LOGIT and PROBIT Models in the Probability Analysis: Change in the Probability of Death of Celiac Disease Patients

Ondřej Šimpach[1] | *University of Economics, Prague, Czech Republic*

Abstract

It is estimated, that in the Czech Republic live about 40 000–50 000 people suffering from celiac disease, which is a disease of gluten intolerance. At the beginning of the independent Czech Republic, the life expectancy at birth of these people was quite low, because just in this period detailed diagnosis of this disease came from abroad. With an increasing age the probability of death of these people grew faster than that of total population. The aim of this study is to analyse the probability of death of x-year old persons during next five years after the general medical examination in 1990 and 1995. Both analyses will be solved using LOGIT and PROBIT models and the hypothesis claiming, that probability of death of x-year old person suffering from celiac disease decreased few years after the gaining of new medical knowledge from abroad will be confirmed or refused.

Keywords

Probability of death, celiac disease, LOGIT, PROBIT, discrete dependent variables

JEL code

C35, C40, I19

INTRODUCTION

Medicine provides to people new knowledge about the diagnosis of specific diseases. By the end of the past regime in the Czech Republic, a detailed diagnosis of disease of the small intestine mucosa, professionally known as celiac disease (see Společnost pro bezlepkovou dietu / Society for the gluten-free diet) was not known. At present, it is estimated, that in the Czech Republic there are about 40 000–50 000 inhabitants, who suffer from this disease. Unfortunately, only about 10–15% of all patients are under medical supervision (see Poradenské centrum pro celiakii a bezlepkovou dietu / Advice center for celiac disease and gluten-free diet). In the past, the people suffering from this disease did not have the sufficient information about correct nutrition and thus they could not eliminate the consequences caused by

[1] University of Economics, Prague, Winston Churchill Sq. 4, 130 67 Prague 3, Czech Republic. E-mail: ondrej.simpach@vse.cz, phone: (+420)224095273, (+420)737665461.

normal diet. Normal diet, which also includes food containing gluten, etched the small intestinal mucosa of these people and most of them had not been able to live as long ordinary people (see e.g. Logan et al., 1989) did. It is important to note, that we do not have separate statistics of people who died of this disease. Deaths are recorded in the summary group "XI - Diseases of the Digestive System", published by the Czech Statistical Office (CZSO).

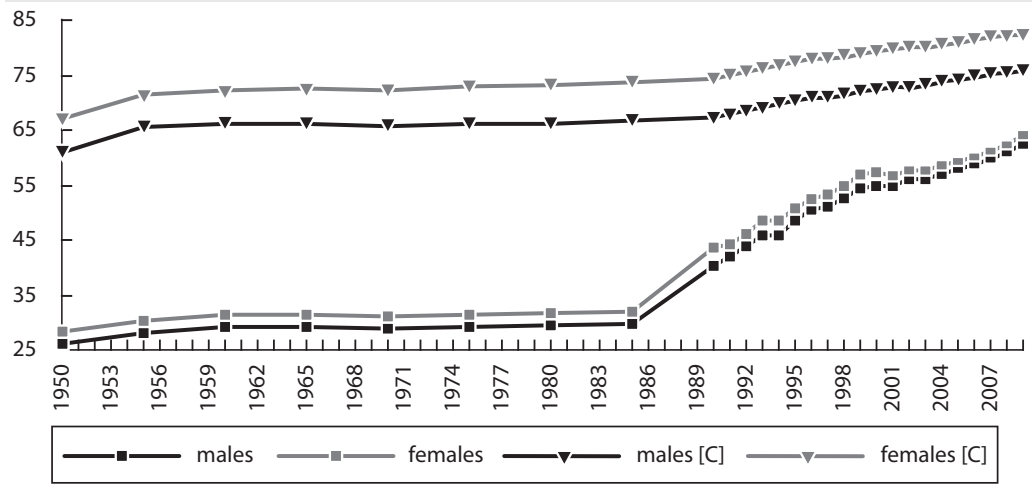
In the Czech Republic, there live about tens of thousands of people who suffer from other diseases. But the diagnoses of these diseases are known for a long time. Taking the developed drugs eliminates partly or completely the consequences of these diseases. This study will make specific probabilistic analysis of death of x-year old individuals. Using the LOGIT and PROBIT models the probabilities of death of x-year old persons suffering from celiac disease during next 5 years after the general medical examination in 1990 and the probabilities of death of x-year old persons suffering from any of other diseases during next 5 years after the general medical examination will be calculated in 1990. Diagnoses of these other diseases were known for longer time than celiac. These are mainly the following:

- Diabetes,
- High blood cholesterol (greater than 5 mmol / l),
- High blood pressure (more than 150 SYS and 95 DIA),
- Bronchial asthma,

  or other diseases.

