Correlates of Multidimensional Indicator of Quality of Life – Fractional Outcome Model Approach

Hanna Dudek[1] | *Warsaw University of Life Sciences – SGGW, Warsaw, Poland*
Wiesław Szczesny[2] | *Warsaw University of Life Sciences – SGGW, Warsaw, Poland*

Abstract

Quality of life indicators need to be measured through a multidimensional framework. In this study, the data from the survey 'Social Diagnosis' is used. The survey encompasses a set of 16 items relating to the evaluation of satisfaction with particular aspects of life. The item's categories are converted into a [0, 1] interval by using a membership function and then they are aggregated into a composite indicator. Fractional output models are applied to assess the impact of various socio-economic and demographic factors on values of this indicator. Such models are useful tools in cases when the response variable ranges between 0 and 1. It is found that satisfaction with life is U-shaped in age. Furthermore, it increases with education and association membership and decreases with disability, urbanisation, and being widowed or divorced. The results of the estimation indicate that the demographic composition of the household, region of residence and source of income all have a statistically significant impact on the quality of life in Poland.

Keywords	**JEL code**
Social diagnosis, quality of life, membership function, fractional outcome models	*I31, C25*

INTRODUCTION

Quality of life is a phrase encountered ever more frequently. It is used in so many contexts and for most different purposes that it is difficult to pin down a universally agreed meaning (Phillips, 2006). In the full sense of the term, Quality of Life (QoL) can be approached from an interdisciplinary perspective – the manner of its use depends on the discipline, and many are involved: sociology, economics, political science, social psychology, medicine, philosophy, marketing, environmental sciences and others (Glatzer, 2004). It has even been claimed that there may be as many definitions of QoL as there are people (Hoe et al., 2011). The recent trend has been to address methodologies that take into account individuals' opinions

[1] Department of Econometrics and Statistics, Warsaw University of Life Sciences – SGGW, Nowoursynowska 159, 02-776 Warsaw, Poland. Corresponding author: e-mail: hanna_dudek@sggw.pl, phone: (+48)225937220.

[2] Department of Informatics, Warsaw University of Life Sciences – SGGW, Nowoursynowska 159, 02-776 Warsaw, Poland. E-mail: wieslaw_szczesny@sggw.pl.

using broadly-designed tools based on questions about the subjective quality of life. Such an approach has an advantage – it prevents the risk of a person's QoL being judged by others, hence avoiding 'diminishing empowering people' in evaluating their own well-being (Rojo-Perez et al., 2015). While in the literature there is a lack of consensus on the meaning of 'quality of life', its multidimensional nature is universally accepted (Betti, 2017; Stiglitz et al., 2009; Eurostat, 2017). When measuring QoL, various domains should be analyzed, including subjective well-being. Indicators of satisfaction with various aspects of personal life are regarded as an important part of monitoring social situation. They enable the comparison of people's feelings against the objective data on living conditions, and thus are an indispensable and crucial element in the multidimensional measurement and analysis of the quality of life (Dudek and Szczesny, 2016).

This study examines the subjective perception of QoL using data from the 2015 survey 'Social Diagnosis – the objective and subjective quality of life in Poland'. It uses methodology first used in a multidimensional poverty analysis, and originally proposed by Cerioli and Zani (1990) and developed by Cheli and Lemmi (1995) and Betti and Verma (1999). It also employs methods of fuzzy set theory (Zadeh, 1965), according to which data on subjective assessments of QoL are converted by a membership function into a [0, 1] interval. Fuzzy set theory has become of particular interest to poverty researchers, since conventional crisp-set applications separating the poor and non-poor are increasingly believed not to adequately capture complex social phenomena like poverty (Neff, 2013).

In order to obtain a synthetic indicator encompassing many areas and aspects of life, weights reflecting the relative importance of satisfaction items are used. Such a framework was first applied in multidimensional poverty analysis (Betti and Verma, 2008; Panek, 2010), but recently also in various other socio-economic areas including job satisfaction (De Battisti et al., 2015) and quality of life (Betti et al., 2016; Betti, 2017). The interesting results obtained by Betti encouraged us to apply his approach to analyze the subjective QoL in Poland.

The present study often refers to Betti's work, where multidimensional fuzzy indicator methodology was first proposed and used to measure QoL (Betti et al., 2016; Betti, 2017). As in those articles, we calculate average values of the QoL indicators for the entire Polish population. As shown in (Betti et al., 2016), estimates based on sub-samples (population groups, regions and the like) can statistically differ from each other. In order to identify such differences, Betti et al. (2016) calculated standard errors for fuzzy indicators of QoL using Jackknife repeated replication. We propose another approach: applying fractional outcome models to explain the indicators. Thus, the main contribution of our research is the use of fractional response models and beta regression models in the fuzzy multidimensional analysis of QoL.

The paper is structured as follows: Following the present introduction, section 1 focuses on the data and methodology. Sub-section 1.1 briefly describes the 'Social Diagnosis' survey, sub-section 1.2 introduces the concept of the fuzzy set approach in a multidimensional measurement of quality of life and sub-section 1.3 gives insights on fractional outcome models. Section 2 presents and discusses the results of the analysis and section 3 provides our conclusions.

1 DATA AND METHODOLOGY

1.1. Data

The empirical analysis in this study is based on a 'Social Diagnosis – the objective and subjective quality of life in Poland' (SD) survey conducted in 2015. The SD is a cyclic survey that collects microdata on Poles' living conditions and quality of life as they report it themselves. The database is available free of charge at the website: <*www.diagnoza.com*>.

The 'Social Diagnosis' research project is undertaken by the members of the 'Council for Social Monitoring'. SD report authors and experts invited to participate by the 'Council' comprise economists, demographers, psychologists, sociologists, insurance specialists and statisticians. Headed by professor Janusz Czapiński, a social psychologist, and professor Tomasz Panek, a statistician, the project focuses on uncovering fundamental facts, behaviours, attitudes and experiences; not just an ordinary descriptive opinion poll, it is a scientific project.

The research was conducted in March and April 2015 by professional interviewers from the Central Statistical Office. The organisation of the questionnaire survey is supervised by the Polish Statistical Association's Office for Statistical Analyses and Research. Two separate questionnaires were used in the SD research.[3] The first provides the information about the household composition and living conditions completed by the interviewer during a meeting with the best-informed household representative. The second questionnaire was completed by all available household members aged 16 and above and contributes the information about their quality of life (Czapiński and Panek, 2015). The 2015 survey involved 11 740 households and 24 324 household members over 16 years of age. Since our study deals with a subjective assessment of QoL, we used the data derived from the second questionnaire, which was completed by 22 208 persons, which is the study's sample size.

