SOCIOECONOMIC DETERMINANTS OF MENARCHE IN RURAL POLISH GIRLS USING THE DECISION TREES METHOD

STANISŁAW MATUSIK*, TERESA ŁASKA-MIERZEJEWSKA† AND MARIA CHRZANOWSKA‡

Statistics and Computer Science Division, Academy of Physical Education, Cracow, Poland, †'Halina Konopacka' Higher School of Physical Culture and Tourism, Pruszków, Poland and ‡Department of Anthropology, Academy of Physical Education, Cracow, Poland

Summary. The aim of this study was to assess the usefulness of the decision trees method as a research method of multidimensional associations between menarche and socioeconomic variables. The article is based on data collected from the rural area of Choszczno in the West Pomerania district of Poland between 1987 and 2001. Girls were asked about the appearance of first menstruation (a yes/no method). The average menarchal age was estimated by the probit analysis method, using second grade polynomials. The socioeconomic status of the girls' families was determined using five qualitative variables: fathers' and mothers' educational level, source of income, household appliances and the number of children in a family. For classification based on five socioeconomic variables, one of the most effective algorithms CART (Classification and Regression Trees) was used. In 2001 the menarchal age in 66% of examined girls was properly classified, while a higher efficiency of 70% was obtained for girls examined in 1987. The decision trees method enabled the definition of the hierarchy of socioeconomic variables influencing girls' biological development level. The strongest discriminatory power was attributed to the number of children in a family, and the mother's and then father's educational level. Using this method it is possible to detect differences in strength of socioeconomic variables associated with girls' pubescence before 1987 and after 2001 during the transformation of the economic and political systems in Poland. However, the decision trees method is infrequently applied in social sciences and constitutes a novelty; this article proves its usefulness in examining relations between biological processes and a population's living conditions.

Introduction

Menarche is a complex event affected, among others, by genetic, hormonal, fetal, early childhood and environmental factors. In the current study these factors were

held constant in order to observe the predictive power of the CART (Classification and Regression Trees) method to determine menarche from socioeconomic factors.

One of the best biological symptoms of socioeconomic stratification of the population is the age of menarche. At the same time, it presents a high level of eco-sensitivity, reflecting even small differences in the living conditions of research population groups (Belsky *et al.*, 1991; Moffitt *et al.*, 1992; Rimpelä & Rimpelä, 1993; Adair, 2001; Ellis & Essex, 2007). Girls from families of higher socioeconomic status are usually characterized by an earlier menarche age in comparison with those brought up in inferior conditions (Bielicki *et al.*, 1986; Hulanicka *et al.*, 1990; Kozieł & Jankowska, 2002; Łaska-Mierzejewska & Olszewska, 2007). This relationship considers arithmetic means of menarche age in the examined groups; while individual girls' results may differ from the rule, the exceptions are ubiquitous.

Studies of the biological effects of social stratification of the rural population in Poland were initiated by Łaska-Mierzejewska in 1967 and repeated again in 1977, 1987 and 2001. Where possible, the research was conducted in the same schools each time (Łaska-Mierzejewska, 1970, 1995; Łaska-Mierzejewska *et al.*, 1982; Łaska-Mierzejewska & Olszewska, 2006).

The present study is based on material concerning one of the areas examined in 1987 and 2001. Under communist rule, the area of Choszczno, located in the West Pomerania district, was distinguished by the highest concentration of state farms in Poland; state-owned farmlands exceeded 50% of the area. After the political and economical transformation of 1989, the state farms were dissolved in 1992, which generated high unemployment among country dwellers who did not own land; in West Pomerania district unemployment amounted to 27% and was the highest in Poland.

The research results from 1987 clearly showed an influence of the economic crisis taking place in Poland between 1978, when food rationing was introduced, and 1989, when it was terminated. The results from 2001 presented an influence of political and economical transformation on the biological condition of the rural population. Between the two studies the educational level of the girls' parents changed significantly and positively, and the percentage of farms with an area of more than 16 hectares increased from 39% in 1987 to 52% in 2001. Household appliance ownership increased markedly over the period: 7% of the families owned a colour TV, automatic washing machine, freezer and car in 1987 as compared with 54% in 2001 (Łaska-Mierzejewska & Olszewska, 2006).

The aim of this study was to assess the usefulness of the decision trees method as a way of studying multidimensional associations between age at menarche and a set of socio-economic variables. For this reason the analysis included material formerly researched considering such dependencies when each of the variables was treated separately and independently from the other features (Łaska-Mierzejewska & Olszewska, 2006), and allowed qualitative assessment of the decision trees method applied in the study.