These diseases were selected because they are frequently represented in the population of the Czech Republic. The probabilities of death will be calculated again for both groups (the group of people suffering from celiac disease and the group of people suffering from any other disease), but the data will be shifted by 5 years forward in time, (i.e. taken in 1995). Based on the results from these other estimates, the hypothesis claiming, that the probabilities of death of x-year old persons suffering from celiac disease decreased in time will be acknowledged. The evolution of life expectancy at birth for males and females suffering from celiac disease is shown in Figure 1. The time series labelled "males" and "females" show the actual life expectancy at birth of total population, which is published by the CZSO. Time series "males_c" and "females_c" show the estimated life expectancy of people suffering from celiac disease. Values for the years 1950–1990 are given only in the 5-year time points, because for the purposes of this study they are not necessary.

Figure 1 Estimates of life expectancy at birth of males and females suffering from celiac disease in 1950–2009



Source: Czech Statistical Office, Šimpach (2011)

After the estimation of LOGIT and PROBIT parameters models of probabilities of deaths of x-year old persons will be graphically interpreted.

1 ASSUMPTIONS OF THE STUDY

For the experiment of nonlinear regression, applied in the first part of this study 243 observations of variables consisting of two samples were obtained. It is important to note, that this is not a representative selection for the application of standard methods of mathematical statistics. The selection was not taken at random. This is the only existing data matrix, obtained by own research. The data matrices, which are prepared by health insurance funds, do not have required form and all needed variables are not recorded. The data contains 118 observations of females and 125 observations of males. (This proportion is set by the fact that the proportion of girls at birth is approximately 0.485 and the proportion of boys at birth is approximately 0.515. For more information see e.g. Pavlík et al., 1986). The selection is obtained from 3 doctors (from 3 different parts of the Czech Republic), whose specialization were adult patients suffering from diseases of digestive system and any other diseases. Selection consists all individual invited in 1990 to general medical examination and their health status was checked in the future.

For consecutive experiment of nonlinear regression, applied in the second part of the study, other 245 observations of patients, consisting of two samples were obtained. The data contains 119 observations of females and 126 observations of males. It is a selection of patients invited in 1995 to the overall medical examination and their health status was checked in the future (but obtained from other 3 doctors from other 3 different parts of Republic). It is almost zero probability that some patients from the first sample are contained in the second sample.

The authors Spector and Mazzeo (1980) put together an example, where they estimated the probability with which a student will succeed in the exam. Based on this example probabilistic LOGIT and PROBIT models were created, which are currently used by many authors in their calculations and publications, such as Hoyos et al. (2010) or Yang and Raehsler (2005). For this study LOGIT and PROBIT models were compiled estimating the probabilities of death of x-year old persons during next five years after the general medical examination, if the disease of the digestive system (known as celiac disease now) was diagnosed for an individual and the probabilities of death of x-year old person during next five years after the general medical examination, if any other disease was diagnosed for a respective individual (e.g. diabetes, high blood cholesterol, high blood pressure, bronchial asthma or other).

Dependent variable Y is an alternative. The value of variable Y equals 1 → the person will die, or the value of variable Y equals 0 → the person will survive. In order to determine the values lying between the two extremes, LOGIT and PROBIT models will be applied where the dependent variable takes values from interval <0, 1>.

The variables in selected data matrices are the following:

- *AGE* – is the age of the person, invited to the general medicinal examination.
- *CIRD* – Constant Increased the Risk of Death, which was designed primarily for the purpose of this study and the calculation of this constant gives the Formula (1). The constant takes values from interval <5, 35>.
- *CEL* – is a binary variable, where the value "0" = a person does not suffer from celiac disease, but suffers from any of the other diseases listed above, or "1" = a person suffers from celiac disease.
- *DEATH_5* – is a binary variable, where the value "1" = a person during next 5 years after the general medicinal examination died as a result of the disease (celiac or other), or "0" = a person during next 5 years after the general medicinal examination did not die as a result of the disease (celiac or other).

As stated above, the constant increasing the risk of death was designed only for the purpose of this study and its calculation is based on Table 1.

Table 1 The responses to doctor during the general medical examination

	$Ш_1$	$Ш_2$	$Ш_3$
Smoker	no	occasionally	periodically
Black coffee	no	occasionally	periodically
Alcohol	no	occasionally	periodically
Sleeping	regularly	irregularly	very bad
Eating	regularly	irregularly	very bad

Source: Own construction

The doctor asked the patient few simple questions during the general medicinal examination. Based on the responses, the table below was created, where verbal responses were replaced by "0" and "1", where "0" = patient's response does not coincide with the word specified in the relevant cell and "1" = patient response coincides with the word specified in the relevant cell. In each row of the table can be only one value "1".

