

Application of Survival Analysis in Studies of Human Ontogeny

Magdalena Kosińska, Anita Szwed

Department of Human Biological Development, Institute of Anthropology, Faculty of Biology, Adam Mickiewicz University, Poznań, Poland

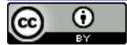
Email: gdua@amu.edu.pl, aniszwed@amu.edu.pl

Received 5 March 2014; revised 20 April 2014; accepted 27 April 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The main goal of this work is to demonstrate the suitability of survival analysis for ontogenetic studies. The research material includes retrospective data of the age of the occurrence of ontogenetic events such as birth (N = 487), menarche (N = 2016) and menopause (N = 3597). In order to study the time of occurrence of ontogenetic events and to indicate the impact of environmental factors the survival analysis was applied. First, the percentiles of functions established for studied events were calculated. Next, the Kaplan-Meier survival curves were derived. In the last step the influence of environmental factors was established and the comparison of groups determined based on the chosen factors was performed. The delivery time shows that 14% of infants were born pre-term. The risk of preterm delivery increases with the severity of factors disrupting pregnancy (from none to coexisting maternal and fetal risk factors) ($\chi^2_4 = 64.695$; $p < 0.001$). In the case of menarche percentile positions indicate that the menarche occurs between the 12th and the 14th year of life as the period in which most girls exceed the puberty threshold. The Cox's proportional hazard model indicates that the time of menarche occurrence is significantly depended ($\chi^2_7 = 41.602$; $p < 0.001$) on the place of the mother's residence and number of children in the family (respectively $p < 0.03$ and $p < 0.001$). The time interval established for 50% occurrence of this experience was designated between the 49th and the 52nd years of life. The time of menopause occurrence is significantly depended on both of considered factors: the educational level and smoking cigarettes ($\chi^2_2 = 42.365$, $p < 0.001$). Survival analysis is suitable for studies of the distribution in time of developmental events. It can be used to indicate the factors which significantly influence the course of development by modifying the duration of developmental stages.

Keywords

Ontogenetic Events, Birth, Menarche, Menopause, Kaplan-Meier Method, Cox's Proportional Hazard Model

1. Introduction

In the center of interest of many biologists, especially dealing with human biology, there are issues related to developmental processes occurring in the individual's life, from fertilization till death *i.e.* [1]. The basic paradigm of developmental phenomena is the statement that an individual during development goes through the subsequent, genetically determined stages. Crossing to the next stage is equivalent to overcoming the selection threshold and achievement the opportunity to reach the next life stages [2]. Progresses from one stage to the next one are considered as the most important events in the individual life, key moments of human development. Reaching each of them reflects the normal course of the development processes. However, it should be emphasized that the organism, in accordance with its genetic properties, actively responds to environmental factors *i.e.* [1] [2]-[5]. One of the consequences of interactions between the genotype and environmental conditions is the diversity in time of achieving the subsequent ontogenetic stages. Thus, researchers dealing with developmental biology are often interested in the time of occurrence of these events and in indicating the reasons of inter-individual diversity *i.e.* [6]-[8].

The existence of such issues creates a necessity for using statistical procedures for which time is the examined variable. The method suitable for studies of the distribution in time of ontogenetic events and the comparison of period that passes between the initial stage (for example the fertilization or birth) and the moment of the event's occurrence, in the relation to biological and environmental conditions, is the survival analysis. These are a set of statistical procedures, which enable the study of the appearance and the distribution of events in time. The outcome variable is the time until an event occurs.

The main goal of this work is to demonstrate the suitability of survival analysis for ontogenetic studies.

2. Materials and Methods

The research material includes retrospective data of the age of the occurrence of ontogenetic events such as birth, entering the puberty and the end of reproductive period. The considered life events reflect the change of the organism's state. The age at birth (gestational age) reflects the time passing from the moment of conception to the delivery (in weeks); the age of entering the puberty is the time passing from the birth to menarche (in years); and the age of ending the reproductive period is the time passing from the birth to menopause (in years). The evaluation of the distribution in time of these events was estimated based on the studies of three independent groups of females. The data of gestational age were taken from 487 live born infants. In the case of the age of menarche data were collected in the group of 2016 girls aged 8 - 23 years, and in the studies of the age of menopause 3597 women between ages of 35 - 60 years were examined.

