On the use of elo rating on harness racing results in the genetic evaluation of trotter

C. Pieramati[1], L. Fusaioli[1], L. Scacco[1],
L. Buttazzoni[2], M. Silvestrelli[1]

[1] Centro di Studio del Cavallo Sportivo, Università di Perugia, Italy
[2] Associazione Nazionale Allevatori Suini, Italy

Corresponding author: Camillo Pieramati. Centro di Studio del Cavallo Sportivo, Università degli Studi di Perugina. Via San Costanzo 4, 06126 Perugia, Italy - Tel. +39 075 5857708 - Fax: +39 075 5857764 - Email: camillo.pieramati@unipg.it

ABSTRACT: The official results of trotters in Italian harness racings have been used to get AM-BLUP estimates of genetic parameters, EBVs and r_{TI} of three groups of traits: speed (racing time, annual best time, best time in career between 2- and 5-year old), earnings (earnings/start, annual earnings, total earnings between 2- and 5-year old) and Elo system traits (underlying performance and final rating). The Elo system has been used for half a century in chess players rating, and it has been modified and fitted to several games and sports: it has already been used for genetic evaluation of sport horses in France. The highest heritability estimates in each group of traits have been found for best time (.430±.014), total earnings (.271±.013) and Elo final rating (.270±.008). The choice of *"k"*, the Elo ratings updating factor, did not show a key role in affecting the results. The underlying performance heritability and repeatability have been estimated .159±.004 and .420±.007 respectively. The Elo-based systems proved to be very promising in objectively evaluating trotters.

Key words: Trotter, Heritability, Repeatability, Elo.

INTRODUCTION – The selection programme of the Italian Trotter Studbook is based on independent culling levels for phenotipic values of three traits: best time/km in career, overall earnings in career and top-race winning (UNIRE, 2005). Genetic evaluations of trotters are in use in several countries: they are usually based on one or more traits which refer to horse speed, to earnings or to rankings. Many of these traits are influenced by the level of the race: furthermore, earnings and rankings need mathematical transformations to become normally distributed, and several strategies have been used for this goal (Ricard *et al.*, 2000). Many games and sports faced the problem of calculating the relative value of the competitors and the rating system developed for chess by A. Elo has been widely adapted (Elo, 1978). This system assumes any performance to be a normally distributed random variable, so that ranking probability for each competitor can be estimated by the difference in ratings from the other ones. The aim of this project is to compare the evaluation on usual traits with an Elo system.

MATERIAL AND METHODS – The dataset of January 1st, 1992-December 31st, 2004 competition results consisted of 1,772,155 performance observations from 41,519 trotters, recorded in 179,820 races. There were 142,989 horses in the pedigree file, and 91.6% of them were from the Italian Studbook. The pedigre file was very complete and for some horse it traced back for 22 generations; therefore more than 91% of the animals had an inbreeding coefficient greater than 0, with mean .045.

Preliminary investigations by multivariate models suggested that best time in 1600-1725 m races and best time in 2000-2200 m races could be used as single trait in a model with a fixed effect for race length, and also that best time at different ages and annual earnings could be analized by repeatability models. Earnings have been normalized by logarithmic transformation and variances standardized within each (sex*year of birth) group (Silvestrelli *et al.*, 1995).

The underlying performances and the final ratings according to the Elo system have been calculated by means of two original ELO-TROT-K and ELO-TROT-PR programs, written in Fortran 95, setting the "classic" values of 2000 and 200 for performance mean and standard deviation, respectively; although a logistic distribution of the under-

lying performance is currently most used in modified Elo systems, we preferred to use a normal distribution. The underlying performance of the first and last trotter in each race, that theorically could be set to +∞ and -∞, has been assessed pretending the horse tied with a phantom. In order to choose a convenient *"k"* factor, that is used in the rating update in the Elo's measures and gives the relative weight of the last performance, the values in the range between .5 and 200 have been explored by a .5 step: the winner prediction rate and Kendall's ***T*** between predicted and observed ranking in each race have been used like indicators of fit.
Genetic parameters have been estimated by MTDFREML and VCE software: this is because a feature of the first package is to calculate EBVs accuracies, and the second program calculates the standard errors of variance components ratios (Boldman *et al.*, 1993; Kovac and Groeneveld, 2003). The estimates for time/km have been calculated by Gibbs Sampling method due to model size limits: in this case, accuracies were calculated by JAA 2.0 program (Misztal and Gianola, 1987); all other estimates were by REML; genetic parameters estimates by MTDFREML and VCE were the same for all traits: since the priors and the optimization method were different (polytope or analytical gradients), and there were only three or four parameters in the optimized function, the probability of a local maximum (minimum) is negligible.