The DS survey uses a two-stage stratified sampling method for selecting households[4]. Census areas were the primary sampling units, sampled with probabilities proportional to the number of dwellings they covered. Urban strata were divided into large towns with more than 100 000 residents, medium-sized towns of 20 000–100 000 and small towns with fewer than 20 000. In the five largest cities, the strata covered individual districts. In the second stage, three dwellings were sampled per census area in large towns, four per area in medium-sized ones, five per area in the smallest towns and six dwellings for rural areas (Czapiński and Panek, 2015). To preserve the representative character on the national study scale and in the identified classification cross-sections, weights for individuals were taken into account in the DS database.

The DS survey questionnaires contain numerous questions about respondent satisfaction with regard to particular areas and aspects of life. The scale of domain satisfaction covers 16 different items exhausting nearly the entire scope of the average person's interests and activities. Czapiński (2015) broke these items down into the following five dimensions:

- social aspects (satisfaction with relationships with closest family members, friends, spouses and children),
- material aspects (satisfaction with the family's financial situation and housing conditions),
- environmental aspects (satisfaction with the situation in the country, place of residence, and level of safety in place of residence),
- health-related aspects (satisfaction with one's health condition, sex-life and way of spending free time),
- self-assessment (satisfaction with one's own achievements, prospects for the future, educational level, work).

Respondents were asked to assess all 16 areas and indicate the extent of their satisfaction with each. There is a range of possible replies: 1) very satisfied 2) satisfied 3) rather satisfied 4) rather not satisfied 5) not satisfied 6) very dissatisfied 7) not applicable.[5] In our study, we assign a value of 3.5 to those individuals who indicated answer '7' and to those who did not give any answers, thus attributing them a neutral position. For 12 items such answers did not exceed 2% of all data (at most 2% of all respondents gave answer 7 or did not give any answers). However, there were individuals who were unmarried, had no children, no sex-life, or who did not work. They had no choice but to answer '7' because they could not assess a spouse, children, sex-life or work. These individuals accounted for 37%, 26%, 27% and 48% of respondents, respectively. So, for 4 of the 16 items, there is a very significant amount of missing information. Thus, we analyse the variant of the data with reduced list of 12 items.

---

3 Questionnaires and instructions for interviewers can be found at the website: <*www.diagnoza.com*> (Czapiński and Panek, 2015).

4 Details on sampling design can be found on the website: <*www.diagnoza.com*> (Czapiński and Panek, 2015).

5 See corresponding questionnaire item in the Appendix.

1.2 The multidimensional indicator of quality of life

In the study we analyze the multidimensional indicator of QoL. Our approach requires the following steps:
1) identify the relevant items and group them into dimensions,
2) convert the items' categories into item scores belonging to a [0, 1] interval,
3) assign weights to the aggregate items' scores in the QoL indicators,
4) calculate the QoL indicators.

As mentioned in the description of the SD research, one of the questionnaires includes 16 questions about satisfaction with particular areas and aspects of life. The answers (replies) to these questions created the items we analyzed in our study. According to SD research head Czapiński (2015), they are grouped into five dimensions: social, material, environmental, health-related and self-assessment. Thus, concerning the first step, omitting 4 items with a significant amount of missing information, we analyze 12 items grouped into 5 dimensions.[6]

In the second step, we construct a membership function for each item. Several methods have been proposed in the literature (Cerioli and Zani, 1990; Cheli and Lemmi, 1995; Betti and Verma, 2008) for how to construct this function. We opt to use the empirical distribution function of each item. Such an approach takes into account a given field's relative position in society. We use the formula fulfilling this requirement (Cheli and Lemmi, 1995):[7]

$$d_{k,j,i} = \frac{1-F(c_{k,j,i})}{1-F(1)}, \tag{1}$$

where: $c_{k,j,i}$ is the category of the j-th item in k-th dimension for the i-th individual, $1 \leq c_{k,j,i} \leq 6$,
F is the corresponding cumulative distribution functions.

The item's categories are ordered from the highest value of QoL to the lowest. Formula (1) converts them into a [0, 1] interval. The item score d can be interpreted as the degree of membership in the fuzzy set of satisfied people. In particular, the value 0 refers to the answer 'very dissatisfied' (c = 6) and the value 1 to 'very satisfied' (c = 1).

In the third step, weights of items were assigned within each of the five dimensions separately. Weights have to be considered as measures of relative importance of the items in the QoL indicators, relative to the other items in the dimension (Guio, 2009). They are essentially value judgements, and several approaches can be followed for defining them (Desai and Shah, 1988; Cerioli and Zani, 1990; Cheli and Lemmi, 1995; Filippone et al., 2001; Lazim and Osman, 2009).[8] In this study, we use the method proposed by Betti and Verma (1999) for two reasons: it assigns less importance to poorly differentiated items and it takes into account data redundancy. To do both, Betti and Verma (1999) defined weights as the product of two components:

$$W_{k,j} = W^a_{k,j} \cdot W^b_{k,j} \tag{2}$$

with the first factor being the coefficient of variation $V_{k,j}$ for j-th item score d in the k-th dimension, i.e.:

6 To identify dimensions, one can use statistical methods, for example factor analysis (Betti et al., 2016; Betti, 2017), but in this study we use a classification applied in the 'Social Diagnosis' Report, according to which there are five dimensions encompassing analyzed items.

7 The analogous formula was used in (Betti, 2017; Betti et al., 2016); the only difference lies in considering opposite ordering of categories c – in Betti's research they are ordered from the lowest value of QoL to the highest.

8 In research (Dudek and Szczesny, 2015) applying SD data methods proposed in papers (Desai and Shah, 1988; Cerioli and Zani, 1990; Betti and Verma, 1999) it was determined that the choice of weights does not significantly affect the distribution of synthetic indicators.

$$W^a_{k,j} = V_{k,j}\,, \tag{3}$$

and the second factor takes into account correlations among item scores:

$$W^b_{k,j} = \left(\frac{1}{1+\sum_{j=1}^{m_k} r_{k,jj'}|r_{k,jj'}<r^*_k}\right)\left(\frac{1}{\sum_{j=1}^{m_k} r_{k,jj'}|r_{k,jj'}\geq r^*_k}\right), \tag{4}$$

where: $r_{k,jj'}$ is the correlation coefficient between the two different scores $d_{k,j}$ and $d_{k,j'}$,
r^*_k is a predetermined cut-off correlation level in the k-th dimension,
m_k is the total number of items in the k-th dimension.

Thresholds r^*_k are determined by the point of the largest gap between the ordered set of correlation values encountered (Betti and Verma, 2008).