Methods

Subjects

Data were from girls aged 9–18 living in rural areas near Choszczno in the West Pomerania district of Poland, and originating from families of different income structure:

- farming families receiving income from their farms (18% of examined families);
- farming families also working elsewhere (21% of examined families);
- families with no arable land (61% of examined families).

Parents' educational level was registered as well as the number of children in a family, household appliance ownership level and amount of land owned.

The girls were asked about their first menstruation (a yes/no method) and the average menarche age was estimated by the probit method. A total of 2354 girls were examined in 2001 and a subgroup of 608 girls was selected from them on the basis of average menarche +/- half of standard deviation. Girls' data from the same area collected in 1987 were analysed to conduct a comparison on the basis of an analogically selected subgroup of 180 persons out of 828 examined.

Variables

Socioeconomic status was determined by five variables:

- fathers' educational level (defined in four categories (1, 2, 3, and 4): primary, basic vocational, secondary and higher);
- mothers' educational level (defined in four categories as above);
- source of income (three categories (1, 2, and 3): farmers, farmers also working elsewhere and non-farmers);
- number of children in a family (1–14 in 2001 and 1–12 in 1987);
- ownership of household appliances: a water supply system, hot water, gas, freezer, colour TV, washing machine, car and fridge in 1997, and video player in 2001 (four categories: the sum in the range 7–8 of the mentioned items was the best category '1'; 5–6 items was category '2'; 3–4 items was category '3'; and the sum not exceeding 2 appliances was category '4').

These five variables were treated as independent variables affecting the dichotomous dependent variable, which was menstruation occurrence (marked 'T' – true) or its absence (consistently marked as italic letter '*F*' – false).

The average age of menarche was estimated by the probit analysis method, using second grade polynomials (Bliss, 1934). Probability p of menarchal age occurrence (dichotomous dependent variable) in a given age range can be determined empirically in a classical way: n is the number of positive answers ('T' – true); $N-n$ is the number of negative answers ('*F*' – false). It is assumed that the analysed phenomenon has a normal distribution dependent on a known parameter t, which is the age of examined girls. It can then be presented as a function dependent on a linear combination of variable t:

$$p = n/N = F^{-1}(b_0 + b_1 t + b_2 t_2), \tag{1}$$

where F^{-1} denotes a reverse function to a normal distribution function ($F^{-1}(z)$ is $100z$-quantile of normal distribution). Values of p belonging to the range [0, 1], independently from values of t are shaped by a linear function (in this case of two variables: t and t^2), whereas unknown parameters b_i are estimated by the maximum likelihood method. The interpretation of parameters b_i ($i=1,2$) is additive, and may express how the probability of menstruation occurrence will change after increasing the age of girls by one year.

Due to the form of distribution, it is obvious that the mean value μ equals the median and it is also possible to determine standard deviation σ for variable t. The probit method is recommended as more suitable for describing biological phenomena than logistic regression (Finney, 1971).

The CART decision trees method

In order to define relations between input socioeconomic variables and the dichotomous variable, the decision trees method was applied (Gatnar, 2001; Larose, 2005). This is an exploratory method used in expert systems, knowledge bases, artificial intelligence and data mining. It is based on the recurrent split of multidimensional space into disjunctive subsets until they reach homogeneity in the field of the distinguished dependent variable. Then, for each subset, there is a local model built based on the relationships of independent input variables. It is worth mentioning that the diagnostic variable (independent) can be both an ordinal variable and a categorical variable, and that is exactly what happened in this work. The decision trees method, unlike discriminatory analysis, does not require meeting relatively strong assumptions, like distributions multinormality of variables and covariance matrix homogeneity in particular subgroups. In the conducted analyses, the division quality measures were selected empirically. Most commonly used heterogeneity measures are Gini's inequality coefficients, G-squared or a value based on Chi-squared. (E. Gatnar presents fifteen different measures of subset division quality; see Gatnar (2001), pp. 31–49.)

The family of decision trees methods are non-parametric methods, i.e. they do not assume either knowledge of distribution shape or kinds of relations between features. A significant advantage of using decision trees is their hierarchical structure and flexibility. In a tree of the binary type, there are two branches from each knot: the leaves denote classes (objects sets), whereas branches are described by simple functions representing features. They were the basis for the division relation: $x<C$, where C is calculated based on the used inhomogeneous measure by the value of discriminatory diagnostic feature. Fulfilling the conditions in a set order, from the tree root (located at the top of the graph) to the leaf, is a way of enabling the following and examination of dependencies between the object to a given concentration and the values of input variables.