In order to study the time passing between the starting point (conception for the evaluation of the gestational age and the birth for remaining analyses) and the moment of studied event the Kaplan-Meier method was used. To determine the set of variables, which have an influence on the time of the occurrence of analyzed ontogenetic events the Kruskal-Wallis test in the case of birth (there were no censored data) and Cox's proportional hazard model for the menarche and menopause were employed. Next the most significant modifiers were chosen to present the differentiation of "survival" functions. In the analysis of the influence of environmental components on the length of prenatal development the maternal and fetal risk factors during the course of pregnancy were used. The number of children in the family and smoking habits were applied for menarche and menopause respectively. The significance of differences between distinguished groups was calculated using χ^2 test.

All statistics were performed on the basis of the statistical software package Statistica 10.0 (Statsoft. Inc. 2014).

3. Results

In order to study the time of occurrence of ontogenetic events and to indicate the impact of environmental factors the survival analysis was applied. First, the percentiles of functions established for studied events were calculated. Next, the Kaplan-Meier survival curves were derived. In the last step the influence of environmental factors were established and the comparison of groups determined based on the chosen factors was performed. For the length of prenatal development the pregnancy clinically diagnosed risk factors were used and the number of children in the family for menarche and smoking habits for menopause were applied.

In the case of the time of birth 50% of newborns were delivered between the 37th and the 40th week of gesta-

tion (**Table 1**). The function estimated for the delivery time (**Figure 1**) shows that 14% of infants were born before the completion of the 37th week of gestation. Out of them 21% were extremely preterm delivered ($x \leq 29$ week of gestation), 29% were very preterm delivered (30 - 32 week of gestation), and preterm delivered (33 - 36 week of gestation) constituted 50% of this group. The distinct slope of curve is visible between the 37th and the 42nd week and the largest number of births was observed during the period between the 38th and 40th week of gestation.

In **Table 2** the influence of variables, commonly treated as predictors of the pregnancy duration, was presented *i.e.* [9] [10]. It was found that the urbanization level and the presence of prenatal clinically diagnosed risk factors should be treated as predictors of the gestational age at the delivery moment.

For the particular categories of the clinically diagnosed risk factor [1/none, 2/premature rupture of membranes, 3/maternal disorders, 4/fetal disorders, 5/coexisting maternal and fetal disorders] functions were estimated and compared (**Figure 2**). Obtained results showed that the risk of preterm delivery increase with the severity of factors disrupting pregnancy (from none to coexisting maternal and fetal risk factors) ($\chi^2 = 64.695$; $p < 0.001$) (**Figure 2**). Although the curve plotted for infants from pregnancies affected by the premature rupture of membranes until the 37th week had the similar course to the curve plotted for infants delivered from pregnancies not affected by any disorders and pathologies. In the group of newborns born from pregnancies not threatened by the occurrence of risk factors the highest frequency of births coincides with the period of term delivery, precisely between the 38th and the 42nd week of gestation. Clearly different course of curves is visible for the time of birth pregnancies burdened with the occurrence of fetal or maternal and fetal risk factors coexistence. In these groups the pronounced slope of curves started much earlier, at the 30th week of gestation.

In the case of menarche percentile positions indicate the interval between the 12th and the 14th year of life as the period in which most girls exceeds the puberty threshold (**Table 1**)—50% of studied girls entered the pubertal stage during these years of life. This result was confirmed by the estimated Kaplan-Meier function and the curve plotted on its base (**Figure 3**). Most pronounced decrease is attributable the time lasting from the 12th up to the 14th year of life. In this period there is the largest number of maturation events. Hence, the age range between 12 and 14 years may be treated as the most typical time for the menarche occurrence. During this period

Table 1. Percentiles of functions established for studied ontogenetic events.

