP information available and with a minimum of 50 daughters calving for the first time after year 2007 (average of 68 daughters, with maximum 438 daughters).

The model equation can be described as follows:

$$Y_{ijklmn} = \mu + HDC_i + C_j + S_k + a_l + \beta_1 age_m + \beta_2 age_m^2 + \gamma_1 s_n + \gamma_2 s_n^2 + e_{ijklmn}$$

where: dependent variables (y_{ijklmn}) were linear type trait scores and fixed effects were HDC_i (herd-date of classification-classifier), C_j (classifier) and S_k (season of calving). The model included linear and quadratic regressions on age at calving ($\beta_1 age_m + \beta_2 age_m^2$) and linear and quadratic regressions on DIM at scoring ($\gamma_1 s_n + \gamma_2 s_n^2$). Random effects were animal (a_l) with a relationship matrix and the residual term (e_{ijklmn}).

The equation of the linear model used to predict the genomic breeding values is the same as that used in routine estimation of breeding values for type traits in Czech Republic. This model was tested by Němcová *et al.* [2011].

The variance-covariance matrix between linear type traits was estimated as described by Němcová *et al.* [2011]. The pedigree for the data set included 385,188 animals and the pedigree for the data subset was 225,504 animals (four generations of ancestors). The DMU program [Madsen and Jensen 2010] was used for the EBV and GEBV predictions, respectively.

For the entire dataset as well as the data subset, breeding values were estimated by uni-trait and multi-trait BLUP animal models using only the pedigree-based relationship matrix (traditional breeding value, EBV) and by BLUPss with the pedigree-based relationship matrix augmented by the genomic relationship matrix (genomic breeding value, GEBV). The single-step method of genomic evaluation [Aguilar *et al.* 2010, Christensen and Lund 2010] was employed for the estimation of GEBV. Genomic relationship matrix \mathbf{G} was calculated according to deviations from the averages of observed allele frequencies and standardized by division by the average value of the diagonal of \mathbf{G} , so that average of diagonal elements = 1 [Forni *et al.* 2011], and shifted, that the elements of additive pedigree relationship matrix only for genotyped animals A22 and elements of \mathbf{G} have the same average [Vitezica *et al.* 2011]. The genomic relationship was arbitrarily assigned weight of 99% .

Four breeding values were estimated for the 101 young bulls in the testing set: EBV post-progeny test (EBV_t) using the phenotypic records from the whole data set, the GEBV post-progeny test ($GEBV_t$) using the phenotypic records from the whole data set plus genomic information, parent average (PA) using the data subset and GEBV predicted using the data subset and genomic information ($GEBV_p$). Merit of the single-step method of genomic evaluation was assessed by comparison of correlations between EBV_t and PA or EBV_t and $GEBV_p$ of young sires from the testing set – (data subset) Pryce *et al.* [2011]. The correlations were calculated using Proc Corr by SAS [2009]. To assess the bias of GEBV, the difference of average EBV_t or PA from EBV of a reference group of proven bulls was calculated to test the possible over-evaluation of young animals. The reference group of proven bulls was composed of 400 sires with minimum of 50 daughters with records in data subset.

Results and discussion

Correlations between uni-trait and multi-trait EBV_t of young bulls by traits are shown in Table 2. Most of correlations were unity or higher than 0.96 (locomotion). Examination of correlations of EBV_t with parent averages (PA) and $GEBV_p$ of young bulls by traits and analysis (Tab. 2) showed that uni-trait as well as multi-trait analysis provided similar correlations for each trait except udder traits and locomotion. For most traits, multi-trait $GEBV_p$ showed a higher correlation with EBV_t than did multi-trait PA. At the same time, the uni-trait correlations between PA and EBV_t and those

Table 2. Correlations between traditional breeding value after progeny test (EBV_t), parent average (PA) and predicted genomic breeding value ($GEBV_p$) of young bulls (testing set) obtained by uni-trait or multi-trait analysis for analysed type traits