The benefit of applying the decision trees method is the graphic and intuitively interpretable way of presenting classification rules, even for complicated models (Matusik, 2005, 2007). Another advantage is an automatic feature ranking list, which is helpful for estimating the statistical associations of classification variables with the examined phenomenon, and assessing their discriminatory ability (discriminatory power) on a scale of 0 to 100.

One of the most effective algorithms, CART (Classification and Regression Trees), was presented by Breiman *et al.* (1984). It relies on considering all combinations of the diagnostic variables in order to find the best division. It is done recurrently in the N-dimensional objects space (i.e. for 2001 the size of the examined girls' subgroup is $N=608$, and $N=180$ for 1987), creating disjunctive subsets until reaching their homogeneity in the field of the distinguished classification variable. (There can be

Table 1. Basic statistical parameters of examined girls, Choszczno, Poland, 1987 and 2001

Parameter	2001	1987
Number of examined girls	2354	828
Average menarche age (*status quo* method)	13.13	13.41
Standard deviation	1.95	1.15
Size of the analysed sample	608	180
Age – minimum in the sample	12.14	12.83
Age – maximum in the sample	14.17	13.98
Menstruation has appeared	309	96
Menstruation has not appeared	299	84

Source: authors' own study.

observed analogies to classical taxonomic methods, especially to the *K*-means method.)

The CART method and exhaustive search for *K*-unidimensional splits with Gini's heterogeneity measure *G* were used to build the models:

$$G(p_1, p_2, ..., p_K) = \sum_{i,j=1, i \neq j}^{K} p_i p_j, \quad \sum_{i=1}^{K} p_i = 1, \tag{2}$$

where $p_i = n_i/N$ is an empirical probability of a feature appearance in an *i*-tuple subset ($i=1,...,K$). Gini's measure (Gini, 1921), while assessing the change of non-homogeneity level, accepts the value of zero when a given knot is homogeneous so it prefers variables dividing the analysed set into subsets that are significantly different considering the examined dependent variable. Division quality measures and options of algorithm stop were selected empirically to obtain the highest possible classification correctness (due to the accepted two levels of dependent variable: menstruation appears or does not appear), while simultaneously keeping transparency of the created model.

Results

In 2001 the examined group of girls was characterised by an average menarche age of 13.13 years (SD=1.95). A distinguished subgroup (by the method of average menarche age +/- half of standard deviation) included 608 girls aged from 12 years and 2 months to 14 years and 2 months. For 1987 there were 180 girls aged from 12 years and 2 months to 14 years and the average menarche age was higher at 13.41 years accompanied by lower age dispersion (Table 1). It can be noticed that in the selected samples the numbers of girls who had menstruation and those who did not are almost the same (51% and 49% in 2001, and 53% and 47% in 1987, respectively).
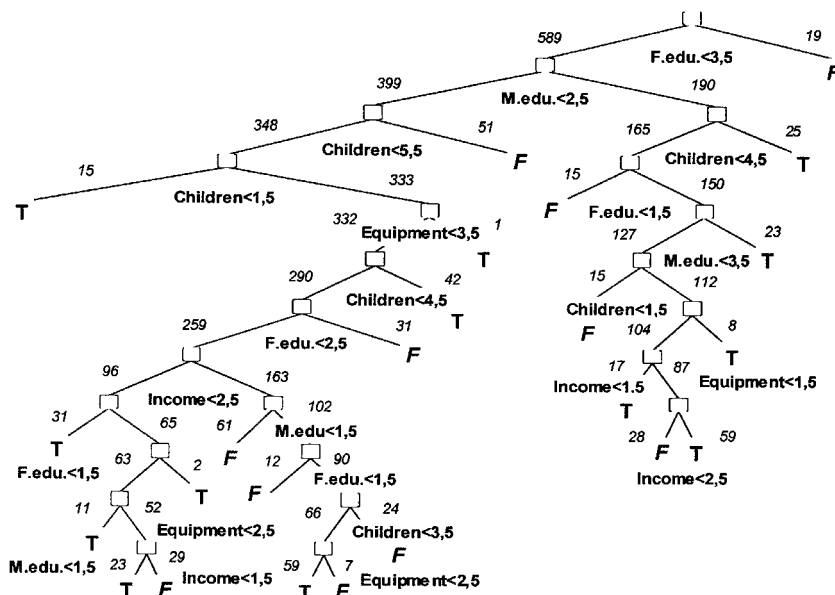
Fig. 1. Classification tree of menarche appearance (T, true; or F, false) in girls aged 12–14 from the area of Choszczno (2001 data). F.edu., father's education status; M.edu., mother's education status; Children, number of children in family; Equipment, household appliance ownership; Income, income source. Source: authors' own study.