RESULTS AND CONCLUSIONS – The genetic parameters estimates for speed and earnings traits are shown in Table 1. The estimates for time/km confirm previous results in the same population with a slight decrease in the ratio, probably due to the driver considered fixed instead of random (Pieramati *et al.*, 2003). For the speed, the estimates are close to values of German trotters, especially in the analysis when an individual races effect is used (Bugislaus *et al.*, 2005). As reported (Ricard *et al*., 2000), from these results it can be confirmed that estimates of speed traits have higher genetic components than earnings traits, and that career records are more heritable than annual or individual records. Pearson's and Sperman's correlations between speed traits EBVs ranged from 0.73 (time/km with best time in career) to 0.96% (annual best time with or without the driver effect); for earnings traits, correlations ranged between .14 (earnings/start with earnings in career when covariated with starts) and .89 (earnings/start with annual earnings when covariated with starts). The earnings in career covariated with starts was the most heritable earnings-based trait, but it was also the least correlated with the others.

Table 1. Estimates of the genetic parameters for speed and earnings traits.

Traits (records)	Fixed effects (levels)	h^2	r
time/km (383,111)	fixtures (21,333), sex (3), age (4), starting type (2), race length (2), driver (2198)	.254±.007	.499±.013
annual best time (68,895)	sex (3), age (4), starting type (2), race length (2), year (13)	.396±.009	.630±.016
	as the model above + driver (2013)	.249±.009	.532–.017
best time in career (20,917)	sex (3), year of birth (10), starting type (2), race length (2)	.430±.014	-
	as the model above + driver (1640)	.240±.014	-
earnings/starts (93,617)	.191±.007	.454±.013	
annual earnings (93,617)		.143±.007	.413±.012
	starts (covariate)	.195±.006	.315±.011
earnings in career (26,647)		.207±.012	-
	starts (covariate)	.270±.013	-

The ELO-TROT-K program was used to optimize the *"k"* factor. The best *"k"* to predict race winner has been found at 33.5: using this value, the percentage of prediction of the winner was 24.6%; the best *"k"* factor to predict the whole race classment was 18.0, and it was lower as expected. The use of this value reached 79.1% of positive Kendall's ***T*** between predicted and observed race results. The plots of *"k"*-correct winner prediction and *"k"*-***T*** showed a smoothed aspect, with a fast raise from 0.5 to the best *"k"*, and a little decrement after this optimal point. It must be noted that this optimal values are close to those used by the most part of Elo-based systems.
The ELO-TROT-PR program was used to get the value of trotter Performance in each race and final Ratings of all

starters. There were 8,586,365 paired comparisons in the competition dataset. In order to verify the effect of the *"k"* factor, seven equally spaced values near the best points were used in estimating genetic parameters both for Elo performance and for Elo final rating: 14.125, 18.0, 21.875, 25.750, 29.625, 33.5, 33.375. Variance ratios increased together with *"k"*: heritability ranged from .250±.008 to .270±.008 for final rating, while for performance heritability ranged from .107±.003 to .159±.004 and repeatability from .291±.005 to .420±.007. The choice of the updating factor did not show any important effect on the EBVs. Correlations were always greater than .99 between EBVs estimated using different *"k"*. Elo performance and rating EBVs were also correlated between .86 and .91. The Elo EBVs were correlated with speed EBVs between -0.57 (Elo performance - best time in career without driver effect) and -0.81 (Elo rating - annual best time without driver effect), and between .13 (Elo performance - earning in career covariated with starts) and .81 (Elo rating - earnings/starts) for earnings traits.
We conclude that a genetic evaluation on time/km can be recommended to the Italian Studbook; also, an Elo-based evaluation could soon replace the culling levels of earnings and winning.

The Authors want to thank UNIRE for providing official results.

REFERENCES – **Boldman**, K.G., Kriese, L.A., Van Vleck, L.D., Van Tassel, C.P., Kachman S.D., 1993. A Manual for use of MTDFREML. A Set of Programs to Obtain Estimates of Variances and Covariances. USDA Agricultural Research Service, Washington, D.C., USA. **Bugislaus**, A.E., Roehe, R., Kalm, E., 2005. Comparison of two different statistical models considering individual races or race tracks for evaluation of German Trotters. Livest. Prod. Sci. 92:69-76. **Elo**, A., 1978. The Rating of Chessplayers, Past and Present. Arco Pub., New York, USA. **Kovac**, M. and Groeneveld, E., 2003. VCE-5 User's Guide and Reference Manual, versione 5.1. Neustadt, Germany. **Misztal**, I., Gianola, D., 1987. Indirect Solution of Mixed Model Equation. J. Dairy Sci., 70:716-723. **Pieramati**, C., Silvestrelli, M., Buttazzoni, L., Verini Supplizi, A., 2003. Genetic Evaluation for Racing Speed in Italian Trotters. 54th Annual Meeting of the EAAP, Rome, Italy. **Ricard**, A., Bruns, E., Cunningham, E.P., 2000. Genetics of Performance Traits. In: A.T. Bowling and A. Ruvinsky (eds.) The Genetics of the Horse. CABI Publishing, New York, USA, pp. 411-438. **Silvestrelli**, M., Pieramati, C., Cavallucci, C., Bonanzinga, M., 1995. The Current Breeding Plans for Saddle Horse, Trotter and Thoroughbred in Italy. 46th Annual Meeting of the EAAP, Prague, Czech Republic. **UNIRE**, 2005. Disciplinare e norme tecniche del libro genealogico del cavallo trottatore italiano. Rome, Italy.