Percentiles	Time of birth (weeks of gestation)	Time of menarche (years)	Time of menopause (years)
25 th (lower quartile)	37	12	49
50 th (median)	39	13	51
75 th (upper quartile)	40	14	52

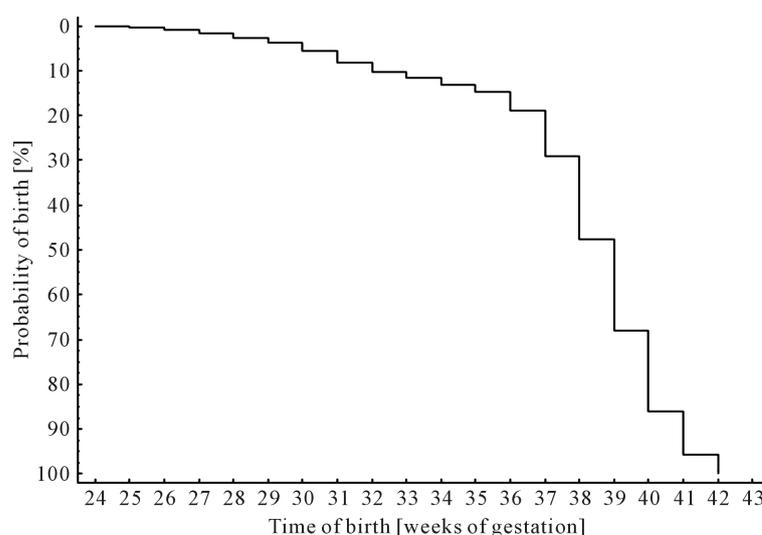


Figure 1. Kaplan-Meier curve applied to age at delivery.

Table 2. The set of variables correlated with the length of prenatal development (the Kruskal-Wallis test).

Variable (N = 487)	Kruskal-Wallis test H value	df	p
Mother's place of residence	25.937	3	<0.001
Mother's work	2.274	2	>0.05
Father's work	1.130	2	>0.05
Mother's education	0.213	2	>0.05
Father's education	3.017	2	>0.05
Mother's age	1.808	4	>0.05
Father's age	1.219	3	>0.05
Birth order	4.605	3	>0.05
Pregnancy order	1.847	3	>0.05
Prenatal (clinically diagnosed) risk factors	64.695	4	<0.001

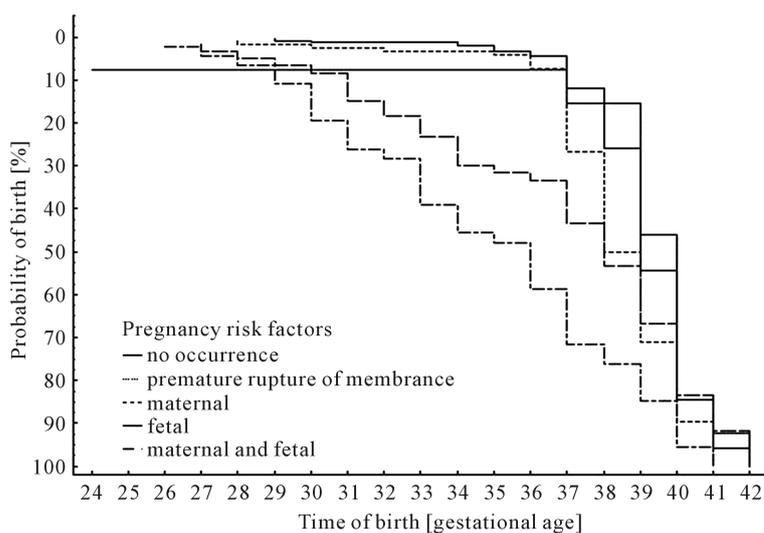


Figure 2. Kaplan-Meier curves applied to the age at delivery of infants distinguished by the categories of pregnancy risk factors.