A relational model of menarche appearance in the form of a decisional tree obtained by the CART method is presented in Fig. 1. The tree of binary divisions is quite extended and has 24 final leaves marking menarche appearance: 'T' (true) or 'F' (false). Above each branch there are numbers resulting from a formerly made division of the subset onto two parts and the relation defining the division. Fulfilling the relation means passing on to a further division on the diagram's left side. On the highest level, at the tree root, which is an input data set for 608 girls, a decisive role was played by divisions based on parents' educational level and the number of children in a family. For example, if the father's educational level was secondary at most (category no more than '3'), the mother's educational level was at least secondary (category no less than '3') and the number of children totalled five or more, in 25 girls out of 190 at the time the study was conducted, menarche occurred (the second upper leaf of the decision tree on the right side in Fig. 1). Looking at the figure, it may be noted that among the girls whose parents' educational level was higher, and their access to household consumer goods was greater, menarche occurred more often.

A predictors ranking list for 2001 is presented in Fig. 2. The number of children in a family was attributed the highest discriminatory power. Parents' educational level and access to household appliances had a smaller yet comparable importance, and the smallest importance was attributed to the family's income source.
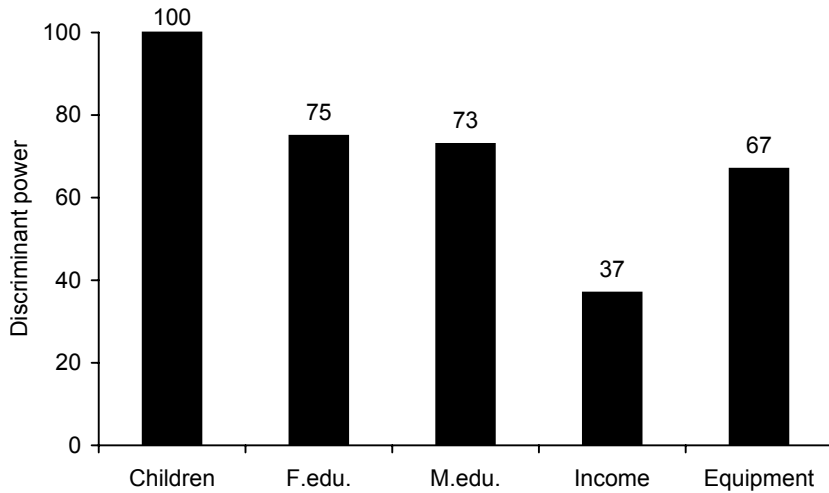
Fig. 2. Importance ranking list of socioeconomic predictors of menstruation appearance in girls aged 12–14 from the area of Choszczno. Sample of 608 persons, 2001 data. See legend to Fig. 1 for definitions of variable abbreviations. Source: authors' own study.

An analogical analysis was conducted using data from 1987. Figure 3 presents a decision tree obtained by the CART method for these data. It comprises thirteen divisions and fourteen final leaves. They indicate the following: if the number of children in a family was higher than four, a subset containing 36 girls was created in the next stage. Then, if an economy variable 'income source' was 1 or 2, a subset of thirteen girls who had not menstruated and whose families' income source was farming or farming and working elsewhere was obtained. Non-farming income sources allowed distinguishing a subset of 23 girls, which was then divided based on the father's educational status and next, the mother's. As a result a group of 36 girls from large families (at least five children) was divided according to the values of three socioeconomic features through consecutive analysis into two separate subsets: one with six girls with menarche and another with the remaining 30 girls (right tree branch). The left tree branch of girls from families with no more than four children (examined sample included 144 of them) was much more numerous. In that group better household appliances and parents' educational level enabled distinguishing subsets of menstruating girls.