Table 2 The responses to doctor during an overall medical examinations in "0/1" format

	$Ш_1$	$Ш_2$	$Ш_3$
Smoker	$Ш_{i,j} = 0/1$	…	…
Black coffee	…		
Alcohol	…		
Sleeping	…		
Eating	…		

Source: Own construction

After recording the responses to Table 2 in the "0/1" format, there was prepared following formula:

$$CIRD = \left(1 \times \sum_{i=1}^{5} Ш_{i,1}\right) + \left(3.5 \times \sum_{i=1}^{5} Ш_{i,1}\right) + \left(7 \times \sum_{i=1}^{5} Ш_{i,3}\right). \quad (1)$$

This Constant can take values from interval <5, 35>, where the extreme value of 5 means, that the patient does not increase the risk of death because of its poor lifestyle and extreme value of 35 means, that the patient increases the risk of death in the worst way possible. Weights 1, 3.5 and 7 were set on the recommendation of the doctors, who provided data matrices. The general rule, arising from the literature, is not here.

Another study, but for case of USA prepared Rubio-Tapia et al. (2009), where they had better data matrices and used other statistical approach. Authors used data matrices including the age of testing and the age at death. Bigger and exhaustive sample allowed for the application of standard methods of mathematical statistics.

1.1 LOGIT model

From the assumptions listed by Christensen (1990) is the probability function:

$$P_i = E(Y = 1|\boldsymbol{X}_i) = \frac{1}{1 + e^{-(\beta_0 + \boldsymbol{\beta}' \boldsymbol{X}_i)}}, \quad (2)$$

modified for the purpose of this study to form:

$$P_i = E(Y = 1|\boldsymbol{X}_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 AGE + \beta_2 CIRD_i + \beta_3 CEL_i)}}, \quad (3)$$

where i is $(i)^{th}$ patient. Let us denote:

$$Z_i = \beta_0 + \boldsymbol{\beta}' \boldsymbol{X}_i, \quad (4)$$

which is modified for the purpose of this study to form:

$$Z_i = \beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i \tag{5}$$

and following formula:

$$P_i = \frac{1}{1 + e^{-Z_i}} = \frac{e^{Z_i}}{1 + e^{Z_i}} = F(Z_i), \tag{6}$$

is distribution function of the logistic distribution. Probability of death of x-years old person who will not die during next five years after the overall medical examination is:

$$1 - P_i = \frac{1}{1 + e^{Z_i}} \tag{7}$$

and therefore:

$$\frac{P_i}{1 - P_i} = e^{Z_i}. \tag{8}$$

By the logarithm we obtain LOGIT:

$$\text{In}\left(\frac{P_i}{1 - P_i}\right) = Z_i = \beta_0 + \boldsymbol{\beta' X_i}, \tag{9}$$

which is for the purpose of this study modified to form:

$$\text{In}\left(\frac{P_i}{1 - P_i}\right) = Z_i = \beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i. \tag{10}$$

To estimate the unknown parameters of the LOGIT model we can't use classical method of least squares, but with quality software we can use the maximum likelihood method. From the common formula of the log-likelihood function, specified e.g. by Christensen (1990):

$$\text{In } L(\beta_0, \boldsymbol{\beta}) = \sum_{i=1}^{N} \left[Y_i \text{ In}\left(\frac{e^{Z_i}}{1 + e^{Z_i}}\right) + (1 - Y_i) \text{In}\left(1 - \frac{e^{Z_i}}{1 + e^{Z_i}}\right)\right], \tag{11}$$

after substitution:

$$\text{In } L(\beta_0, \boldsymbol{\beta}) = \sum_{i=1}^{N} \left[Y_i \text{ In}\left(\frac{e^{\beta_0 + \boldsymbol{\beta' X_i}}}{1 + e^{\beta_0 + \boldsymbol{\beta' X_i}}}\right) + (1 - Y_i) \text{In}\left(1 - \frac{e^{\beta_0 + \boldsymbol{\beta' X_i}}}{1 + e^{\beta_0 + \boldsymbol{\beta' X_i}}}\right)\right], \tag{12}$$

and for the purpose of this study is:

$$\begin{aligned} \text{In } L(\beta_0, \beta_1, \beta_2, \beta_3) &= \sum_{i=1}^{N} \left[DEATH\_5_i \text{ In}\left(\frac{e^{\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i}}{1 + e^{\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i}}\right)\right. \\ &\left. + (1 - DEATH\_5_i) \text{ In}\left(1 - \frac{e^{\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i}}{1 + e^{\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i}}\right)\right]. \end{aligned} \tag{13}$$