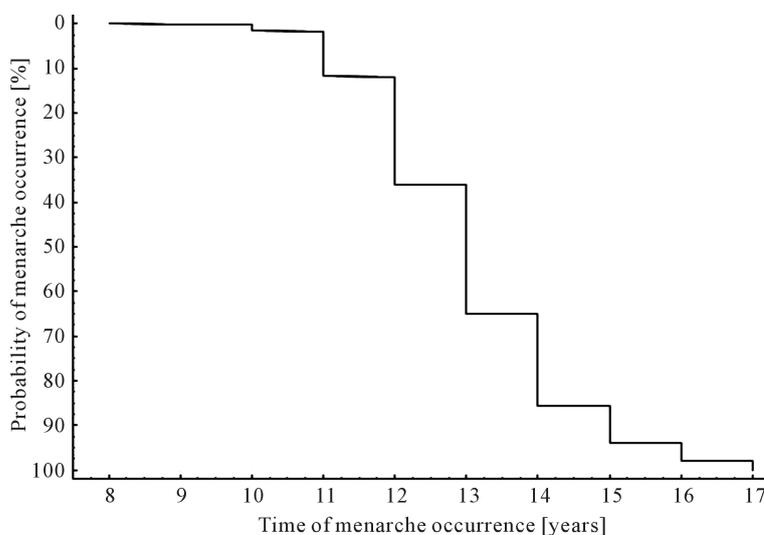


Figure 3. Kaplan-Meier curve applied to age of menarche.

of time the largest number of studied event occurrence was noted (particularly in the 13th year of life). Approximately 10% of girls constitute the group of early matured up to the 11th year of life. Less than 2% of girls reached maturity before this age. The earliest menarche was in the 9th year of life. As expected, most pronounced decrease is attributable the time lasting from the 12th up to the 14th year of life. In this period there is the largest number of maturation events. However, it should be noted that the completion of 11th year of life may already be regarded as characteristic for the start time of the studied event. Late menarche (at the 15th year or even later) was estimated for approximately 5% of girls.

The Cox's proportional hazard model indicate that the time of menarche occurrence is significantly depended ($\chi^2_7 = 41.602$; $p < 0.001$) on the place of mother's residence and number of children in the family (respectively $p < 0.03$ and $p < 0.001$) (Table 3).

Then curves distinguished by categories of the number of children in the family were compared (Figure 4). The family size clearly differentiates the age of menarche ($\chi^2_3 = 53.182$, $p < 0.001$). As it was mentioned less than 2% of girls reached maturity up to the age of 11 years. Sections of the curves plotted for a period of time before that age included many censored data. The curve period where the earliest menarche was noted in the 9th year of life, but it was observed only in two girls. The differences among plotted curves are clearly visible from the 11th year. Girls from families with one child experienced the maturation in the earliest age and girls from families with more children much later. The percentage of late matured girls increase with the family size. For example in the 15th year of life approximately only 5% of only child and 10% of girls from families with four or more children did not have menarche, respectively in the 16th year of life 2 and 10 percent.

Table 3. The set of variables correlated with the time of menarche occurrence.

Variable (N = 2016)	Beta	SE	Test t value	Wald stat.	p
Mother's education	0.026	0.042	0.624	0.389	>0.05
Father's education	0.047	0.042	1.118	1.251	>0.05
Mother's place of residence	-0.146	0.065	-2.253	5.078	<0.003
Father's place of residence	0.093	0.066	1.413	1.997	>0.05
Mother's work	0.058	0.038	1.522	2.315	>0.05
Father's work	-0.037	0.048	-0.771	0.595	>0.05
Number of children in the family	-0.115	0.025	-4.656	21.675	<0.001

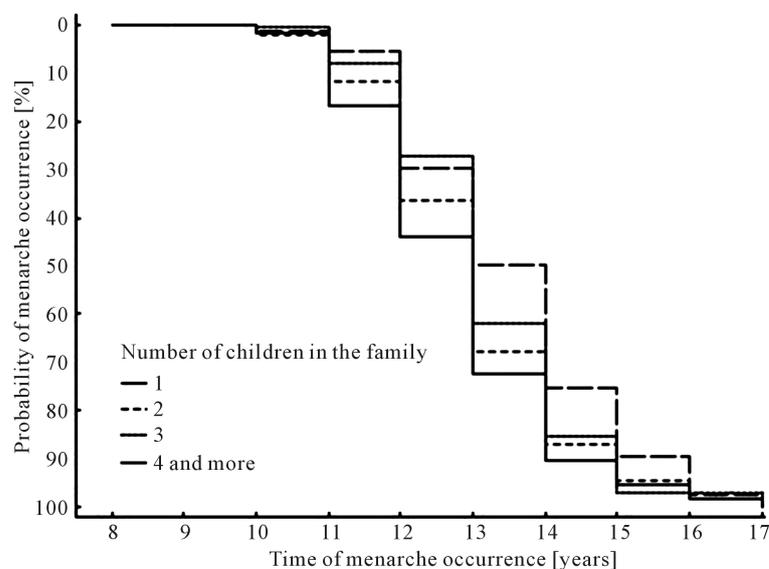


Figure 4. Kaplan-Meier curves applied to the age of menarche occurrence of girls distinguished by the categories of family size.