Using Formulas (3)–(4) results in weight $W_{k,j}$ being directly proportional to the variability of the $d_{k,j}$ and inversely proportional to its correlation with items in the k-th dimension. The low value of the factor $W^a_{(k,j)}$ means that item score $d_{k,j}$ discriminates individuals poorly, while the low value of the factor $W^b_{(k,j)}$ means that $d_{k,j}$ is highly correlated with other item scores in k-th dimension, thus reducing the effect of redundancy (Betti, 2017). Weights are normalized to unity by setting:

$$w_{k,j} = \frac{w_{k,j}}{\sum_{j=1}^{m_k} w_{k,j}}\,. \tag{5}$$

In the fourth step we calculate the QoL indicator. First, the sub-indicators in each dimension are calculated. For an i-th individual, aggregation over a set of item scores in a k-th dimension ($k = 1, 2, \ldots, K$) is given by formula (Betti et al., 2016; Betti, 2017):

$$S_{k,i} = \sum_{j=1}^{m_k} w_{k,j}\, d_{k,j,i}\,, \tag{6}$$

where: $d_{k,j,i}$ – the value of j-th item score in the k-th dimension for the i-th individual,
$w_{k,j}$ – normalised weight for j-th item score in the k-th dimension,
m_k – the total number of items in the k-th dimension.

Next, an overall QoL indicator for the i-th individual is calculated as the mean of sub-indicators $S_{k,i}$:

$$S_i = \frac{1}{K}\sum_{k=1}^{K} S_{k,j}\,, \tag{7}$$

where K is the number of dimensions.

In the next step of the analysis, to gain a deeper insight into the subject matter, we try to explain the values of the indicator S by socio-economic and demographic factors.

1.3 Fractional outcome models

The aim of our research is to estimate a model with the dependent variable S ranged between 0 to 1, inclusive. To handle these data properly, one should take the bounded nature of the response into account. A comprehensive survey of the models and estimation methods suitable to deal with fractional response variables can be found in (Carrasco et al., 2014; Ramalho et al., 2011). The use of linear regression model can generate predictions outside the unit interval. Moreover, it is conceptually flawed to assume normal distribution for a response variable in the [0, 1] range. As Papke and Wooldridge (1996) pointed out, the drawbacks of a linear model for fractional data are analogous to the drawbacks to a linear probability model for binary data. One way to handle this for response variables' values belonging to a closed unit interval is to apply a fractional response model (FRM). Papke and Wooldridge introduced such a model in a paper in 1996 (Papke and Wooldridge, 1996).

Fractional regression is a model of the mean of the dependent variable y conditional on covariates $\boldsymbol{x}$, which we denote by $E(y|\boldsymbol{x}) = \mu_x$ Because y is in the [0, 1] interval, to ensure that μ_x also belongs to it [0, 1], in an FRM it is assumed that:

$$\mu(\boldsymbol{x}_i) = G(\boldsymbol{x}_i^{'}\boldsymbol{\beta}), \tag{8}$$

where: $\mu(\boldsymbol{x}_i) = E(y_i|\boldsymbol{x}_i)$
$G(^\bullet)$ is a known function with $0 < G(z) < 1$ for $z \in R$,
$\boldsymbol{x}_i$ is a vector of explanatory variables representing the characteristics of individual i,
$\boldsymbol{\beta}$ is a vector of parameters to be estimated.

Typically, non-linear functional forms used for G are chosen to be a cumulative distribution function (cdf). The two most popular examples used in FRM are the logistic function $(z) = \Lambda(z) = \frac{\exp(z)}{1+\exp(z)}$ and $G(z) = \Phi(z)$, where Φ is the standard normal cumulative distribution function. Note that G is the inverse function for the so-called link function that specifies the link between the random and systematic components. It indicates how the expected value of the response variable relates to the linear predictor of explanatory variables. For a discussion on link functions in fractional outcome models, see (Smithson and Verkuilen, 2006; Ramalho et al., 2011).

The nonlinear estimation of an FRM's parameters is performed via maximization of the log-likelihood. The Bernoulli log-likelihood function for the FRM is of the form:

$$lnL = \textstyle\sum_{i=1}^{n} v_i y_i \, lnG(\boldsymbol{x}_i^{'}\boldsymbol{\beta}) + v_i(1 - y_i)\ln\left(1 - G(\boldsymbol{x}_i^{'}\boldsymbol{\beta})\right), \tag{9}$$

where: y_i is the dependent variable for the i-th individual,
x_i are the covariates for individual i, and
v_i denotes sample weight of the i-th individual,
n is the sample size.

To obtain robust estimation of an FRM, the quasi-maximum likelihood (QML) is used. It is important that the QML estimator does not require full distributional assumption of the dependent variable for consistency. The only information that it needs is the conditional mean to be correctly specified for consistent parameter estimates. The QML estimator of $\boldsymbol{\beta}$ is consistent and asymptotically normal, regardless the distribution of the dependent variable, conditional on the predictors (Papke and Wooldridge, 1996). To test the correct link specification of the conditional mean function, Ramsey's RESET test, more common in econometrics literature, can be applied.

The partial effects in an FRM of a given variable, say X_j, are given by:

$$\frac{\partial E(y_i|\boldsymbol{x}_i)}{\partial x_{ji}} = \beta_j g(\boldsymbol{x}_i^{'}\boldsymbol{\beta}), \tag{10}$$

where: $g(\boldsymbol{x}_i^{'}\boldsymbol{\beta}) = \dfrac{\partial G(\boldsymbol{x}_i^{'}\boldsymbol{\beta})}{\partial(\boldsymbol{x}_i^{'}\boldsymbol{\beta})}$,

x_{ji} is a value of j-th explanatory variable for i-th individual.

Hence, the significance and the direction of the marginal effects may be analyzed simply by examining the significance and sign of β_j (Ramalho and Vidigal da Silva, 2013).

FRMs have been applied in a variety of disciplines, including the social sciences, health sciences and economics. To see how FRMs have been used, see (Cardoso et al., 2010; Czarnitzki and Kraft, 2004; Flores et al., 2015) to name a few.

A beta regression models (BRMs) may be a valid alternative to FRMs. Though the beta distribution has been known in statistics for about a century, the research that has been done on BRM is relatively recent. BRMs have gained traction thanks to their flexibility for modelling dependent variables ranging to the open unit interval. For papers introducing these models, see (Paolino, 2001; and Ferrari and Cribari-Neto, 2004). BRMs are applied across variety fields, including finance, medicine, psychology and economics, for examples, see (Grzybowska and Karwański, 2015; Karwański et al., 2015; Rogers et al., 2012; Smithson and Verkuilen, 2006; Zanin, 2017).