Trait	uni- EBV_t × multi- EBV_t	Uni-trait analysis		Multi-trait analysis	
		EBV_t	EBV_t	EBV_t	EBV_t
		× PA	× $GEBV_p$	× PA	× $GEBV_p$
Angularity	0.99	0.55	0.56	0.52	0.51
BCS	1.00	0.48	0.55	0.43	0.51
Chest width	0.99	0.45	0.49	0.44	0.48
Body depth	1.00	0.38	0.43	0.37	0.43
Stature	1.00	0.28	0.31	0.30	0.31
Rump angle	1.00	0.45	0.49	0.45	0.49
Rump width	1.00	0.47	0.55	0.46	0.55
Rear legs rear view	0.97	0.36	0.33	0.27	0.26
Rest legs set (side view)	0.99	0.49	0.49	0.51	0.51
Foot angle	0.98	0.46	0.43	0.48	0.46
Bone quality	1.00	0.49	0.59	0.47	0.57
Locomotion	0.96	0.40	0.26	0.28	0.22
Fore udder attachment	1.00	0.11	0.18	0.11	0.17
Rear udder height	0.99	-0.03	0.12	0.20	0.31
Udder depth	1.00	0.03	0.14	0.10	0.21
Rear udder width	0.99	0.37	0.41	0.32	0.39
Central ligament	1.00	0.08	-0.06	0.47	0.51
Teat length	1.00	0.10	0.16	0.23	0.32
Front teat placement	1.00	-0.18	-0.16	0.42	0.45
Rear teat position	1.00	0.26	0.33	0.43	0.43

Correlations of value $\geq 0.20 = P < 0.05$.

Correlations of value $\geq 0.25 = P < 0.01$.

between $GEBV_p$ and EBV_t were often slightly higher than corresponding correlations from multi-trait analyses. Averaged over all traits except those scoring the udder, uni-trait correlations with EBV_t were 0.44 for PA and 0.46 for $GEBV_p$, while the multi-trait correlations were 0.42 and 0.44 respectively.

For udder traits, correlations between EBV_t and PA were higher from the multi-trait analyses than from the uni-trait analyses, except for rear udder width and fore udder attachment. Averaging over udder traits, the uni-trait correlations with EBV_t were 0.09 for PA and 0.14 for $GEBV_p$, while multi-trait correlations were 0.29 and 0.33, respectively. The low uni-trait average is caused by rear udder height and front teat placement, which provided negative uni-trait correlations between current EBV_t and PA. On the contrary, for locomotion and rear legs rear view, correlations between EBV_t and PA from the uni-trait analyses were substantially higher than the correlations related to multi-trait analyses. For udder traits, multi-trait analysis results fluctuated less than results from uni-trait analysis and were more in line with results for other characters.

In this study, there was no clear relationship between the heritability of linear traits and correlations between EBV_p , PA and $GEBV_p$. There are indications in presented

results, that lowly heritable traits such as locomotion exhibit smaller increases in accuracy of evaluation from including genomic information in the analysis than do more highly heritable traits. On the other hand, it cannot be said that highly heritable traits necessarily exhibited a beneficial response to genomic prediction. For example, the highly heritable trait stature showed only a slight increase in the correlations of PA or of $GEBV_p$ to EBV_t of test bulls.

Differences in average EBV_t estimated for 101 test bulls between uni-trait and multi-trait analyses were slight for all of the linear type traits. Regarding PA, the largest differences of averages between uni-trait and multi-trait analyses occurred for udder traits. The resulting average $GEBV_p$ and $GEBV_t$ of young bulls were very similar. Considerably higher differences were noted between average EBV_t and $GEBV_p$ for young bulls. Average values of EBV_t for 101 young bulls are close to average of reference group of proven bulls, which indicate, that no direction but stabilizing selection is applied. The differences of average EBV_t and PA from the average EBV_t of the reference group of proven bulls were small. They fell within the range of -0.21 to 0.35 for uni-trait analyses and from -0.23 to 0.42 for multi-trait analyses. The largest differences between PA and EBV_t of proven bulls were for BCS (0.35; 0.27; 0.42; 0.23) and stature (-0.10; 0.26; -0.11; 0.25) for uni-trait and multi-trait analyses, respectively. Small discrepancies occurred for some udder traits: angularity, rear udder height, rear udder width, central ligament, teat length and rear teat position, either only for multi-trait analysis or just for uni-trait analysis. The differences of average $GEBV_t$ or $GEBV_p$ from the average EBV_t of proven bulls were higher than those of EBV_t or PA. They fell within the range from -0.78 to 1.02. The largest contrast in differences was again detected for BCS, stature and angularity and then for rear udder height, rear udder width, central ligament, teat length and rear teat position.

