Definition of subgroups for fixed regression in the test-day animal model for milk production of Holstein cattle in the Czech Republic

L. Zavadilová, E. Němcová, J. Přibyl, J. Wolf

Research Institute of Animal Production in Prague, Czech Republic

ABSTRACT: The investigation was based on roughly 3.9, 2.7 and 1.7 million test-day records from first, second and third lactation, respectively, sampled from 596 200 Czech Holstein cows between the years 1991 and 2002. Breeding values were estimated from multi-lactation random-regression test-day models which contained the fixed effect of herd-test day, fixed regression on days in milk and random regressions on the animal level and the permanent environmental effect. Third degree Legendre polynomials (with four coefficients) were used for both the fixed and random regressions. The models differed in fixed regression. In Analysis I, 96 subclasses were defined according to age at calving, season and year of calving within lactation. In Analysis II, days open were additionally included as a grouping factor resulting in 480 subclasses. Rank correlations over 0.98 between both analyses were observed for breeding values for sires. Grouping according to Analysis I was recommended.

Keywords: test-day model with random regression; fixed regression subgroups; lactation curves; milk yield; dairy cattle

Genetic evaluation models in dairy cattle that use test-day records instead of 305-day lactation records are of great interest in the area of cattle breeding for production traits (Reents *et al.*, 1998; Emmerling *et al.*, 2000b; Lidauer *et al.*, 2000; Schaeffer *et al.*, 2000). The models for production traits are known as test-day models (TDM) and can account for the effects of test day, number of test-days per lactation and intervals between test-day records, and provide information about persistency by modelling the pattern of the lactation curve. Different types of TDM are described in the literature (Swalve, 2000).

The random regression (RR) TDM (Schaeffer and Deckers, 1994) can account for many environmental factors that could affect cows differently during the lactation. The lactation curve is split into two parts: a fixed part (average lactation curve) and a random animal specific part (deviation from the average curve). For the correction on the lactation stage, an appropriate sub-model is accounted for, nested within sub-groups. Different strategies are proposed for the definition of subgroups for fixed regression on the stage of lactation.

Frequently used factors are season of calving and/or classes of age at calving (Candrák *et al.*, 1997; Reents *et al.*, 1998; Strabel and Misztal, 1999; Emmerling *et al.*, 2000b; Lidauer *et al.*, 2000; Schaeffer *et al.*, 2000). In addition, the Canadian estimation procedure involved classes according to region and time period (Schaeffer *et al.*, 2000). Reproduction traits (e.g. days open, calving interval) are completely ignored there. In Finland the variable "days carried calf" is included in the model (Lidauer *et al.*, 2000). Reents *et al.* (1998) reported a model where calving interval is employed. Candrák *et al.* (1997) involved the breed effect.

The aim of this paper was to define subgroups for fixed regression according to age at calving, year and season of calving and days open in random regression test-day models and to test the influence

Supported by the National Agency for Agricultural Research of the Ministry of Agriculture of the Czech Republic (Project QC 1235).

of differently defined subgroups on the predicted breeding values (BV).

MATERIAL AND METHODS

Test-day data for milk production of the Czech Holstein breed were extracted from the Czech national milk-recording database. The data included roughly 3.9, 2.7 and 1.7 million records for first, second and third lactation, respectively, sampled from 596 200 cows between the years 1991 and 2002 (Table 1).

Random regression test-day models were applied for the prediction of BVs and the estimation of lactation curves for fixed-regression subgroups. The total number of animals in the pedigree was 1 210 278. Data were analysed by a multiple-lactation model in which TD yields in the first, second and third lactations were considered as different traits. The model was assumed to be the same for each lactation.

In matrix notation, the multiple lactation random regression TD model can be written as

$$y = Hc + Xb + Wp + Za + e$$

where: y = the vector of observations in 3 lactations

c = the vector of fixed contemporary groups effects defined as herd-test-day date-lactation subclasses

b = the vector of fixed regression coefficients nested within particular subclasses, see Analysis I and II

p = the vector of random regression coefficients for animal permanent environmental (PE) effects

a = the vector of random regression coefficients for animal additive genetic effects

H = the incidence matrix that relates contemporary groups to observations

X, W, Z = matrices of covariances involving number of days in milk associated with a cow on a given test date and correspondingto the observations.