Figure 4 presents a ranking list of predictors (the socioeconomic diagnostic variables) on a scale of 1 to 100. Similarly to 2001, it indicates that the strongest discriminatory power for the menstruation appearance variable was attributed to the mother's educational level and the number of children in a family, and subsequently the father's educational level, household wealth, whereas the least was attributed to income source. A decreasing role of educational level and household equipment in 2001, in relation to 1987, may mean the levelling of living conditions of schoolgirls from the rural surroundings of Choszczno.
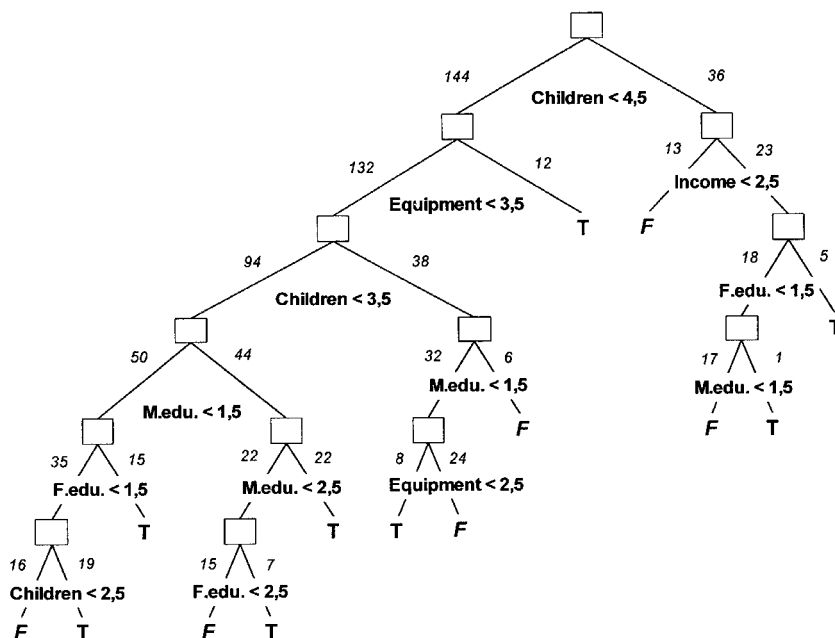
Fig. 3. Classification tree of menarche appearance (T, true; or F, false) in girls aged 12–14 from the area of Choszczno (1987 data). See legend to Fig. 1 for definitions of variable abbreviations. Source: authors' own study.
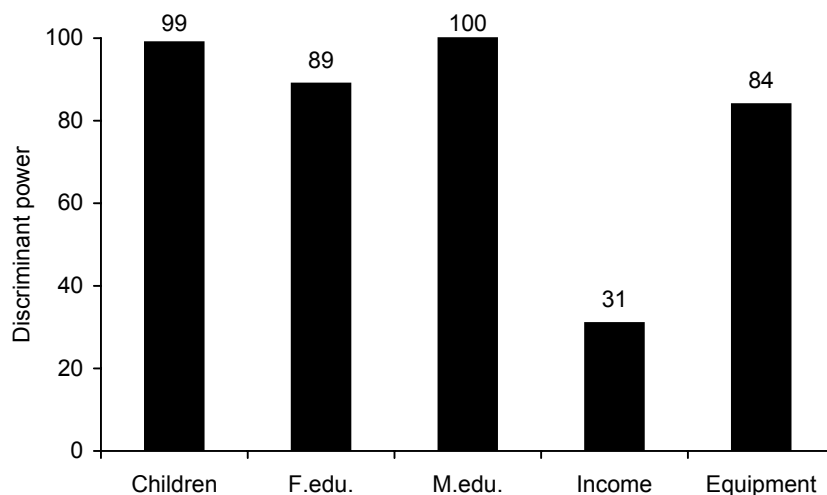

Fig. 4. The ranking list of the importance of the socioeconomic predictors of menarche appearance in girls aged 12–14 from the area of Choszczno. Sample of 180 persons, 1987 data. See legend to Fig. 1 for definitions of variable abbreviations. Source: authors' own study.

Table 2. Menstruation appearance misclassification matrix obtained by CART method

	2001 (*N*=608)		1987 (*N*=180)	
	Yes (T)	No (F)	Yes (T)	No (F)
Yes (T)		122		25
No (F)	115		32	

Source: authors' own study.

A complexity of the decision model is a function of the problem considered and the level of modelling accuracy in order to achieve the highest possible conformity to the observed menarche appearance in the examined samples. The tree obtained through dichotomous relations brings the examined reality closer, so goodness-of-fit measured by the number of faulty classifications (Table 2) is significant. Classification efficiency of the CART method could especially be confirmed due to the fact that menarche appearance or absence in the examined samples was almost equally distributed.