1.2 PROBIT model

To estimate the unknown parameters of the PROBIT model we can not use classical methods of least squares either, but we can use a universal maximum likelihood method. From the common formula of the log- likelihood function, specified e.g. by Christensen (1990):

$$\text{In } L\ (\beta_0, \boldsymbol{\beta}) = \sum_{i=1}^{N} [Y_i \text{ In } (F(Z_i)) + (1 - Y_i) \text{ In } (1 - F(Z_i))], \tag{14}$$

after substitution:

$$\text{In } L\ (\beta_0, \boldsymbol{\beta}) = \sum_{i=1}^{N} [Y_i \text{ In } (F(\beta_0 + \boldsymbol{\beta' X_i})) + (1 - Y_i) \text{ In } (1 - F(\beta_0 + \boldsymbol{\beta' X_i}))], \tag{15}$$

for the purpose of this study:

$$\begin{aligned} &\text{In } L\ (\beta_0, \beta_1, \beta_2, \beta_3) \\ &\quad = \sum_{i=1}^{N} [DEATH\_5_i \text{ In } (F(\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i)) \\ &\quad + (1 - DEATH\_5_i) \text{ In } (1 - F(\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i))] \end{aligned} \tag{16}$$

and distribution function:

$$F(\beta_0 + \boldsymbol{\beta' X_i}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\beta_0 + \boldsymbol{\beta' X_i})} e^{\frac{-z^2}{2}} dz, \tag{17}$$

after substitution for the purpose of this study is:

$$F(\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\beta_0 + \beta_1 AGE_i + \beta_2 CIRD_i + \beta_3 CEL_i)} e^{\frac{-z^2}{2}} dz. \tag{18}$$

2 ESTIMATES OF UNKNOWN MODELS PARAMETERS

Estimating the unknown parameters of nonlinear regression models is no problem today. The software uses an iterative method. The software itself selects the initial value. To estimate the parameters of LOGIT and PROBIT model Statgraphics Centurion XVI version 16.1.11 and Gretl 1.8.7 build 2010-01-24 were used.

2.1 LOGIT model, year 1990 versus 1995

Based on the methodology showed above the estimates of unknown parameters of LOGIT models for males in 1990 and 1995 as well as for females in 1990 and 1995 were calculated. Model for males, who were invited to the general medical examination in 1990 is showed in Table 3, model for females, who were invited to the general medical examination in 1990 is showed in Table 4. Using data acquired in 1995, there were calculated estimates of parameters for both males (see Table 5) and females (see Table 6) once again.

Table 3 LOGIT model for males, 1990

Param	Estimate	St. Error	Est. Odds Rat.
C	–15.3537	3.61058	
AGE	0.327231	0.0842361	1.37899
CIRD	0.362252	0.0811236	1.42556
CEL	–8.18332	1.8898887	0.000266545

Source: Own construction

Table 4 LOGIT model for females, 1990

Param	Estimate	St. Error	Est. Odds Rat.
C	–15.7816	3.71599	
AGE	0.329121	0.0853405	1.38975
CIRD	0.361625	0.0857813	1.43566
CEL	–8.19201	1.89216	0.000276858

Source: Own construction

Table 5 LOGIT model for males, 1995

Param	Estimate	St. Error	Est. Odds Rat.
C	–9.17489	1.90414	
AGE	0.162236	0.0322365	1.18122
CIRD	0.189985	0.0488745	1.21113
CEL	–1.85112	0.5899631	0.159863

Source: Own construction

Table 6 LOGIT model for females, 1995

Param	Estimate	St. Error	Est. Odds Rat.
C	–9.38463	1.9487	
AGE	0.161341	0.0391378	1.17509
CIRD	0.195672	0.0471524	1.21613
CEL	–1.84446	0.522231	0.158111

Source: Own construction

We can see that all estimates of unknown parameters are statistically significant. The values from Tables 3–6 can be used to achieve the formulas (3), (10) and (13).

Now we express if the model is estimated well as a whole. As evaluation criteria we can choose for example McFadden's coefficient of determination, as set out e.g. by Freese and Long (2006), whose formula is:

$$McFadden's\ R^2 = 1 - \frac{LLF_{ur}}{LLF_r}, \tag{19}$$

where LLF_{ur} is the value of unlimited maximum likelihood function with all explanatory variables and LLF_r is the value of restricted maximum likelihood function without explanatory variables – a model with constant only. For the purpose of this study can be written as:

$$McFadden's\ R^2 = 1 - \frac{\ln L(b_0, b_1, b_2, b_3,)}{\ln L(b_0)}. \tag{20}$$

We can also use McFadden's corrected coefficient of determination, which also reflects a number of redundant variables in the model. Its formula can be written for the purpose of this study as:

$$McFadden's\ R^2_{adj.} = 1 - \frac{\ln L(b_0, b_1, b_2, b_3,) - K}{\ln L(b_0)}, \tag{21}$$

where K is the number of explanatory variables. McFadden's adjusted coefficient of determination is suitable for comparing of individual models. Table 7 shows the values of these evaluation criteria.