The expectations and covariance matrices are:

$$E\begin{pmatrix} y \\ p \\ a \\ e \end{pmatrix} = \begin{pmatrix} Hc + Xb \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

and

$$Var\begin{pmatrix} p \\ a \\ e \end{pmatrix} = \begin{pmatrix} P & 0 & 0 \\ 0 & G & 0 \\ 0 & 0 & R \end{pmatrix}$$

where

$$P = I \otimes P_0,\ G = A \otimes G_0,\ R = \Sigma^+ Rij$$

where: A = the additive genetic relationship matrix

P_0, G_0 = covariance matrices for the PE and genetic regression coefficients, respectively

R_{ij} = the covariance matrix for cow i on a given test day j

The corresponding mixed model equations (MME) for this model are

$$\begin{pmatrix} H'R^{-1}H & H'R^{-1}X & H'R^{-1}W & H'R^{-1}Z \\ X'R^{-1}H & X'R^{-1}X & X'R^{-1}W & X'R^{-1}Z \\ W'R^{-1}H & W'R^{-1}X & W'R^{-1}W + P^{-1} & W'R^{-1}Z \\ Z'R^{-1}H & Z'R^{-1}X & Z'R^{-1}W & Z'R^{-1}Z + G^{-1} \end{pmatrix} \begin{pmatrix} \hat{c} \\ \hat{b} \\ \hat{p} \\ \hat{a} \end{pmatrix} = \begin{pmatrix} H'R^{-1}y \\ X'R^{-1}y \\ W'R^{-1}y \\ Z'R^{-1}y \end{pmatrix}$$

The model applied for analyses was:

$$y_{nkitjl} = HTD_{ni} + \sum_{m=1}^{4} \beta_{nkm} z_{tm} + \sum_{m=1}^{4} a_{njm} z_{tm} + \sum_{m=1}^{4} p_{njm} z_{tm} + e_{nkitjl}$$

where: y_{nkitjl} = the record l on cow j belonging to subclass k for fixed regression, in t days in milk DIM) within lactation n measured within herd-test day class i

HTD_{ni} = the fixed effect of herd-test day i in lactation n

β_{nkm} = fixed regression coefficients specific to subclass k in lactation n

a_{njm} = random regression coefficients specific to animal j in parity n

p_{njm} = random regression coefficients specific to the permanent environmental (PE) effect of cow j in lactation n

e_{nkitjl} = the residual effect for the given observation

z_{tm} ($m = 1, \ldots, 4$) are covariates associated with DIM, assumed to be the same for both fixed and random regressions

Third degree Legendre polynomials (with four coefficients) were used for both the fixed and random regressions on the scale from 6 to 305 DIM. Let $\mathbf{p}_\mathbf{j}$ represent the vector of 12 random permanent environmental regression coefficients (4 coefficients in each of the 3 lactations) for cow j with

Table 1. Summary of data

	Lactation I			Lactation II			Lactation III		
	No.	Mean	SD	No.	Mean	SD	No.	Mean	SD
Milk (kg)	3 977 875	18.8	6.50	2 661 978	21.0	8.40	1 712 695	21.20	8.68
Days in milk	3 977 875	147.6	82.22	2 661 978	145.9	81.90	1 712 695	145.4	81.69
Days open	3 304 399	136.9	82.71	2 220 076	132.5	76.72	1 404 415	131.6	75.46
Age at calving (days)	3 794 681	840.5	91.07	2 525 028	1 256.6	123.37	1 611 228	1 660.0	148.66

SD = standard deviation

the covariance matrix **P**. The PE covariance matrix for all cows is then **I**⊗**P**. For $\mathbf{a}_j$, the vector of the 12 random regression coefficients for animal *j*, the covariance matrix is **G. A**⊗**G** is then the genetic covariance matrix for all animals with **A** being the additive relationship matrix. Different residual variances were allowed for different lactations and time periods within lactation, defined as 7 to 45 DIM, 46 to 115 DIM, 116 to 265 DIM and 266 to 305 DIM (Jamrozik *et al.*, 1998). Residual effects on different DIM were assumed to be uncorrelated both within and between cows. The used model was a special case of the model presented by Jamrozik *et al.* (1998) for the multiple-lactation, multiple-trait situation. MT-AM, programme BLUP90IOD (Tsuruta *et al.*, 2001) was used for the solution of mixed model equations.