In 2001 66.1% of girls were classified properly. In 1987, 25 girls without menstruation and 32 girls already menstruating were classified incorrectly (for 2001 the values were 122 and 115, respectively). Better efficiency (70.6%) was obtained for girls examined in 1987. It may prove that the socioeconomic determinants of menarche occurrence played a more important role and affected the biological development in the examined groups more strongly, and that is why these variables were more efficient at classification by the decision trees method.

A tree depth is connected with the number of splits and is controlled by algorithm stop parameters. Classification accuracy can be increased by its modification, but it usually decreases model clearness and increases the complication level of relations leading towards leaves that are assumed to make homogeneous subsets. Generally, it can be admitted that in both cases the CART method's efficiency was demonstrated for assessing the importance of socioeconomic determinants of rural girls' biological development.

Measurement data were statistically processed using Statistica (version 7) and Microsoft Excel.

Discussion

The purpose of this study was to assess the usefulness of the multivariable decision trees method to examine the menarche of teenagers from families of different socioeconomic status and to learn the discriminatory power of particular social variables describing such status.

The highest discriminatory power in the decision trees method was shown by the number of children in a family and mother's educational level. A very important role of these two factors in determining body height was indicated by studies conducted

on Polish conscript soldiers between 1965 and 1995 (Bielicki *et al.*, 1997). From among four factors considered by the authors, the number of children was the most significant in 1965 and 1986, whereas in 1995 mother's educational level became the leading feature.

In studies of rural girls (Łaska-Mierzejewska & Olszewska, 2006, 2007), the number of children more strongly differentiated menarche age than any other variables considered. A spectacular result referring to the number of children in a family was achieved in a 2001 study that focused on four examined areas (*N*=9599) (Łaska-Mierzejerwska & Olszewska, 2004). Menarche age increased monotonically following family enlargement by each subsequent child from one to seven, or more. Only children started menstruating at the age of 12.99, and their peers from families with seven children started menstruating at 13.52 years.

High discriminatory values in the decision trees method were indicated by parents' educational status, favouring mother's educational level. One feature study of the same sample and a complex study of all research areas showed a significantly lower role of the variable in assessing menarche age in comparison with the number of children in a family. In 2001 the difference of menarche age among inhabitants of Choszczno from families of extreme categories of father's educational level amounted to 0.18 of a year, and for mothers 0.1 of a year; and in a complex study of all areas the differences were 0.22 and 0.13, respectively.

The next social variable determining family's social status and statistically associated with the menarche of teenagers was the ownership of household appliances. The acquisition of appliances significantly improved between 1987 and 2001. In the Choszczno area the percentage of families owning all of the following – a colour TV, freezer, fridge (1987) or video (2001), automatic washing machine and a car – increased from 7% to 54%, and the percentage of families without any of those decreased from 44% to only 0.1%. Whether or not a particular item was owned, and especially the absence of several items, significantly differentiated menarche age. These factors substantially determined menarche age in the above-mentioned studies.

A relatively low discriminatory value of the income source in a rural family in comparison with other analysed variables needs wider elaboration. There are probably two reasons for this: increasing differences in menarche age among girls from rural families of different income sources and some dissimilarity of socioeconomic situation of the inhabitants of Choszczno area in comparison with other examined areas. The differences in menarche age between the earliest maturing girls from non-farming families and the latest menstruating farmers' daughters became smaller between subsequent researches: from 0.53 in 1967, 0.44 in 1977, 0.33 in 1987 and just 0.15 in 2001. The differences refer to the complex study of all areas. The pattern of menarche age differences among girls from non-farming families maturing latest and farmers' daughters maturing earliest remained the same in consecutive studies in each of the examined areas, apart from Choszczno, in 1987 and 2001. The economic crisis in Poland between 1977 and 1987 affected differently inhabitants of various urban areas (Hulanicka *et al.*, 1990; Hulanicka & Waliszko, 1991) and various social groups. Among the country dwellers it affected mostly the non-farming population, and delayed menarche age among girls from the group by 0.33 year; in farmers' and food producers' daughters there was an increase in menarche age of 0.08

year. After the political and economical transformation in Poland of 1989, a levelling of the biological losses of the crisis time took place. Between 1987 and 2001, the acceleration of menarche age in the farming group reached 0.36, and in the non-farming group it was only 0.24 (Łaska-Mierzejewska & Olszewska, 2007).