Table 7 Values of model evaluation criteria

LOGIT model for	McFadden's R^2 (in %)	McFadden's $R^2_{adj.}$ (in %)
males 1990	70.2291	65.2763
females 1990	70.3652	65.9664
males 1995	36.2791	31.2406
females 1995	38.3223	33.3214

Source: Own construction

Further evaluation of this model is available in Tables 8–11. In determining whether the model can be simplified, notice that the highest P-value for the likelihood ratio tests is 0.0002. Because the P-value is less than 0.05, it is irrelevant to remove any variable from the model.

Table 8 Likelihood rat. test, males 1990

Factor	Chi-Sq.	Df	P-Value
AGE	32.1326	1	0.0000
CIRD	37.6458	1	0.0000
CEL	77.6632	1	0.0000

Source: Own construction

Table 9 Likelihood rat. test, females 1990

Factor	Chi-Sq.	Df	P-Value
AGE	31.2349	1	0.0000
CIRD	38.8651	1	0.0000
CEL	78.8339	1	0.0000

Source: Own construction

Table 10 Likelihood rat. test, males 1995

Factor	Chi-Sq.	Df	P-Value
AGE	22.6632	1	0.0000
CIRD	23.1131	1	0.0000
CEL	15.5439	1	0.0001

Source: Own construction

Table 11 Likelihood rat. test, females 1995

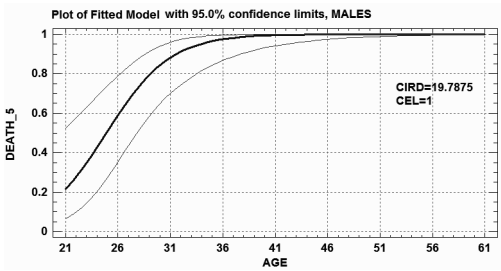
Factor	Chi-Sq.	Df	P-Value
AGE	21.5910	1	0.0000
CIRD	22.1363	1	0.0000
CEL	14.2464	1	0.0002

Source: Own construction

Based on the estimated parameters, there were constructed Figures 2–9, where the vertical axis is the probability of death of x-year old person during next five years after the general medical examination, the horizontal axis expresses a person's age. Value of CIRD in graphs is mean of the population.

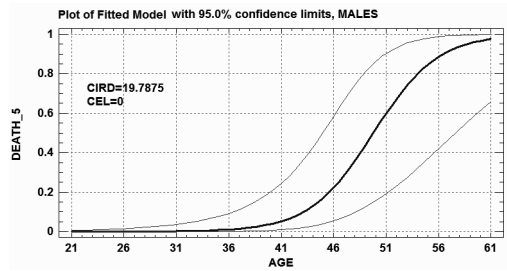
Figure 2 shows the probabilities of death of x-year old males during next 5 years after the general medical examination in 1990, if the males had diagnosed the disease of the digestive system (known as celiac disease now). Figure 3 shows the probabilities of death of x-year old males during next five years after the general medical examination in 1990, if the males had diagnosed any other disease (e.g. diabetes, high blood cholesterol, high blood pressure, bronchial asthma or other).

Figure 2 Probability of death, males, 1990, celiac disease – yes



Source: Own construction

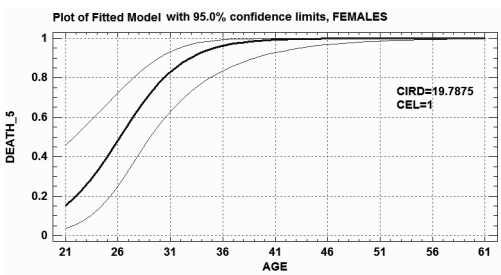
Figure 3 Probability of death, males, 1990, celiac disease – no



Source: Own construction

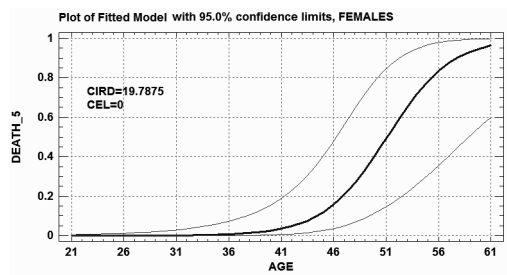
The probabilities for females with disease of the digestive system in 1990 are shown in Figure 4 and the probabilities for females with any other disease (e.g. diabetes, high blood cholesterol, high blood pressure, bronchial asthma or other) are shown in Figure 5.