The used models differed only in the definition of the subclasses for fixed regression. The subclasses were defined in two ways:

Analysis I. The subclasses were defined according to the age at calving, season and year of calving. This gave 96 subgroups of fixed regression within lactation.

Analysis II. Days open were included in the definition of subclasses beyond the factors used in Analysis I. In addition, cows were assigned to one of five levels for days open within lactation. The number of subclasses therefore increased to 480 subgroups within lactation.

For the definition of the levels for the fixed regression subgroups see Table 2. The (co)variance parameters were estimated by Gibbs sampling on a sample of data (Dědková *et al.*, 2002).

Table 2. Definition of levels for fixed regression subgroups

	Lactation I	Lactation II	Lactation III
Age at calving (in days)	500–774	800–1 149	1 100–1 549
	775–899	1 150–1 349	1 550–1 749
	900–1 100	1 350–1 650	1 750–2 200
	and missing	and missing	and missing
	All lactations		
Season of calving (months)	Feb, Mar, Apr	Aug, Sep, Oct	
	May, Jun, Jul	Nov, Dec, Jan	
Year of calving	1994; 1996; 1998; 2000		
	1995; 1997; 1999; 2001 and 2002		
Days open	28–59	over 120	
	60–89	missing	
	90–119		

RESULTS AND DISCUSSION

Fixed regression

In both types of analyses, the estimated lactation curves differed between the defined subgroups in a similar way. The additional inclusion of days open in Analysis II did not influence the differences between lactation curves in particular subgroups due to the other effects.

Figures 1 to 5 demonstrate examples of the influence of lactation, calving season, age at calving, year of calving and days open on the regression functions. Only results for the first lactation (except demonstrating the difference between lactations) and for Analysis II were considered, where levels for days open were included in the definition of subgroups for the fixed regression.

As can be seen in Figure 1, the milk yield in the second and third lactations was higher than in the first one. The shape of the lactation curve for subgroups was essentially the same in the second and third lactations. With regard to the first lactation the decline of the curve after the peak was steeper in later lactations.

Seasonal differences in the shape of the lactation curves were apparent. Lactation curves for the calving seasons (Figure 2) differed in the shape especially at the beginning of lactation. The lactation curves for the first and second seasons of calving (Nov, Dec, Jan–Feb, Mar, Apr) were similar likewise those for the third and fourth seasons (May, Jun, Jul–Aug, Sep, Oct). This is partly in accordance with previous results of Dědková and Němcová (2003); they detected a different shape of phenotypic lactation curves for cows that calved from December to May and those that calved from July to October. Furthermore, in line with the present results, they found that the lactation yield was lowest for the cows calved in June and July and highest for those that calved from October to December.

Tekerli *et al.* (2000) reported the same grouping of calving months into seasons of calving because of differences in persistency and in the peak of the lactation. Kučera *et al.* (1999) suggested the maximum production of milk for cows that calved from December to March and the lowest production for those that calved from June to September. A similar development in milk production between the months of calving was reported by Strandberg and Lundberg (1991).

With respect to the similarity of the shape of the lactation curves and of the lactation yield, a restriction to only two seasons of calving, i.e. winter and summer period, could be suggested.

For the age at calving (Figure 3), the lactation curves differed more in the level of the yield than in the shape. A higher milk production was observed for the cows with a higher age at calving. This is in accordance with previous findings (Dědková and Němcová, 2003). In an RR test-day model, similar differences in the lactation curves between fixed regression subgroups according to age at calving were suggested by Jamrozik *et al.* (1997).