The non-farming group in the Choszczno area was affected by mass unemployment resulting from the closing down of state farms, as the region was characterized by the highest concentration of state-owned land in comparison with other examined areas. It can be conjectured that the discriminatory value of the rural family income source would have been higher had another region of Poland been investigated.

The method of decision trees is infrequently applied in social sciences in Poland and is a sort of novelty (Matusik, 2004, 2007; Matusik & Woźniacka, 2007). This study attempts to draw attention to this method using existing data (Łaska-Mierzejewska & Olszewska, 2006), to show some of its characteristics, advantages and existing limitations. The algorithm itself, published in the mid-1980s, could be applied effectively thanks to advancing computer techniques.

The deliberately undertaken challenge, considering a methodological type of work, was to show an almost identical probability of menarche occurrence or its absence in the examined subgroups. With such assumptions it was easier to prove effectiveness of the applied classification method which, the authors conclude, was accomplished. It should be emphasized that the determinants were only socioeconomic features expressed in a few categories. Their statistical associations on adult population growth, body height and pubescence pace have been analysed by many authors (among others Bielicki *et al*., 1986; Hulanicka *et al*., 1990; Charzewski *et al*., 1991, 1998), and the results of the study are consistent.

Due to increasing difficulties in obtaining information on menarche age, the models also have a practical advantage of high probability of correct assessment of girls' biological development level, based on relatively easily accessible information considering socioeconomic status of their families.

The models are also of cognitive importance. For example, the 2001 model implies that 59 girls out of 309 (i.e. 19% – right branch in Fig.1) were from families where both parents had secondary educational level (or, possibly, father's educational level was basic vocational), the number of children was between two and four, the family owned all household appliances and their income source was from non-farming activities. An analogical, though more complex, deduction can be made for another group of 59 girls (left branch in Fig. 1), as well as for 2001 data. A similar dependency analysis obtained from the 1987 model indicates that the most numerous group of 22 girls (i.e. 23% from among 96 examined – left branch in Fig. 3) who reached menarche age, came from families of no more than three children, rather rich in material possessions (category 1, 2 or 3), in which mothers had secondary or higher educational level. The obtained classification results are coherent and consistent with existing knowledge on the association of living conditions with girls' menarche age.

Conclusions

This work, based on 2000 and 1987 data, shows the mechanics of the decision trees method based on the CART (Classification and Regression Trees) algorithm, with

Gini's heterogeneity measure, presenting relational models based on binary divisions, classifying pubescence advance in girls from rural areas of Poland (near Choszczno, West Pomerania district). The models indicated classification correctness for 66–70% of cases, relying on five commonly accessible socioeconomic variables only. Better classification efficiency was obtained for 1987, which may indicate a decreased association of socioeconomic differentiation affecting the level of biological development in rural girls from West Pomerania district, Poland, in 2001. The application of the decision trees method enabled the definition of the hierarchy of nominal, quantitative socioeconomic variables influencing the girls' biological development level. The strongest discriminatory power was attributed to the number of children in a family and mother's level of education, followed by father's level of education and variables connected with family's economical status as well as its living conditions.

The results of this study are also consistent with the results of the analysis of the same data using the mean of menarche age and depending on the girls' place on the scale of each of the analysed social variables (Łaska-Mierzejewska & Olszewska, 2006). It proves the usefulness of the decision trees method in examining relations between biological processes taking place in populations and the conditions in which the populations exist.

References

Adair, L. S. (2001) Size at birth predicts age at menarche. *Pediatrics* **107**, e59, 1–7.

Belsky, J., Steinberg, L. & Draper, P. (1991) Childhood experience, interpersonal development, and reproductive strategy; an evolutionary theory of socialization. *Child Development* **62**, 647–670.

Bielicki, T., Szklarska, A., Welon, Z. & Brajczewski, C. (eds) (1997) Nierówności społeczne w Polsce. Antropologiczne badania poborowych w trzydziestoleciu (1965-1995). *Monografie Zakładu Antropologii PAN Wrocław* **16**.

Bielicki, T., Waliszko, H., Hulanicka, B. & Kotlarz, K. (1986) Social class gradients in menarcheal age in Poland. *Annals of Human Biology* **13**, 1–11.

Bliss, C. I. (1934) The method of robits. *Science* **79**, 38–39.

Breiman, L., Friedman, J., Olshen, R. & Stone, C. (eds) (1984) *Classification and Regression Trees*. Wadsworth & Brooks/Cole, Monterey, CA.