Multi-trait analyses had the advantage over uni-trait ones of providing compact and stable results for udder traits. Generally, the accuracy and stability of prediction will be higher because a multi-trait model combines information from different traits. Because of high correlations between EBV_t from uni-trait and multi-trait analyses, we consider the analyses to be essentially equivalent. Multi-trait analyses, however, provided stable results for all evaluated characters, especially for udder traits. For this reason, we prefer it to the uni-trait analysis. In comparison to our results Tsuruta *et al.* [2011] found R^2 in the wider range from 0.10 to 0.34 for uni-trait models and from 0.12 to 0.35 for multi-trait models. Particularly with regard to the upper limit, they found high values of R^2 for more traits than we found but some traits also showed very low R^2 . In marked contrast to our results, Tsuruta *et al.* [2011] reported an increase in R^2 by 16% for $GEBV$ in comparison to EBV . A possible explanation is higher number of genotyped bulls and larger data set than were used in the current study.

Relatively low correlations between EBV_t and PA or $GEBV_p$ (ranging from 0.03 to 0.57) can be explained partly by the small scale of the data set and partly by the nature of the recorded traits. For lactation milk yield, Přebyl *et al.* [2012] reported correlations around 0.5 for EBV and 0.52 or 0.54 for $GEBV$. They found increases

of 0.02 to 0.05 in the correlations due to the inclusion of genomic relationship information. We observed a higher increase for udder depth and rear udder height (0.11) than were reported by Přebyl *et al.* [2012] for milk yield. This increase is on the lower threshold of results from the literature [Hayes *et al.* 2009, Van Raden *et al.* 2009, Aquilar *et al.* 2010]. At the same time, our results showed a decrease for angularity, rear legs rear view and locomotion. Both, the reference population that was used by Přebyl *et al.* [2012] – 210 sires – and the reference population – 221 sires – used in our analysis were small. However, the differences between results of the two studies for linear type traits are more likely to be explained by differences in the genetic control of particular traits [Calus 2010]. The conformity of differences of average EBV_p , PA, $GEBV_t$ and $GEBV_p$ of young bulls from the EBV of proven bulls for some traits indicates good agreement between results in both datasets. The over-valuation of young animals resulting from PA or GEBV calculated in the data set was not manifested except for angularity, BCS, stature and some udder traits.

The results presented in this study show that the use of multi-trait single-step genomic evaluation is advantageous for the Czech Holstein population. The multi-trait evaluation gives less variable, more compact and stable breeding values than uni-trait evaluation, and single step genomic evaluation resulted in higher correlations between breeding values of test sires and their final breeding values based on genotypic information and phenotypes of daughters.

Acknowledgement. We thank to V. Čermák, Ing. Z. Růžička, Ing. J. Šplíchal and Ing. M. Verner from Czech Moravian Breeding Corporation, Prague, Czech Republic, and Doc. J. Motyčka and Ing. L. Vondrášek from Holstein Cattle Breeders Association of the Czech Republic, Prague, Czech Republic for preparation of data files. We are grateful to Dr. Per Madsen, University of Aarhus, Denmark, and to Prof. Ignacy Misztal, University of Georgia, USA, for methodical support and for provision of computer programs. The access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the programme “Projects of Large Infrastructure for Research, Development, and Innovations” (LM2010005) is highly acknowledged. Thanks are due to W.D. Hohenboken (Philomath, Oregon, USA) for editing the English of the paper and for valuable comments.