The next important event during the women's life is the end of reproductive period-menopause. The high proportion of studied women reported the end of the procreation period (63%). The time interval established for 50% occurrence of this experience was designated between the 49th and the 52nd years of life (**Table 1**). This period of time was also indicated by the Kaplan-Meier function (**Figure 5**). The curve illustrating the distribution of age of the reproductive expiration is characterized by the distinct slope starting at the 48th year of life and maintaining for up to the 53rd years of life. The largest number of observations falls into the period between the 50th and the 53rd year of life. Only 3% of women were classified to the group of women experienced the early menopause. Until the completion age 47 years there is mild and slight decrease of the plotted curve. Within the time designated by a starting point of studies (35 years) and the age 47 age only about 6% of surveyed women had the menopause. In turn, after the completion of the 53rd year only 5% of women still retained menstrual cycle.

As with the above-discussed events the influence of environmental factors was estimated. In this case Cox's proportional hazard model indicates that the time of menopause occurrence is significantly depended on both of considered factors: the educational level and smoking cigarettes ($\chi^2 = 42.365$, $p < 0.001$) (**Table 4**).

The curves plotted for women according to their smoking habits (**Figure 6**) showed that smoking cigarettes influences the age at menopause ($\chi^2 = 67.050$, $p < 0.001$). Women who had never smoked experienced menopause at the latest age. In group of smoking women menopause was significantly accelerated compared to non-smokers and also to casual smokers. The decrease of curves plotted for women smoking at least 10 cigarettes per day started earlier, at about 3 years with respect to non-smoking women.

4. Discussion

Human biologists, especially dealing with auxological issues are interested in developmental processes occurring in the individual's life. Ontogenetic development involves irreversible directional changes that occur in genetically determined sequence. These changes are the important events in the individual's life. The age at which these events occurred is interpersonal varied. The diversity is the result of the fact that the occurrence of them is depended on genetic properties and the action of environmental factors [11]. Therefore, in developmental studies, frequently considered issue is the time which passes between the initial stage (for example birth) and the

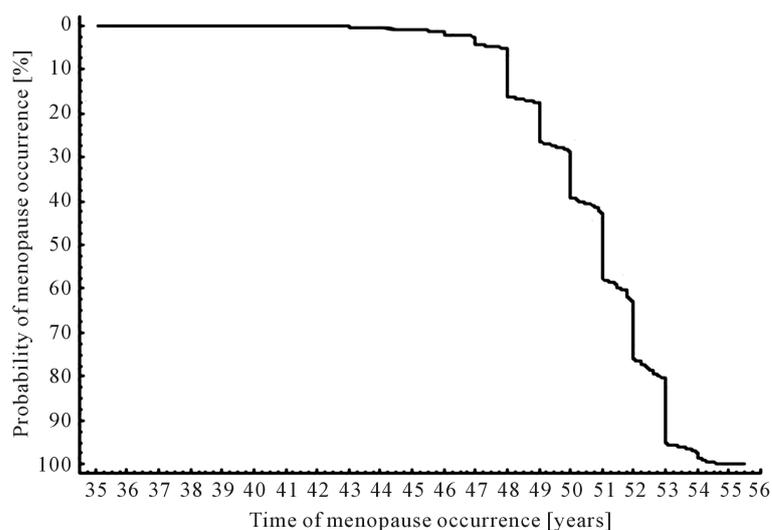


Figure 5. Kaplan-Meier curve applied to age of menopause occurrence.

Table 4. The set of variables correlated with the time of menarche occurrence.