BRMs are based on the assumption that the dependent variable y is beta-distributed and that its mean is related to a set of explanatory variables through a linear predictor with unknown coefficients and a link function. They also include a precision parameter which may be constant or depend on a set of regressors through a scale-link function as well. The density of a beta-distributed dependent variable y conditional on covariates (explanatory variables) **x** can be written as (Ferrari and Cribari-Neto, 2004):

$$f(y, \mu_x, \psi) = \frac{\Gamma(\psi)}{\Gamma(\mu_x\psi)\Gamma((1-\mu_x)\psi)}\, y^{\mu_x\psi-1}\,(1-y)^{(1-\mu_x)\psi-1}, \tag{11}$$

where: $\mu_x = E(y|\boldsymbol{x})$ is the mean of the dependent variable y conditional on covariates **x**,
ψ scales the conditional variance according to:

$$\mathrm{Var}(y|\boldsymbol{x}) = \frac{\mu_x(1-\mu_x)}{1+\psi}. \tag{12}$$

The parameter ψ is known as the precision parameter[9] since, for fixed μ_x, the larger the ψ, the smaller the conditional variance of y. Note also that conditional variance of y is a function of μ_x which renders the regression model based on this parameterization naturally heteroskedastic (Cribari-Neto and Zeileis, 2010).

A BRM is a model of $\mu_x = \mathrm{E}(y|\boldsymbol{x})$. It is appropriate when y takes values from the (0, 1) interval to ensure that μ_x is also in (0, 1), link function for the conditional mean is used. As for FRMs, it is assumed that the mean μ_x is given by Formula (8), thus the partial effects in the BRM are given by (10).

According to (11), the log-likelihood function is of the form:

$$lnL = \textstyle\sum_{i=1}^{n} v_i\,(ln\Gamma(\psi) - ln\Gamma(\mu_x\psi) - ln\Gamma((1-\mu_x)\psi) + (\mu_x\psi - 1)ln_i + ((1-\mu_x)\psi - 1)ln(1-y_i)), \tag{14}$$

where: y_i is the dependent variable for the i-th individual,
μ_x is given by Formula (8),
ψ is the precision parameter,
n is the sample size,

9 Precision parameter may be constant or depend on set of regressors through a scale-link function (Smithson and Verkuilen, 2006).

v_i denotes the sample weight of the i-th individual.
Parameter estimation is performed by maximum likelihood (ML), simply replacing μ_x with (8).

In our study we try to explain the values of fractional variable *S*, being the QoL indicator, by explanatory variables using a FRM and a BRM. All computations are performed using STATA 14. In order to ensure a representative character on the national scale and in the identified classification cross-sections, we use a sample weight for each individual.

To compare a goodness of fit of the models to the data, we calculated simple measures by taking the observed (y) value minus its corresponding predicted conditional mean ($\hat{y}$). A lot of measures based on such differences can be obtained. The goodness of fit of models in our research was evaluated using the root mean square error (RMSE) and the mean absolute error (MAE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n}}, \tag{15}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|. \tag{16}$$

These are common statistics used to assess models. Large values indicate a poor fit.

2 RESULTS AND DISCUSSION

As described in Section 1, we considered five dimensions encompassing 12 items. All items were converted by membership function (1) into item scores. To calculate weights for them, we applied the procedure *mdepriv*[10] – a Stata command written by Pi Alperin and Van Kerm (2014). These weights were used to calculate the QoL indicators given by Formulas (2)–(5). Table 1 reports descriptive statistics for the overall summary QoL indices *S* and the indices S_1, S_2, S_3, S_4, S_5 corresponding to the five dimensions.

Table 1 Descriptive statistics for overall and dimension-specific QoL indices

Descriptive statistic	Overall (S)	Social (S_1)	Material (S_2)	Environmental (S_3)	Health-related (S_4)	Self-assessment (S_5)
Mean	0.3917	0.4002	0.3993	0.3732	0.3978	0.3914
Standard deviation	0.1754	0.2575	0.2325	0.1975	0.2290	0.2203
Median	0.3740	0.4139	0.4268	0.3463	0.3852	0.3483
Maximum	1	1	1	1	1	1
Minimum	0	0	0	0	0	0
Skewness	0.6202	0.7283	0.5568	0.5841	0.5280	0.5049
Kurtosis	3.3934	2.8453	2.8686	3.2607	2.7736	2.6840

Source: Authors' calculations

As shown in Table 1, Poles, on average, were best satisfied in relationships with other people (social aspects) and least satisfied with environmental aspects. Mean values of all QoL indices stand at about 0.4 with a standard deviation of about 0.2. A minimum of 0 and a maximum of 1 for the indices *S* and S_1, S_2, S_3, S_4, S_5 means that there existed individuals who were very dissatisfied with each aspect of life and others who were very satisfied. All QoL indices exhibit slightly positive asymmetry, indicating distributions with

[10] We found weak or moderately strong positive correlations among all pairs of item scores in a given dimension.

an asymmetric tail extending toward more positive values. Skewness values close to 0 and kurtosis values close to 3 indicate that distributions of the QoL indices do not differ much from the normal distribution.

The next part of the study explores statistical significance and the impact of various socio-demographic factors on the QoL indicator (*S*). As described in Section 1, we applied the FRMs and BRMs using the logit and the probit link function. Because beta-regression is designed to model values on the interval (0,1) we coded values 0 as 0.0001 and values 1 as 0.9999. There were 3 observations with a value of 0 and 60 observations with a value of 1. We considered a number of socio-economic and demographic variables that can shed light on QoL. Akaike (AIC) and Bayesian (BIC) information criteria were used to compare alternative models with various sets of explanatory variables. See the Appendix for a description of these variables.

Table 2 reports the estimation results for the FRMs and the BRMs with logit and probit variants. We found that for both BRMs, AIC and BIC information criteria clearly indicate the choice of explanatory variables presented in Table 2, while results for the FRMs are not so explicit – the AIC criterion prefers the same set of variables used in the BRMs, but the BIC criterion prefers the set of variables without the variable describing the class of respondents' place residence. To compare the results obtained with the various models, we used the same explanatory variables in each of them.