Figure 4 Probability of death, females, 1990, celiac disease – yes



Source: Own construction

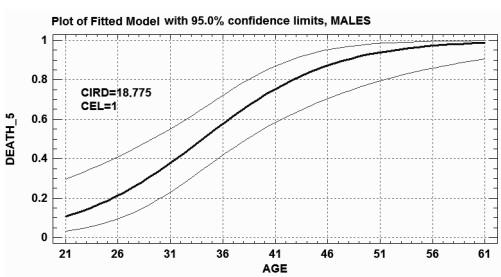
Figure 5 Probability of death, females, 1990, celiac disease – no



Source: Own construction

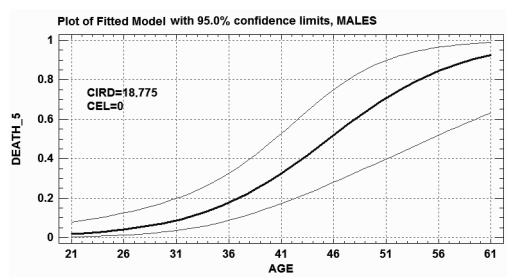
As time went on, the probabilities of death of x-years old person decreased. Decreased for both males and females, but for females a little more. Probability of death of x-year old males during next 5 years after the general medical examination in 1995 (if the males had diagnosed the celiac disease) shows Figure 6. Figure 7 shows the probability of death of x-year old males during next five years after the general medical examination in 1995, if the males had diagnosed any other disease.

Figure 6 Probability of death, males, 1995, celiac disease – yes



Source: Own construction

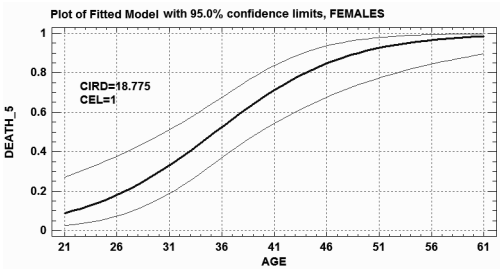
Figure 7 Probability of death, males, 1995, celiac disease – no



Source: Own construction

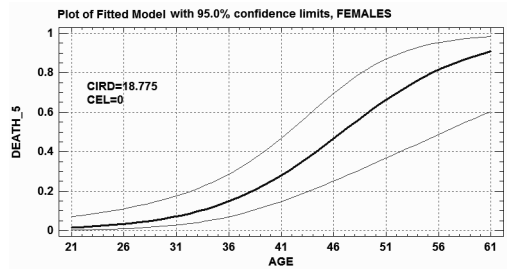
The same situation, but for females, is shown in Figure 8 in case of celiac disease and in Figure 9 in case of any other disease.

Figure 8 Probability of death, females, 1995, celiac disease – yes



Source: Own construction

Figure 9 Probability of death, females, 1995, celiac disease – no



Source: Own construction

2.2 PROBIT model, year 1990 versus 1995

Similarly as in the case of LOGIT model, there were calculated the estimates of the unknown parameters of PROBIT model for the same data. Model for males, who were invited to the general medical examination in 1990 is shown in Table 12, model for females in 1990 is shown in Table 13. Using data acquired in 1995, estimates of parameters for both males (see Table 14) and females (see Table 15) were calculated once again.

Table 12 PROBIT model for males, 1990

Param	Estimate	St. Error
C	–8.7076	1.90321
AGE	0.179963	0.0443266
CIRD	0.212333	0.0441222
CEL	–4.82122	1.0121123

Source: Own construction

Table 13 model for females, 1990

Param	Estimate	St. Error
C	–8.94966	1.95924
AGE	0.1862	0.0458694
CIRD	0.206262	0.0455631
CEL	–4.80235	1.01696

Source: Own construction

Table 14 PROBIT model for males, 1995

Param	Estimate	St. Error
C	–4.85958	0.983271
AGE	0.0900233	0.0236552
CIRD	0.1023366	0.0245888
CEL	–0.987444	0.2796356

Source: Own construction

Table 15 PROBIT model for females, 1995

Param	Estimate	St. Error
C	–4.96564	1.00653
AGE	0.0815844	0.0208831
CIRD	0.11225	0.0252715
CEL	–0.974493	0.282501

Source: Own construction

All estimates of unknown parameters are statistically significant. The values from Tables 12–15 could be used to achieve to the formulas (16) and (18). Table 16 shows the values of models evaluation criteria and from partial likelihood ratio tests is evident (see Tables 17–20), that the models are really estimated as far as possible well.