Variable (N = 3597)	Beta	SE	Test t value	Wald stat.	p
Mother's education	-0.068	0.023	-2.972	8.833	<0.003
Smoking cigarettes	0.161	0.026	6.257	39.152	<0.0021

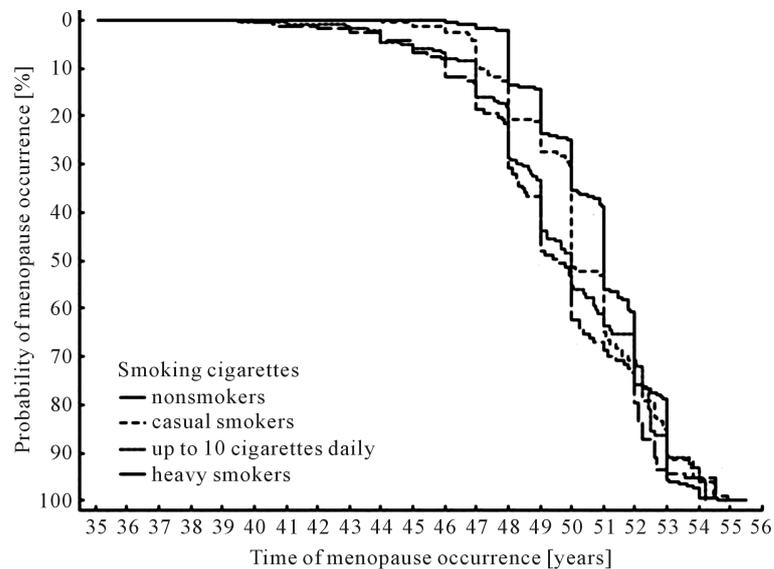


Figure 6. Kaplan-Meier curves applied to the age of menarche occurrence of girls distinguished by the categories of family size.

moment of life event occurrence. In this meaning time refers to the age of an individual when an event occurs. In biological studies, to resolve these issues many statistical methods are used. Most often, they are allowed to determine the average age of the studied phenomenon, evaluate the impact of environmental variables, and examine the interactions between the examined variables. However, frequently there is a necessity for using a statistical method for which time is the examined variable.

A collection of statistical procedures for which the outcome variable of interest is time (exactly the time until the studied event occurs) is survival analysis [12]. These procedures enable the study of the appearance and the distribution of events in time. In medicine it is mainly applied for the estimation of mortality and morbidity [13] [14]. In biology the survival analysis is applied by researchers dealing with environmental biology. For example, to evaluate the survival of the examined animal species, or their offspring, in various environmental conditions *i.e.* [15]. Despite the name relating to survival this analysis may be applied in studies, which do not relate to survival in the literal sense. It may be used in human biology studies, which focus on the issues associated with the time of occurrence of ontogenetic events *i.e.* [1] [16] [17]. Events which reflect the changing of organism' state, thus represent important moments of ontogeny.

The studies, which have been carried out, showed that "survival" analysis is suitable for studies of the distribution in time of developmental events. It can be used to indicate the factors which significantly influence the course of development by modifying the duration of developmental stages. Moreover, there is a possibility of comparison of the functions describing the time of studied moment occurrence, according to these factors. The advantage of survival analysis is the possibility of considering censored data. Sometimes in studies of developmental biology researchers have certain information about the time of the event occurrence but they do not know this time exactly. The advantage of presented statistical procedures is the ability to rely on censored data. It is only known that the time until the occurrence of the phenomenon is at least as long as the period that the studied person has been followed. In the presented studies every data of delivery time were completed. However, in the case of studies relating to time of menarche and menopause occurrence many of collected data were censored. They were data obtained from girls who did not have the menarche up to the moment of examination. For those girls the study ended before this ontogenetic event. Similar situation concerns the studies of menopause occurrence. Not every woman has completed a period of the investigated reproductive stage of life.

References

- [1] Bogin, B. (1988) *Patterns of Human Growth*. Cambridge University Press, Cambridge.
- [2] Szwed, A. and Kosińska, M. (2012) Biological Maturity at Birth, the Course of the Subsequent Ontogenetic Stages and Age at Menarche. *HOMO-Journal of Comparative Human Biology*, **63**, 292-300.