Table 2 Estimates of fractional regression and beta regression models

Variable	FRM with logit link function		FRM with probit link function		BRM with logit link function		BRM with probit link function	
	b	*S(b)*	*b*	*S(b)*	*b*	*S(b)*	*b*	*S(b)*
Age	–0.0293***	0.0020	–0.0181***	0.0013	–0.0301***	0.0023	–0.0187***	0.0014
Age2	0.0003***	2E-5	0.0002***	1E-05	0.0003***	2E-05	0.0002***	1E-05
Disability	–0.1833***	0.0186	–0.1126***	0.0113	–0.1778***	0.0207	–0.1094***	0.0126
Association membership	0.1069***	0.0190	0.0665***	0.0118	0.1081***	0.0218	0.0673***	0.0136
Civil state								
Married	Ref.	–	Ref.	–	Ref.	–	Ref.	–
Unmarried	–0.0293	0.0211	–0.0180	0.0131	–0.0261	0.0251	–0.0160	0.0156
Widowed	–0.1383***	0.0250	–0.0839***	0.0153	–0.1283***	0.0258	–0.0779***	0.0158
Divorced/separated	–0.1118***	0.0344	–0.0693***	0.0211	–0.1081***	0.0355	–0.0669***	0.0218
Education								
1 (primary)	Ref.	–	Ref.	–	Ref.	–	Ref.	–
2 (basic vocational)	0.1935***	0.0208	0.1184***	0.0127	0.1886***	0.0209	0.1155***	0.0128
3 (secondary)	0.2813***	0.0213	0.1730***	0.0115	0.2929***	0.0224	0.1803***	0.0137
4 (higher)	0.4409***	0.0235	0.2723***	0.0144	0.4499***	0.0251	0.2781***	0.0154
Class of place of residence								
Town bigger then 20 000[12]	Ref.	–	Ref.	–	Ref.	–	Ref.	–
Very small town	0.0918***	0.0204	0.0568***	0.0126	0.0942***	0.0227	0.0583***	0.0135
Village	0.0486***	0.0152	0.0300***	0.0094	0.0632***	0.0181	0.0390***	0.0112
Regions								
Central	Ref.	–	Ref.	–	Ref.	–	Ref.	–
South	0.0776***	0.0196	0.0485***	0.0121	0.0685***	0.0204	0.0429***	0.0126
East	–0.0721***	0.0185	–0.0444***	0.0114	–0.0742***	0.0193	–0.0456***	0.0119

[12] Differences between very big towns, big towns, medium-sized towns and small towns were not statistically significant even at the 0.1 level, therefore we used aggregation to towns bigger than 20 000.

Table 2 Estimates of fractional regression and beta regression models (continuation)

Variable	FRM with logit link function		FRM with probit link function		BRM with logit link function		BRM with probit link function	
	b	*S(b)*	*b*	*S(b)*	*b*	*S(b)*	*b*	*S(b)*
Northwest	0.0775***	0.0217	0.0484***	0.0134	0.0952***	0.0270	0.0593***	0.0167
Southwest	–0.0022	0.0232	–0.0012	0.0144	–0.0062	0.0246	–0.0038	0.0152
North	0.1254***	0.0215	0.0778***	0.0133	0.1635***	0.0257	0.1015***	0.0160
Household type								
MC without children	Ref.	–	Ref.	–	Ref.	–	Ref.	–
MC with 1 child	–0.0394*	0.0214	–0.0246*	0.0132	–0.0307	0.0255	–0.0193	0.0158
MC with 2 children	–0.0091	0.0227	0.0059	0.0141	–0.0165	0.0259	–0.0105	0.0161
MC with 3+ children	–0.0915***	0.0268	–0.0568***	0.0166	–0.1014***	0.0287	–0.0630***	0.0177
Single-parent	–0.1966***	0.0274	–0.1214***	0.0169	–0.2086***	0.0279	–0.1289***	0.0172
Multi-family	0.0059	0.0250	0.0037	0.0155	0.0199	0.0291	0.0123	0.0180
One-person	–0.0583**	0.0267	–0.0365**	0.0165	–0.0802***	0.0278	–0.0500***	0.0172
Non-family	–0.1390*	0.0764	–0.0973*	0.0471	–0.1672**	0.0750	–0.1047**	0.0462
The socio-economic group								
Employees	Ref.	–	Ref.	–	Ref.	–	Ref.	–
Entrepreneurs	–0.0152	0.0307	–0.0090	0.0191	–0.0332	0.0343	–0.0202	0.0213
Farmers	–0.0136	0.0228	–0.0084	0.0141	–0.0337	0.0240	–0.0208	0.0149
Retirees	–0.0272*	0.0166	–0.0167*	0.0103	–0.0315*	0.0179	–0.0194*	0.0110
Pensioners	–0.0742**	0.0301	–0.0458**	0.0184	–0.0684**	0.0352	–0.0424**	0.0216
Living on unearned sources	–0.1647***	0.0430	–0.1027***	0.0263	–0.1663***	0.0510	–0.1038***	0.0311
Constant	0.0343	0.0618	0.0210	0.0383	0.0860	0.0709	0.0538	0.0440
Scale parameter	–	–	–	–	1.8476***	0.0254	1.8476***	0.0254

Note: *b* are estimates, *S*(*b*) – their standard errors. All standard errors are robust (with heteroscedasticity-robust asymptotic variance). * means statistical significance at 0.10, ** – statistical significance at 0.05, *** – statistical significance at 0.01.
Source: Authors' calculations

It is evident that most of the explanatory variables are statistically significant at the 0.01 level. In addition, almost all coefficients in all models have the same sign and statistical significance. This means that the impact of the socio-economic and demographic variables on quality of life can be interpreted in the same way for the FRMs and the BRMS. All of the interpretations presented here were made under the assumption of *ceteris paribus*.

We have determined that age had a negative sign while its squared term had a positive sign, implying a U-shaped effect. In other words, people tend to be more satisfied with life when they are younger and older than when they are middle-aged. A number of other researchers have reached the same conclusion (Blanchflower and Oswald, 2008; Sanfey and Teksoz, 2007; Pierewan and Tampubolon, 2015). Our investigation indicates that Poles were the least satisfied with their life at around age 54, a higher age than the turning point for most developed countries, which is typically in the forties (Blanchflower and Oswald, 2008).

As in other studies, we found that being a member of a political party or union has a positive effect on QoL, while being disabled has a negative one (Wang and VanderWeele, 2011; Christoph, 2010).

A widowed individual is likely to be less satisfied than one who is married. The same can be said of those who are divorced or separated. This confirms the findings of other studies (Sanfey and Teksoz, 2007; Pierewan and Tampubolon, 2015).

Education may be one of the most important and consistent determinants of QoL. As a human capital indicator, this covariate predicts the well-being. A number of studies have also investigated the relation between education and QoL (Betti et al., 2016; Sanfey and Teksoz, 2007). In general, the impact of education on satisfaction with one's life is ambiguous across the studies we analysed: there is no clear correlation. Malešević Perović (2010) found a positive correlation, while Clark and Oswald (1994) uncovered a negative one. Still, others have observed a mixed correlation: Betti et al. (2016) found that people with a middle level of education were the most satisfied. Finally, Sanfey and Teksoz (2007) stated that there is no correlation between happiness and education in transition countries. In our study, QoL tended to rise alongside the level of education.