The estimated lactation curves according to year of calving (Figure 4) showed a lower variability than the appropriate phenotypic lactation curves from uncorrected data (Dědková and Němcová, 2003). The

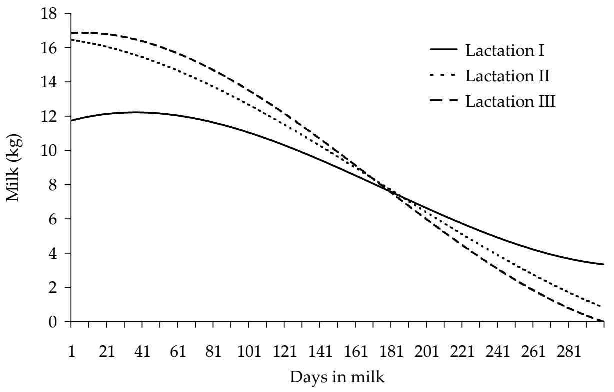

Figure 1. Example of predicted lactation curves in the first three lactations (second level of age at calving – see Table 2; season of calving: May to July; year of calving: 1998; days open: 90–119 days)

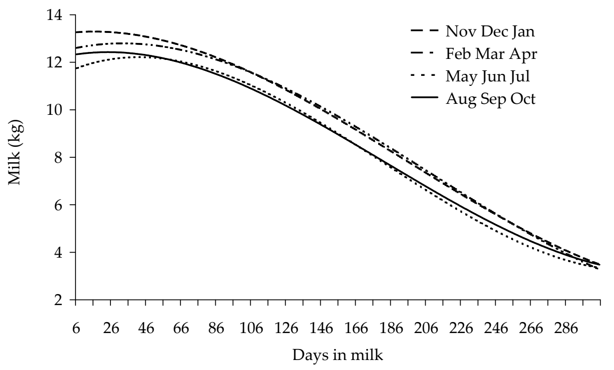

Figure 2. Example of predicted lactation curves for season of calving (first lactation; second level of age at calving: 775–899 days; year of calving: 1998; days open: 90–119 days)

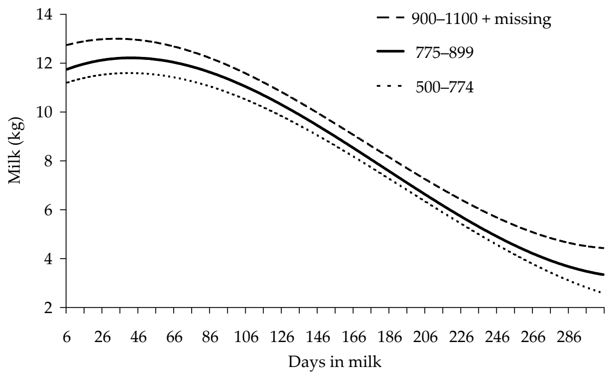

Figure 3. Example of predicted lactation curves for age at calving (first lactation; season of calving: May to July; year of calving: 1998; days open: 90–119 days)

increase in the peak of milk production was not very apparent in comparison with the phenotypic lactation curves. Furthermore, the decline after the peak was steeper for lactation curves with a higher peak in phenotypic lactation curves. However, the involvement of the time period (Schaeffer *et al.*, 2000) or calving year (Lidauer *et al.*, 2000) into the definition of subgroups seems to be necessary because of changes in management.

The shape of lactation curves estimated by RR TDM for different levels of days open (Figure 5) resembled the phenotypic lactation curves for these groups (Dědková and Němcová, 2003). The shorter the period of days open, the steeper the decline of lactation curves. A low yield was found for subgroups of cows without days open recorded. This is probably due to selection for milk production because milk production is influenced by pregnancy. This positive correlation between days open and yield variables is the reason why Emmerling *et al.* (2000a) did not recommend the inclusion of days open in the model, but they proposed instead to integrate a third degree polynomial of "days carried calf" as suggested by Lidauer *et al.* (2000).

Correlations between BVs

The correlations between BVs for the first, second and third lactations are shown in Table 3. They were essentially the same within both types of analyses. High values occurred between successive lactations. The highest correlations were found between the second and third lactations.

The rank correlations between both analyses (Table 4) were over 0.98 for all predicted BVs and over 0.99 if only the BVs for sires were considered. In Figure 6, the scatter plot between BVs of sires predicted in Analysis I and Analysis II is shown. The listings of the top 100 bulls with more than 10 daughters showed minimal re-ranking (maximal re-ranking 33, average re-ranking for BVs in the first lactation: 1.69). It appears that the inclusion of days open in the definition of subgroups for the correction on the lactation stage by fixed regression did not markedly influence the ranking of breeding values. For that reason the grouping used in Analysis I should be preferred, as in Analysis I the number of observations in the individual classes will be considerably higher which will yield more precise estimates of the regression coefficients. The use of a more detailed grouping would be justified only if great changes in the ranking of animals are expected.