Table 16 Values of model evaluation criteria

PROBIT model for	McFadden's R^2 (in %)	McFadden's $R^2_{adj.}$ (in %)
males 1990	70.5539	65.6010
females 1990	70.5433	65.1132
males 1995	34.7826	29.7441
females 1995	34.0030	30.3366

Source: Own construction

Table 17 Likelihood rat. test, males 1990

Factor	Chi-Sq.	Df	P-Value
AGE	33.1369	1	0.0000
CIRD	38.6366	1	0.0000
CEL	77.9966	1	0.0000

Source: Own construction

Table 18 Likelihood rat. test, females 1990

Factor	Chi-Sq.	Df	P-Value
AGE	33.0536	1	0.0000
CIRD	39.4311	1	0.0000
CEL	79.5142	1	0.0000

Source: Own construction

Table 19 Likelihood rat. test, males 1995

Factor	Chi-Sq.	Df	P-Value
AGE	18.3636	1	0.0000
CIRD	22.3147	1	0.0000
CEL	13.3289	1	0.0003

Source: Own construction

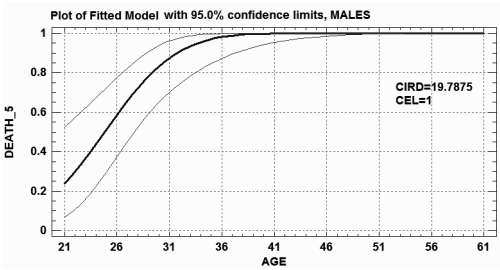
Table 20 Likelihood rat. test, females 1995

Factor	Chi-Sq.	Df	P-Value
AGE	18.6966	1	0.0000
CIRD	21.6042	1	0.0000
CEL	12.4251	1	0.0004

Source: Own construction

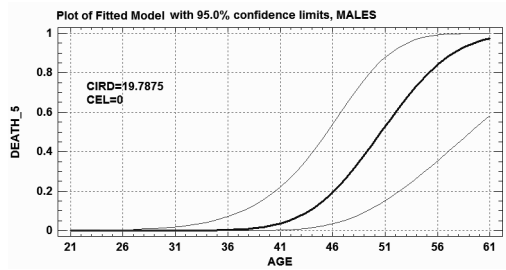
In the same system as in LOGIT models are organized outputs of probabilities of death of x-years old persons. These outputs result from Figures 10–17.