REFERENCES

1. AGUILAR I., MISZTAL I., JOHNSON D.L., LEGARRA A., TSURUTA S., LAWLOR T.J., 2010 – Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* 93, 743-752.
2. AGUILAR I., MISZTAL I., TSURUTA S., WIGGANS G.R., LAWLOR T.J., 2011 – Multiple trait genomic evaluation of conception rate in Holsteins. *Journal of Dairy Science* 94, 2621-2624.
3. CALUS M.P.L., 2010 – Genomic breeding value prediction: methods and procedures. *Animal* 4, 157-164.

4. CHRISTENSEN O.F., LUND M.S., 2010 – Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution* 42, 2.
5. HAYES B.J., BOWMAN P.J., CHAMBERLAIN A.J., GODDARD M.E., 2009 – Invited review: Genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science* 92, 433-443.
6. FORNI S., AGUILAR I., MISZTAL I., 2011 – Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genetics Selection Evolution* 43, 1.
7. MADSEN P., JENSEN J., 2010 – DMU – a package for analysing multivariate mixed models Version 6, release 50, 30 pp Retrieved on 10 January 2011, from <http://dmu.agrsci.dk>
8. MISZTAL I., LEGARRA A., AGUILAR I., 2009 – Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science* 92, 4648-4655.
9. MISZTAL I., LAWLOR T.J., SHORT T.H., VAN RADEN P.M., 1992 – Multiple-trait estimation of variance components of yield and type traits using an animal model. *Journal of Dairy Science* 75, 544-554.
10. NĚMCOVÁ E., ŠTÍPKOVÁ M., ZAVADILOVÁ L., 2011 – Genetic parameters for linear type traits in Holstein cattle. *Czech Journal of Animal Science* 56, 157-162.
11. PRYCE J.E., GREGLER B., BOLORMAA S., BOWMAN P.J., EGGER-DANNER C., FUERST C., EMMERLING R., SÖLKNER J., GODDARD M.E., HAYES B.J., 2011 – Short communication: Genomic selection using a multi-breed, across-country reference population. *Journal of Dairy Science* 94, 2625-2630.
12. PŘIBYL J., HAMAN J., KOTT T., PŘIBYLOVÁ J., ŠIMEČKOVÁ M., VOSTRÝ L., ZAVADILOVÁ L., ČERMÁK V., RŮŽIČKA Z., ŠPLÍCHAL J., VERNER M., MOTYČKA J., VONDRÁŠEK L., 2012 – Single-step prediction of genomic breeding value in a small dairy cattle population with strong import of foreign genes. *Czech Journal of Animal Science* 57, 151-159.
13. SAS, 2002 – Release 9.1 (TS1M3) of the SAS® System for Microsoft® Windows®. SAS Institute, Inc., Cary, USA.
14. SU G., MADSEN P., NIELSEN U.S., MANTYSAARI E.A., AAMAND G.P., CHRISTENSEN O.F., LUND M.S., 2012 – Genomic prediction for Nordic Red Cattle using one-step and selection index blending *Journal of Dairy Science* 95, 909-917.
15. TSURUTA S., MISZTAL I., AGUILAR I., LAWLOR T.J., 2011 – Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. *Journal of Dairy Science* 94, 4198-4204.
16. VAN RADEN P.M., VAN TASSELL C.P., WIGGANS G.R., SONSTEGARD T.S., SCHNABEL R.D., TAYLOR J.F., SCHENKEL F.S., 2009 – Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science* 92, 16-24.
17. VITEZICA Z.G., AGUILAR I., MISZTAL I., LEGARRA A., 2011 – Bias in genomic predictions for populations under selection. *Genetic Research* 93, 357-366.
18. ZAVADILOVÁ L., NĚMCOVÁ E., ŠTÍPKOVÁ M., 2011 – Effect of type traits on functional longevity of Czech Holstein cows estimated from a Cox proportional hazards model. *Journal of Dairy Science* 94, 4090-4099.
19. ZAVADILOVÁ L., ŠTÍPKOVÁ M., 2012 – Genetic correlations between longevity and conformation traits in the Czech Holstein population. *Czech Journal of Animal Science* 57, 125-136.
20. ZINK V., ŠTÍPKOVÁ M., LASSEN J., 2011 – Genetic parameters for female fertility, locomotion, body condition score, and linear type traits in Czech Holstein cattle. *Journal of Dairy Science* 94, 5176-5182.