- <http://dx.doi.org/10.1016/j.jchb.2012.01.003>
- [3] James-Todd, T., Tehranifar, P., Rich-Edwards, J., Titievsky, L. and Terry, M.B. (2010) The Impact of Socioeconomic Status across Early Life on Age at Menarche among a Racially Diverse Population of Girls. *Annals of Epidemiology*, **20**, 836-842. <http://dx.doi.org/10.1016/j.annepidem.2010.08.006>
- [4] INSERM Collective Expertise Centre (2007) INSERM Collective Expert Reports Paris, Institut national de la santé et de la recherche 2000. Growth and Puberty Secular Trends, Environmental and Genetic Factors. <http://www.ncbi.nlm.nih.gov/books/NBK10786>
- [5] Latinen, J., Power, Ch. and Järvelin, M.-R. (2001) Family Social Class, Maternal Body Mass Index, Childhood Body Mass Index, and Age at Menarche as Predictors of Adult Obesity. *American Journal of Clinical Nutrition*, **74**, 287-294.
- [6] Kramer, M.S., Wilkins, R., Gouletf, L., Séguinf, L., Lydonc, J., Kahnd, S.R., McNamarae, H., Dassaf, C., Dahhoua, M., Masseh, A., Minere, L., Asselinh, G., Gauthierf, H., Ghanemf, A., Benjamine, A. and Platta, R.W. (2009) Investigating Socio-Economic Disparities in Preterm Birth: Evidence for Selective Study Participation and Selection Bias. *Paediatric and Perinatal Epidemiology*, **23**, 301-309. <http://dx.doi.org/10.1111/j.1365-3016.2009.01042.x>
- [7] Łaska-Mierzejewska, T. and 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. <http://dx.doi.org/10.1016/j.jchb.2003.07.001>
- [8] Brambilla, D.J. and McKinlay, S.M. (1989) A Prospective Study of Factors Affecting Age at Menopause. *Journal of Clinical Epidemiology*, **2**, 1031-1039. [http://dx.doi.org/10.1016/0895-4356\(89\)90044-9](http://dx.doi.org/10.1016/0895-4356(89)90044-9)
- [9] Lumley, J. (1993) The Epidemiology of Preterm Birth. *Bailliere Clinical Obstetrics and Gynaecology*, **7**, 477-498. [http://dx.doi.org/10.1016/S0950-3552\(05\)80445-6](http://dx.doi.org/10.1016/S0950-3552(05)80445-6)
- [10] Lumley, J. (2003) Defining the Problem: The Epidemiology of Preterm Birth. *BJOG*, **110**, 3-7. <http://dx.doi.org/10.1046/j.1471-0528.2003.00011.x>
- [11] Pawlińska-Chmara, R. and Szwed, A. (2007) Cigarette Smoking and Age at Natural Menopause of Women in Poland. *Internet Journal of Biological Anthropology*, **2**.
- [12] Kleinbaum, D.G. Survival Analysis. A Self-Learning Text. Springer-Verlag, New York.
- [13] Xie, J., Brayne, C. and Matthews, F.E. (2008) Survival Times in People with Dementia: Analysis from Population Based Cohort Study with 14 Year Follow-Up. *BMJ*, **336**, 258-262. <http://dx.doi.org/10.1136/bmj.39433.616678.25>
- [14] Salles, G.F., Bloch, K.V. and Cardoso, C.R.L. (2004) Mortality and Predictors of Mortality in a Cohort of Brazilian Type 2 Diabetic Patients. *Diabetes Care*, **27**, 1299-1305. <http://dx.doi.org/10.2337/diacare.27.6.1299>
- [15] Kosiński, Z. and Ksit, P. (2006) Comparative Reproductive Biology of Middle Spotted Woodpeckers and Great Spotted Woodpeckers in a Riverine Forest. *Bird Study*, **53**, 237-246.
- [16] Szwed, A., John, A., Czapla, Z. and Kosińska, M. (2013) Influence of Socioeconomic Factors on Age at Menarche of Polish Girls. *Anthropologischer Anzeiger*, **70**, 455-470. <http://dx.doi.org/10.1127/0003-5548/2013/0338>
- [17] Kosińska, M. (2006) Two Methods for Estimating Age of Newborns in Catch-Up Growth Studies. *Early Human Development*, **82**, 575-582. <http://dx.doi.org/10.1016/j.earlhumdev.2005.12.004>