Charzewski, J., Łaska-Mierzejewska, T., Piechaczek, H. & Łukaszewska, L. (1991) Wiek menarche dziewcząt warszawskich w 1976-1986 roku. *Wychowanie Fizyczne i Sport* **35**(2), 15–29.

Charzewski, J., Lewandowska, J., Piechaczek, H., Syta, A. & Łukaszewska, L. (1998) Wiek menarche dziewcząt warszawskich 1986-1997. *Wychowanie Fizyczne i Sport* **42**(1), 61–66.

Ellis, B. J. & Essex, M. J. (2007) Family environments, adrenarche, and sexual maturation: a longitudinal test of a life history model. *Child Development* **78**, 1799–1817.

Finney, D. J. (1971) *Probit Analysis*. Cambridge University Press, Cambridge.

Gatnar, E. (2001) *Nieparametryczna metoda dyskryminacji i regresji*. Wydawnictwo Naukowe PWN, Warszawa.

Gini, C. (1921) Measurement of inequality of incomes. *The Economic Journal* **31**, 124–126.

Hulanicka, B., Brajczewski, C., Jedlińska, W., Sławińska, T. & Waliszko, A. (1990) City–town–village. Growth of children in Poland in 1988. *Monografie Zakładu Antropologii PAN Wrocław* **7**.

Hulanicka, B. & Waliszko, A. (1991) Deceleration of age at menarche in Poland. *Annals of Human Biology* **18**, 507–513.

Kozieł, S. & Jankowska, E. A. (2002) Effect of low versus normal birthweight on menarche in 14-year-old Polish girls. *Journal of Paediatrics and Child Health* **38**, 268–271.

Larose, D. T. (2005) *Discovering Knowledge in Data. An Introduction to Data Mining*. Wiley-Interscience.

Łaska-Mierzejewska, T. (1995) Age at menarche as an indicator of the socio-economic situation of rural girls in Poland in 1967, 1977 and 1987. *American Journal of Human Biology* **7**, 651–656.

Łaska-Mierzejewska, T. (1970) Effect of ecological and socio-economic factors on the age at menarche, body height and weight of rural girls in Poland. *Human Biology* **42**, 284–292.

Łaska-Mierzejewska, T., Milicer, H. & Piechaczek, H. (1982) Age at menarche and its secular trend in urban and rural girls in Poland. *Annals of Human Biology* **9**, 227–233.

Łaska-Mierzejewska, T. & Olszewska, E. (2004) The maturation rate of girls living in rich and poor rural regions of Poland before and after the transformation of 1989. *Homo – Journal of Comparative Human Biology* **55**, 129–142.

Łaska-Mierzejewska, T. & Olszewska, E. (2006) Changes in the biological status of Polish girls from a rural region, associated with economic and political processes in the period 1967–2001. *Journal of Biosocial Science* **38**, 187–202.

Łaska-Mierzejewska, T. & Olszewska, E. (2007) Anthropological assessment of changes in living conditions of the rural population in Poland in the period 1967–2001. *Annals of Human Biology* **34**, 262–276.

Matusik, S. (2004) Społeczno-ekonomiczne uwarunkowania poziomu umieralności niemowląt w gminach województwa małopolskiego. *Annales Universitatis Mariae Curie-Skłodowska. Sectio D Medicina* **59** (Supplement XIV), 48–54.

Matusik, S. (2005) Modeling a level of development in Malopolskie using decision trees. *International Advances in Economic Research* **11**(3), 343–344.

Matusik, S. (2007) Demographic factors in economic development models for the Malopolskie communes in 2002. In Welfe, W. & Wdowiński, P. (eds) *Modelling Economies in Transition 2006*. AMFET Monographs, Łódź, 153–169.

Matusik, S. & Woźniacka, R. (2007) The decision trees method model of growth increment rate of long bones in children of differentiate physical activity. *Polish Journal of Environmental Studies* **16** (5C, Part II), 371–374.

Moffitt, T. E., Caspi, A., Belsky, J. & Silva, P. A. (1992) Childhood experience and the onset of menarche: a test of a sociobiological model. *Child Development* **63**, 47–58.

Rimpelä, A. H. & Rimpelä, M. K. (1993) Towards an equal distribution of health? Socioeconomic and regional differences of the secular trend of the age of menarche in Finland from 1978 to 1989. *Acta Paediatrica* **82**, 80–87.