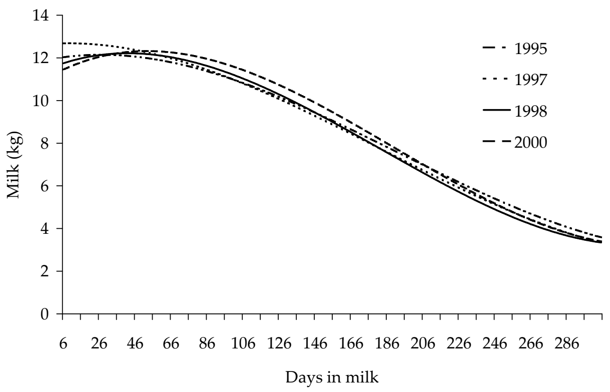

Figure 4. Example of predicted lactation curves for years of calving (first lactation; second level of age at calving: 775–899 days; season of calving: May to July; days open: 90–119 days)

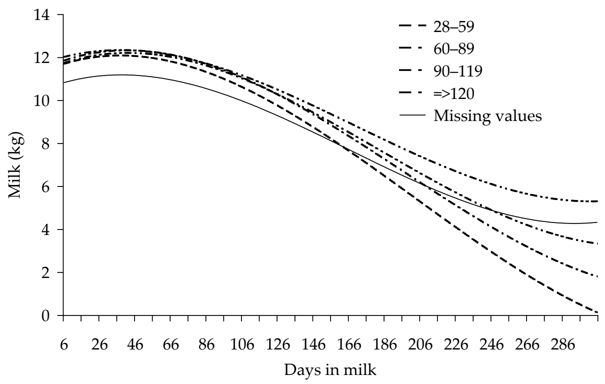

Figure 5. Example of predicted lactation curves for levels of days open (first lactation; second level of age at calving: 775–899 days; year of calving: 1998, season of calving: May to July)

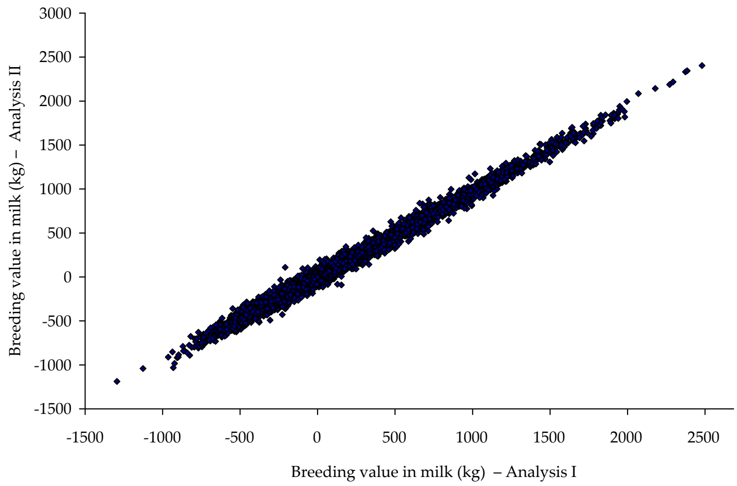

Figure 6. Relationship between the average breeding values of sires predicted in Analyses I and II

Table 3. Spearman rank (above diagonal) and Pearson correlations (below diagonal) between estimated breeding values for Analyses I and II

Lactation	Analysis I			Analysis II		
	lactation I	lactation II	lactation III	lactation I	lactation II	lactation III
I	1.00	0.70	0.60	1.00	0.70	0.59
II	0.78	1.00	0.83	0.77	1.00	0.82
III	0.66	0.88	1.00	0.66	0.87	1.00

Table 4. Spearman rank and Pearson correlations between both analyses for predicted breeding values of all animals and sires only

	All animals			Sires		
	lactation I	lactation II	lactation III	lactation I	lactation II	lactation III
Spearman	0.985	0.987	0.993	0.990	0.992	0.993
Pearson	0.990	0.990	0.990	0.995	0.997	0.998

CONCLUSIONS

The shape of the lactation curve of Holstein cattle is influenced by all analysed factors, i.e. season and year of calving, days open and age at calving. The assignment of month of calving into four seasons of calving seems to be a good solution because there are differences between them in the shape as well as in the production. Involving days open classes into the test-day model is not recommended.