Also in line with other studies (Gerdtham and Johannesson, 2001; Requena, 2016), our results show that living in the countryside or in towns with less than 20,000 inhabitants improved the perception of QoL. Requena (2016) observed that in wealthier countries, rural living standards are high enough to create a higher level of subjective well-being; while in less developed countries the rural environment cannot compete with urban resources for creating subjective well-being. Also in agreement with other research, we found territorial differences in the QoL (Cracolici et al., 2014; Malešević Perović, 2010). Comparing the Central Region, where Poland's capital city Warsaw is located, the South, the Northwest and the North exhibit significantly better QoL, while the East and the Southwest were perceived as significantly worse and not significantly worse, respectively.

With regard to type of household, we stated that the composition of the household affected the perception of QoL. Married couples with three children, single-parent families, non-family households perceived their situation as significantly worse than married couples. In this respect, we did not find a significant difference between married couples and the remaining types (i.e. married couples with one child and with two children and multi-family households). The impact of the composition of the household on subjective well-being has been confirmed by many studies. For example, Cracolici et al. (2014) found that couples with no children were better off than others, while Betti et al. (2016) found that one-person households were in a worse situation than others.

Our results show the impact of a socio-economic group identified on the basis of the household's main source of income. Others reported similar findings on the influence of socio-economic group membership on a subjective perception of QoL (Cracolici et al., 2014; Wang and VanderWeele, 2011). Setting employees as the reference group, we found retirees, pensioners and those living on unearned sources other than retirement pay and pension to be in a significantly worse situation, while the self-employed and farmers exhibited not significantly worse position. The members of households living on unearned sources other than retirement pay and pensions were often the unemployed and poor. Such households generally assess various aspects of life with more pessimism than others. Sen (1997) mentioned a variety of reasons that unemployment may impact the QoL, including a lack of purpose in life, a lower social status and sense of self-esteem and a reduced sense of freedom and financial control.

Unlike the studies carried out for data from various European countries (Corazzini et al., 2012; Pierewan and Tampubolon, 2015), we found that in our study, gender does not reveal different patterns in explaining QoL.

Because this study is the first to explain QoL through the application of fractional outcome models, we considered various types of such models. As previously stated, the results concerning the estimates of significance and the impact of socio-economic and demographic variables obtained by the models considered in our study are very similar. In the next step we compared the models' goodness of fit. The predictive accuracy of the models is assessed using two performance measures: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Models with lower RMSE and MAE more accurately estimate the QoL indicator.

Table 3 Values of goodness of fit measures

Goodness of fit measure	FRM with logit link function	FRM with probit link function	BRM with logit link function	BRM with probit link function
RMSE	0.1748	0.1748	0.1756	0.1756
MAE	0.1396	0.1396	0.1406	0.1406

Source: Authors' calculations

The results reported in Table 3 show that the FRMs exhibit both RMSE and MAE only slightly better than the respective errors of all the BRMs. It should be also stressed that the Ramsey's RESET test reveals no misspecification of the conditional mean function in all estimated models. Thus, it cannot be determined to what extent one model is superior to another.

CONCLUDING REMARKS

This study has examined a new methodological framework for assessing the subjective perception of life by using methods of fuzzy set theory proposed by Betti (Betti et al., 2016; Betti, 2017). The main contribution of this analysis is its application of fractional outcome models to explain the quality of life through various socio-economic and demographic factors.

The data employed for the analysis came from the 'Social Diagnosis' survey conducted in 2015, a good deal of which was devoted to aspects of personal life. According to Betti's approach, the ordered data on subjective assessments were converted by a membership function into a [0,1] interval and then the synthetic QoL indicator encompassing all the aspects of life under consideration was computed. Because all of the QoL indicator values lay in the unit interval, we proposed to explain them using fractional outcome models. We applied a fractional regression model proposed by Papke and Wooldridge (1996) and a beta regression model developed by Ferrari and Cribari-Neto (2004). We included various socio-economic variables and demographic factors as explanatory variables: age, gender, education, civil status, disability, association membership, place of residence, household type and main source of income. We found that the QoL was U-shaped in age, minimizing around the age of 54. Furthermore, the perception of QoL increases with education, association membership, and decreases with disability, urbanisation, and being widowed or divorced. Results of our estimation indicate that the demographic composition of the household, region of residence and source of income all had a statistically significant impact. Our findings are largely in line with other studies.

It should be stressed that our study omits sociological nuances of the definition of 'quality of life' concept. Our goal is to demonstrate the potential for using modern methods to identify factors affecting the multidimensional indicator of QoL. The application of fractional outcome models has many advantages. Such models allow the assessment of whether given socio-economic and demographic factor is associated with response variable bounded by 0 and 1 while controlling the outcomes overlapping associations with other explanatory variables. Also, their ability to capture non-linearities is an important advantage.

We hope that our study with using fractional outcome models approach can provide some insight into the subjective perception of the quality of life. We plan various extensions of our study. Future research could apply panel data models for controlling unobserved heterogeneity of individuals and monitoring changes of QoL over time.

References

BETTI, G. AND VERMA,V. Measuring the degree of poverty in a dynamic and comparative context: A multidimensional approach using fuzzy set theory. *Proceedings of the ICCS-VI.* Lahore, Pakistan, 1999, 11, pp. 289–301.

BETTI, G. AND VERMA, V. Fuzzy measures of the incidence of relative poverty and deprivation: a multi-dimensional perspective. *Statistical Methods and Applications*, 2008, 17(2), pp. 225–250.

BETTI, G., SOLDI, R., TALEV I. Fuzzy multidimensional indicators of quality of life: The empirical case of Macedonia. *Social Indicators Research*, 2016, 127(1), pp. 39–53.

BETTI, G. Fuzzy measures of quality of life in Germany: a multidimensional and comparative approach. *Quality and Quantity*, 2017, 51(1), pp. 23–34.

BLANCHFLOWER, D. G. AND OSWALD, A. J. Is well-being U-shaped over the life cycle? *Social Science & Medicine*, 2008, 66, pp. 1733–1749.

CARDOSO, A. R., FONTAINHA, E., MONFARDINI, C. Children's and parents' time use: empirical evidence on investment in human capital in France, Germany and Italy. *Review of Economics of the Household*, 2010, 8, pp. 479–504.