Figure 10 Probability of death, males, 1990, celiac disease – yes



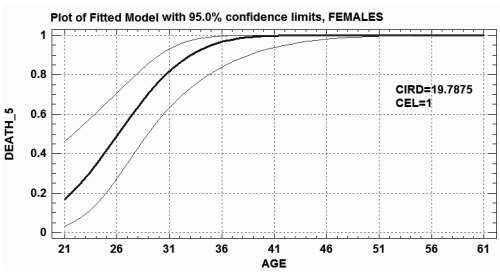
Source: Own construction

Figure 11 Probability of death, males, 1990, celiac disease – no



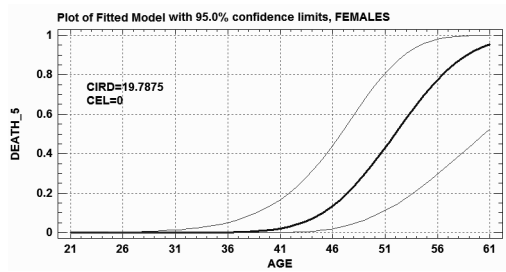
Source: Own construction

Figure 12 Probability of death, females, 1990, celiac disease – yes



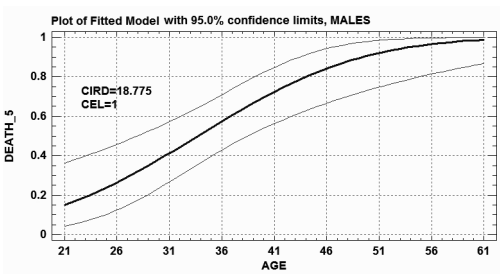
Source: Own construction

Figure 13 Probability of death, females, 1990, celiac disease – no



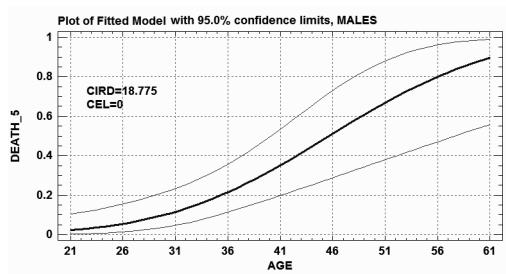
Source: Own construction

Figure 14 Probability of death, males, 1995, celiac disease – yes



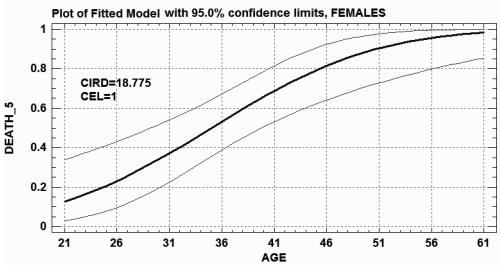
Source: Own construction

Figure 15 Probability of death, males, 1995, celiac disease – no



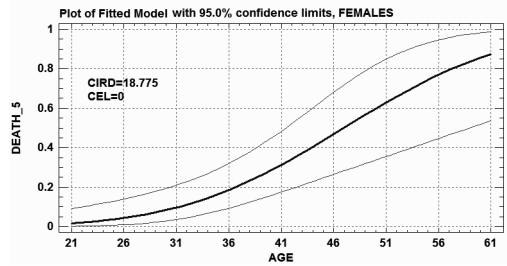
Source: Own construction

Figure 16 Probability of death, females, 1995, celiac disease – yes



Source: Own construction

Figure 17 Probability of death, females, 1995, celiac disease – no



Source: Own construction

CONCLUSION

The aim of this study was to analyse the probability of death of x-year old persons during next five years after the general medical examination in 1990 and 1995. Both analyses were solved using LOGIT and PROBIT models and tried to confirm the hypothesis claiming, that the probability of death of x-year old person suffering from celiac disease decreased few years after the gaining of another medical knowledge from abroad. At the beginning of the independent Czech Republic, the life expectancy at birth of people suffering from celiac disease was quite low, because the detailed diagnosis of this disease from abroad came later. With increasing age, the probability of death of these people grew faster than at the total population. This difference is already much smaller.

LOGIT and PROBIT models in this study do not differ much from each other, the results are almost comparable. Historical data from this last period is not possible to separate and the presented outputs of this study are based on specific research. Even if some assumptions for the application of methods of mathematical statistics are broken, it is possible to say, that the key hypothesis was confirmed.

References

Alergie na lepek (Allergy to Gluten) [online]. Společnost pro bezlepkovou dietu o. s. (Society for the Gluten-Free Diet). <http://celiak.cz/o-nemoci/alergie-na-lepek>.

CZSO. *Naděje dožití při narození v letech 1950–2010* (Life Expectancy at Birth between the Years 1950–2010) [online]. Czech Statistical Office. <http://www.czso.cz/csu/redakce.nsf/i/nadeje_doziti_pri_narozeni_v_letech_1950_2011>.

Diagnostika celiakie (Diagnosis of Celiac Disease) [online]. Poradenské centrum pro celiakii a bezlepkovou dietu (Advice Center for Celiac Disease and Gluten-Free Diet). <http://www.bezlepkovadieta.cz/celiakie-celiakalni-sprue/261-3/diagnostika-celiakie>.

FREESE, J., LONG, J.S. *Regression Models for Categorical Dependent Variables Using Stata*. College Station: Stata Press, 2006.

HOYOS, D., MARIEL, P., MEYERHOFF, J. *Comparing the Performance of Different Approaches to Deal with Attribute Non-Attendance in Discrete Choice Experiments: a Simulation Experiment. BILTOKI 201001,* Universidad del País Vasco – Departamento de Economía Aplicada III (Econometría y Estadística), 2010.

CHRISTENSEN, *R. Log-Linear Models.* New York: Springer-Verlag, 1990.

LOGAN, R. F., RIFKIND, E. A., TURNER, I. D., FERGUSON A. Mortality in Celiac Disease. *Gastroenterology,* 1989, 97(2), pp. 265–271.

PAVLÍK, Z., RYCHTAŘÍKOVÁ, J., ŠUBRTOVÁ, A. *Základy demografie* (Fundamentals of Demography). 1st Ed. Prague: Academia, 1986.

RUBIO-TAPIA, A. et al. Increased Prevalence and Mortality in Undiagnosed Celiac Disease [online]. *Gastroenterology,* 2009, 137, pp. 88–93. [cit. 16.8.2012]. <http://www.celiac.org/images/stories/PDF/Increase%20Prevalence_%20Tapia%20and%20Murray%202009.pdf>.

SPECTOR, L. C., MAZZEO, M. Probit Analysis and Economic Education. *Journal of Economic Education,* Spring, 11, 1980, pp. 37–44.

ŠIMPACH, O. Dopad získání lékařských poznatků na naději dožití u osob s celiakií a srovnání s nadějí dožití celkové populace (The Impact of the Acquisition of Medical Knowledges on the Life Expectancy of People with Celiac Disease and a Comparison with the Life Expectancy of the General Population). Prague, 05.12.2011–06.12.2011. In: *RELIK 2011 Reprodukce lidského kapitálu* (Reproduction of Human Capital). Slaný: Melandrium, 2011, pp. 1–9.

YANG, CH. W., RAEHSLER, R. D. An Economic Analysis on Intermediate Microeconomics: An Ordered PROBIT Model. *Journal for Economic Educators,* Vol. 5, No. 3, 2005.