Acknowledgement

Thanks are due to Prof. I. Misztal (University of Georgia, USA) and Dr. J. Jamrozik (CGIL, University of Guelph, Canada) for making available program packages.

REFERENCES

Candrák J., Kadlečík O., Schaeffer L.R. (1997): The use of test-day model for Slovak cattle populations. In: Proc. 48th Annual Meeting of the European Association for Animal Production, Vienna, Austria, August 25–28.

Dědková L., Jamrozik J., Schaeffer L.R. (2002): Genetic parameters for test-day model with random regressions for production traits of Czech Holstein cattle. In: XX. Genetické dny, 12.–13. 9. 2002, Brno. 117–121.

Dědková L., Němcová E. (2003): Factors affecting the shape of lactation curves of Holstein cows in the Czech Republic. Czech J. Anim. Sci., *48*, 395–402.

Emmerling R., Dempfle L., Götz K.-U. (2000a): Validation of fixed effects in a test day model. In: Proc. 51st Annual Meeting of the European Association for Animal Production, August 21–24 2002, Den Haag, The Netherlands.

Emmerling R., Lidauer M., Mäntisaari E.A. (2000b): Multiple lactation random regression test-day model for Simmental and Brown Swiss in Germany and Austria. Interbull Bulletin, No. 29, 111–115.

Jamrozik J., Schaeffer L.R., Deckers J.C.M. (1997): Genetic evaluation of dairy cattle using test yields and random regression model. J. Dairy Sci., *80*, 1217–1226.

Jamrozik J., Schaeffer L.R., Grignola F. (1998): Genetic parameters for production traits and somatic cell score of Canadian Holsteins with multiple trait random regression model. In: Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, 303–306.

Kučera J., Hyánek J., Mikšík J., Čermák V. (1999): Vliv období otelení na mléčnou užitkovost dojnic českého strakatého skotu. Czech J. Anim. Sci., *44*, 343–350.

Lidauer M., Mäntisaari E.A., Strandèn I., Pösö J. (2000): Multiple-trait random regression test-day model for all lactations. In: Proc. Interbull Meeting, Vienna, Austria, Int. Bull Evaluation Service, Bulletin No. 16, 75–79.

Reents R., Dopp L., Reinhardt F., Schmutz M. (1998): Beschreibung des Zuchtwertschätzverfahrens für Milchleistung (Testtagstiermodell). In: Aktueller Stand und neue Entwicklungen der Zuchtwertschätzung beim Rind. DGfZ – Schriftenreihe, Heft 11, 16–33.

Schaeffer L.R., Deckers J.C.M. (1994): Random regressions in animal models for test-day production in dairy cattle. In: Proc. 5th World Congr. Genet. Appl. Livest. Prod., Guelph, Canada, 443–446.

Schaeffer L.R., Jamrozik J., Kistemaker G., VanDoormaal B. (2000): Experience with a test-day model. J. Dairy Sci., *83,* 1135–1144.

Strabel T., Misztal I. (1999): Genetic parameters for first and second lactation milk yields of Polish Black and White cattle with random regression test-day models. J. Dairy Sci., *82*, 2805–2810.

Strandberg E., Lundberg C. (1991): A note on the estimation of environmental effects on lactation curves, Anim. Prod., *53*, 399–402.

Swalve H.H. (2000): Theoretical basis and computational methods for different test-day genetic evaluation methods, J. Dairy Sci., *83,* 1115–1124.

Tekerli M., Akinci Z., Dogan I., Akcan A. (2000): Factors affecting the shape of lactation curves of Holstein cows from the Balikesir province of Turkey. J. Dairy Sci., *83*, 1381–1386.

Tsuruta S., Misztal I., Strandén I. (2001): Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. J. Anim. Sci., *79,* 1116–1172. http://nce.ads.uga.edu/~ignacy

Received: 04–05–04
Accepted after corrections: 04–10–18

Corresponding Author

Ing. Ludmila Zavadilová, CSc., Research Institute of Animal Production, P.O.Box 1, 104 01 Prague-Uhříněves, Česká republika
Tel. +420 267 009 574, fax +420 267 710 779, e-mail: dedkova.ludmila@vuzv.cz