CARRASCO, J. M. F., FERRARI, S. L. P., ARELLANO-VALLE, R. B. Errors-in-variables beta regression models. *Journal of Applied Statistics*, 2014, 41(7), pp. 1530–1547.

CERIOLI, A. AND ZANI, S. A fuzzy approach to the measurement of poverty. In: DAGUM, C., ZENGA M., eds. *Income and Wealth Distribution, Inequality and Poverty.* Berlin: Springer, 1990.

CHELI, B. AND LEMMI, A. A totally fuzzy and relative approach to the multidimensional analysis of poverty. *Economic Notes*, 1995, 24, pp. 115–134.

CHRISTOPH, B. The relation between life satisfaction and the material situation: A re-evaluation using alternative measures. *Social Indicators Research*, 2010, 98(3), pp. 475–499.

CLARK, A. E. AND OSWALD, A. J. Unhappiness and Unemployment. *Economic Journal*, 1994, 104(424), pp. 648–659.

CORAZZINI, L., ESPOSITO, L., MAJORANO, F. Reign in hell or serve in heaven? A cross-country journey into the relative vs absolute perceptions of wellbeing. *Journal of Economic Behavior & Organization*, 2012, 81(3), pp. 715–730.

CRACOLICI, M. F., GIAMBONA, F., CUFFARO, M. Family structure and subjective economic well-being: some new evidence. *Social Indicators Research*, 2014, 118, pp. 433–456.

CRIBARI-NETO, F. AND ZEILEIS, A. Beta Regression in R. *Journal of Statistical Software*, 2010, 34(2), pp. 1–24.

CZAPIŃSKI, J. Indywidualna jakość i styl życia. *Contemporary Economics*, 2015, 9/4, pp. 200–331.

CZAPIŃSKI, J., PANEK, T., et al. Diagnoza Społeczna 2015. *Warunki i Jakość Życia Polaków – Raport.* Warsaw, 2015, available at the website: <*www.diagnoza.com*>.

CZARNITZKI, D. AND KRAFT, K. Firm leadership and innovative performance: evidence from seven EU countries. *Small Business Economics*, 2004, 22, pp. 325–332.

DE BATTISTI, F., MARASINI, D., NICOLINI, G. A measure of job satisfaction by means of fuzzy set theory. *Statistica Applicata – Italian Journal of Applied Statistics*, 2015, 23(3), pp. 361–374.

DESAI, M. AND SHAH, A. An econometric approach to the measurement of poverty. *Oxford Economic Papers*, 40(3), 1988, pp. 505–522.

DUDEK, H. AND SZCZESNY, W. Subjective perception of quality of life – multidimensional analysis based on fuzzy sets approach. *Wrocław University of Economics Research Papers, Quality of Life, Human and Ecosystem Well-being*, 2016, 435, pp. 55–68.

DUDEK, H. AND SZCZESNY, W. Zastosowanie funkcji przynależności w analizie subiektywnego postrzegania jakości życia. *Ekonometria*, 2015, 50(40), pp. 62–78.

EUROSTAT. *Quality of life indicators – measuring quality of life* [online]. Eurostat, 2017. <*http://ec.europa.eu/eurostat/statistics-explained/index.php/Quality_of_life_indicators_-_measuring_quality_of_life*>.

FERRARI, S. L. P. AND CRIBARI-NETO, F. Beta Regression for Modeling Rates and Proportions. *Journal of Applied Statistics*, 2004, 31(7), pp. 799–815.

FILIPPONE, A., CHELI, B., D'AGOSTINO, A. Addressing the interpretation and the aggregation problems in totally fuzzy and relative poverty measures. *ISER Working Paper Series number 2001–22*, University of Essex, 2001.

FLORES, G., INGENHAAG, M., MAURER, J. An anatomy of old-age disability: Time use, affect and experienced utility. *Journal of Health Economics*, 2015, 44, pp. 150–160.

FUWA, N., ITO, S., KUBO, K., KUROSAKI, T., SAWADA, Y. Gender discrimination, intrahousehold resource allocation, and importance of spouses' fathers: evidence on household expenditure from rural India. *Developing Economies*, 2006, 44(4), pp. 398–439.

GERDTHAM, U.-G. AND JOHANNESSON, M. The relationship between happiness, health, and socioeconomic factors: results based on Swedish microdata. *Journal of Socio-Economics*, 2001, 30(6), pp. 553–557.

GLATZER, W. Challenges for Quality of Life. In: GLATZER, W., VON BELOW, S., STOFFREGEN, M., eds. *Challenges for Quality of Life in the Contemporary World.* Dordrecht, Boston, London: Kluwer, 2004.

GRZYBOWSKA, U. AND KARWAŃSKI, M. Application of Mixed Models and Families of Classifiers to Estimation of Financial Risk Parameters. *Quantitative Methods in Economics*, 2015, 16(1), pp. 108–115.

GUIO, A.-C. What can be learned from deprivation indicators in Europe? *Eurostat methodologies and working paper*, Luxembourg: Eurostat, 2009.

HOE, J., ORRELL, M., LIVINGSTON, G. Quality of life measures in old age. In: ABOU-SALEH, M. T., KATONA, C., KUMAR, A., eds. *Principles and practice of geriatric psychiatry.* London: Wiley, 2011.

KARWAŃSKI, M., GOSTKOWSKI, M., JAŁOWIECKI, P. Loss given default modeling. *Journal of Risk Model Validation*, 2015, 9(3), pp. 23–40.

LAZIM, M. A. AND OSMAN, M. T. A. A new Malaysian quality of life index based on fuzzy sets and hierarchical needs. *Social Indicators Research*, 2009, 94(3), pp. 499–508.

MALEŠEVIĆ PEROVIĆ, L. M. Life satisfaction in Croatia. *Croatian Economic Survey*, 2010, 12(1), pp. 45–81.

NEFF, D. Fuzzy set theoretic applications in poverty research. *Policy and Society*, 2013, 32, pp. 319–331.

PANEK, T. Multidimensional approach to poverty measurement: Fuzzy measures of the incidence and the depth of poverty. *Statistics in Transition*, 2010, 11(2), pp. 361–379.

PAOLINO, P. Maximum likelihood estimation of models with beta-distributed dependent variables. *Political Analysis*, 2001, 9(4), pp. 325–346.

PAPKE, L. E., WOOLDRIDGE, J. M. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 1996, 11, pp. 619–632.

PHILLIPS, D. *Quality of Life: Concept, Quality, Practice.* London: Routledge, 2006.

PI ALPERIN, M. N., VAN KERM, P. MDEPRIV: Stata module to compute synthetic indicators of multiple deprivation. *Statistical Software Components S457806*, Boston College Department of Economics, 2014.

PIEREWAN, A. C. AND TAMPUBOLON, G. Happiness and health in Europe: A multivariate multilevel model. *Applied Research in Quality of Life*, 2015, 10(2), pp. 237–252.

RAMALHO, J. J. S. AND VIDIGAL DA SILVA, J. Functional form issues in the regression analysis of financial leverage ratios. *Empirical Economics*, 2013, 44(2), pp. 799–831.

RAMALHO, E. A, RAMALHO, J. J. S., MURTEIRA, J. M. R. Alternative estimating and testing empirical strategies for fractional regression models. *Journal of Economic Surveys*, 2011, 25(1), pp. 19–68.

REQUENA, F. Rural–urban living and level of economic development as factors in subjective well-being. *Social Indicators Research*, 2016, 128(2), pp. 693–708.

ROGERS, J. A, POLHAMUS, D., GILLESPIE, W. R. et al. Combining patient-level and summary-level data for Alzheimer's disease modeling and simulation: a beta regression meta-analysis. *Journal of Pharmacokinetics and Pharmacodynamics*, 2012, 39(5), pp. 479–498.

ROJO-PÉREZ, F., FERNÁNDEZ-MAYORALAS, G., RODRÍGUEZ-RODRÍGUEZ, V. Global Perspective on Quality in Later Life. In: GLATZER, W., CAMFIELD, L., MØLLER, V., ROJAS M., eds. *Global Handbook of Quality of Life. International Handbooks of Quality-of-Life.* Dordrecht: Springer, 2015.

SANFEY, P. AND TEKSOZ, U. Does transition make you happy? *Economics of Transition*, 2007, 15(4), pp. 707–731.

SEN, A. Inequality, unemployment and contemporary Europe. *International Labour Review*, 1997, 136(2), pp. 155–72.

SMITHSON, M. AND VERKUILEN, J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 2006, 11(1), pp. 54–71.

STIGLITZ, J. E., SEN, A., FITOUSSI, J. *Report by the commission on the measurement of economic performance and social progress.* 2009, retrieved from: <*www.stiglitz-sen-fitoussi.fr*>.

WANG, P. AND VANDERWEELE, T. J. Empirical research on factors related to subjective wellbeing of Chinese urban residents. *Social Indicators Research*, 2011, 101(3), pp. 447–459.

ZADEH, L. A. Fuzzy sets. *Information and Control*, 1965, 8, pp. 338–353.

ZANIN, L. The effects of various motives to save money on the propensity of Italian households to allocate an unexpected inheritance towards consumption. *Quality and Quantity*, 2017, 51(4), pp. 1755–1775.

APPENDIX

Table A1 Items in the individual questionnaire concerning respondent satisfaction with regard to particular areas and aspects of life[13]

Please, assess the specific areas of your life and state to what extent you are satisfied with them. Please, give your answers by crossing the box next to the appropriate digit for the given area of life. The specific digits mean:

1 – VERY SATISFIED
2 – SATISFIED
3 – RATHER SATISFIED
4 – RATHER NOT SATISFIED
5 – NOT SATISFIED
6 –VERY NOT SATISFIED
7 – not applicable

To what extent are you satisfied with:

1.	your relations with your close family members	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □
2.	the financial situation of your family	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □
3.	your relations with friends (a group of friends)	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □
4.	your health condition	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □
5.	your life achievements	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □
6.	the situation in the country	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □
7.	your housing conditions	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □
8.	the town/city you live in	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □
9.	your future prospects	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □
10.	your sex life	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □
11.	your education	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □
12.	the manner in which you spend your free time	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □
13.	your work	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □
14.	children	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □
15.	marriage	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □
16.	safety in your town/city of residence	1 □ 2 □ 3 □ 4 □ 5 □ 6 □ 7 □

Source: Own construction based on (Czapiński and Panek, 2015)

Table A2 List and description of explanatory variables

Variable	Description
Age	The individual's age
Age2	The individual's age squared
Female	1 if the individual is female
Civil state	Four groups of formal civil states are considered:
married	1 if married
unmarried	1 if unmarried
widowed	1 if widowed
divorced/separated	1 if divorced or separated
Education	The educational level achieved by the individual is classified as:
1 (primary)	primary or lower
2 (basic vocational)	basic vocational or lower-secondary

[13] All items of questionnaire can be found on the website: <*www.diagnoza.com*> (Czapiński and Panek, 2015).

Table A2 List and description of explanatory variables (continuation)

Variable	Description
3 (secondary)	secondary
4 (higher)	higher or post-secondary
Disability	1 if the individual is disabled
Association membership	1 if the individual is a member of any organization, party or clubs
Class of place of residence	The class of place of residence is divided into urban and rural areas, with urban areas further subdivided by resident size units:
Very big town	Towns over 500 000 residents
Big town	Towns with 200 000–500 000
Medium-sized town	Towns with 100 000–200 000 residents
Small town	Towns with 20 000–100 000 residents
Very small town	Towns up to 20 000 residents
Village	Rural areas
Regions	Regions are the first level NUTS regions of the European Union. They include corresponding second-level sub-regions:
Central	Łódź, Mazovia
South	Lesser Poland, Silesia
East	Lublin, Subcarpathian, Świętokrzyskie, Podlaskie
Northwest	Greater Poland, West Pomerania, Lubusz
Southwest	Lower Silesia, Opola Voivodeship
North	Kuyavian-Pomeranian, Varmia-Masuria, Pomerania
Household type	Household type was established on the basis of the number of families and biological family type
MC without children	married couples with no children
MC with 1 child	married couples with one child
MC with 2 children	married couples with two children
MC with 3+ children	married couples with three or more children
Single-parent	single-parent families
Multi-family	multi-family households
One-person	non-family one-person households
Non-family	non-family multi-person households
The socio-economic group	The socio-economic group is identified on the basis of the household's main source of income. The following groups of households are taken into account:
Employees	households where the sole or main (dominant) source of income is from gainful employment in the public or private sector and from performing home-based work or on the basis of agency agreements
Self-employed	households whose exclusive or main (prevailing) source of income is self-employment (other than from private farming)
Farmers	households where the sole or main (dominant) source of income is from a farm with agricultural land exceeding 1 ha (including users of plots up to 1 ha of agricultural land and owners of domestic animals but no agricultural land if the livestock is the sole or main source of income)
Retirees	households where the sole or main (dominant) source of income is a retirement pension
Pensioners	households where the sole or main (dominant) source of income is a form of disability welfare support
Living on unearned sources	households where the sole or main (dominant) source of income are sources other than paid work (except for retirement pension, disability benefit or other type of pension)

Source: Own construction