CUMULATIVE INCIDENCE FUNCTION IN STUDIES
ON THE DURATION OF THE UNEMPLOYMENT EXIT PROCESS

Beata Bieszk-Stolorz, Ph.D., Associate Prof.

University of Szczecin
Faculty of Economics and Management
Institute of Econometrics and Statistics
Mickiewicza 64, 71-101 Szczecin, Poland
e-mail: beatus@wneiz.pl

Abstract

When we analyse the employment seeking process, an event that ends the observation of a given individual is their employment. The remaining observations are considered to be censored: the observations concluded before the end of the study or the cases of deregistering for other causes (e.g. old-age pension, taking up residence in a foreign country, starting further education). The act of taking up income-generating work can take various forms: taking up a job, setting up a business or taking advantage of subsidised job programmes. Jobseekers are often deregistered from poviat labour offices because they refuse to take up an offered job or fail to report to the office in due time. All the above events are forms of competing risk. The purpose of this paper is to use the cumulative incidence function to assess the probability of the unemployment exit with regard to different types of the competing risk. When competing-risk events occur, a solution sometimes is used where the remaining endpoint events are considered censored observations. Such a solution leads to an overestimation of probability. The results implicate that the beneficiaries' will to find employment was not a principal reason for a registering decision. The study is based on the individual data of jobseekers registered in the Poviat Labour Office in Szczecin.

Keywords: cumulative incidence function *CIF*, competing risk, unemployment

JEL classification: C41, J64

Introduction

Analysis of the duration of unemployment is one of the most important elements of unemployment research. Unemployment brings negative effects to both the national economy and the well-being of households. This is why governments can and should use adequate tools to alleviate these effects, focusing in particular on the duration of unemployment. Unfortunately, there are no ready-made solutions to the problem. Both unemployment rates and its mean duration vary from country to country, disregarding their development level. Governments offer diverse forms of support and activation programmes directed to unemployed citizens in general as well as to their specific groups. Finding a job is just one of many causes why an individual leaves the labour office register. The registered unemployed have the opportunity to retire, apply for an invalidity pension, receive early retirement benefits or enrol for full time studies. One of the most common causes of de-registration is the unjustified refusal to accept a job offer. All the above events are forms of competing risk. Interesting methods of estimating competing risk can be found among the methods used in the survival analysis.

The purpose of this paper is to use the cumulative incidence function (CIF) to assess the probability of unemployment exit with regard to different types of the competing risk. The study is based on the individual data of jobseekers registered in the Poviat Labour Office in Szczecin. The paper presents the problem of competing risks in studies from the analysis of duration. The methods and data used in the study were then described. In the empirical part of the paper, the forms of exiting the unemployment of persons registered in the Poviat Labour Office in Szczecin were examined. Finally, the most important conclusions of the study are presented.

1. Competing risk in the survival analysis – literature review

The survival analysis methods, commonly used in demography and medical sciences to study to measure human life expectancy, can also be applied in studies on the duration of social and economic phenomena. What is analysed here is the individual's survival time in a specific state (random variable T) until a specific endpoint event occurs. This can be, for example, company lifetime, a spell of unemployment or credit repayment time. We can use the survival analysis methods to examine the duration of firms (Markowicz, 2013), population's economic activity (Landmesser, 2009), poverty duration in households (Sączewska-Piotrowska, 2015b), credit risk (Wycinka, 2015) or the duration of unemployment (Bieszk-Stolorz, Markowicz, 2015).

A survival analysis can be complicated due to the issues related with censoring. Applied tools force the appropriate method of collecting statistical data. Individual data with consideration of time of being in a particular state is essential. An observation is censored when the survival time of a subject of interest is only partially known. In the case of the survival analysis the most common form is right censoring where the study ends or the subject withdraws before the endpoint event has occurred or when the subject disappears from the field of the survey for an unknown cause (Sączewska-Piotrowska, 2015a). In scientific research, right censoring is a condition when the endpoint event occurs which precludes the actual event from happening (Pepe, 1991), thus being a competing event. Kalbfleisch and Prentice (1980, p. 164) identified three different problems appearing in the analysis of data with competing risk:

1. The estimation of the relationship between covariates and the rate of occurrence of cause-specific failure.
2. The study of the interrelation between failure types under a specific set of conditions.
3. The estimation of failure rates for certain types of failure given the removal of some or all other failure types; this is regarded as a classic competing risk analysis.

Gooley, Leisenring, Crowley and Storer (1999) define the competing risk as an event whose occurrence either precludes the occurrence of another event under examination or fundamentally alters the probability of occurrence of this other event. We must assume that the occurrence of a specified event does not affect the probability of the occurrence of any other events (Crowder, 1994, 1996, 1997). The unit of interest is exposed to different kinds of risk at the same time interval, but it is assumed that the eventual failure of the unit is due to only one of these risks, which is called "a cause of failure" (Aly, Kochar, McKeague, 1994). In medical studies a typical example is the analysis of the cause of death. For example, in the case of death by cancer may occur due to the relapse or during the process of remission (caused by the administered therapy). Making a distinction between these two causes has a fundamental meaning for the whole process of treatment. In engineering, the competing risk is associated with the influence of individual components on the whole system. The failure of any of these components results in the failure of the system in general. Therefore, the observation covers the time of failure as well as the faulty component and the extent to which it affected the system. In the study of the labour market different routes of unemployment exit are different forms of competing risk. Assessment of the probability occurrence of a particular form of exit from unemployment is important from the point of view of an effective labour market policy.

2. Research methods

Let T and C be continuous random variables describing the time to event and the time to censoring, respectively. When there are K types of competing risks, the observation encompasses the pairs (X, δ), where $X = \min(T, C)$ and $\delta = 0, 1, \ldots, K$. If a given observation is censored, then $\delta = 0$ and $\delta = 1, \ldots, K$ for the observations ending with an event (one of the K competing-risk ones). In this context, one of the K events can be considered the event of elementary importance, while all the remaining ones – the competing-risk events.

The cumulative incidence function denoted as $CIF_k(t)$ is the probability of the occurrence of the event due to k before or at the time t. It is defined as (Klein, Moeschberger, 2003, p. 52):

$$CIF_k(t) = P(t \leq T, \delta = k) = \int_0^t S(u) h_k(u)\, du = \int_0^t S(u)\, dH_k \text{ for } k = 1, 2, 3, \ldots, K \tag{1}$$

where $H_k(t)$ is a specified (for a fixed k) function of the cumulative hazard function and $S(t)$ is a survival function. In the literature there is also the term cumulative incidence curves (*CIC*) (Kleinbaum, Klein, 2005).

Let $t_1 < t_2 < \ldots < t_i < \ldots < t_n$ be event times. Similarly to the standard cumulative hazard function in the survival analysis, the cumulative hazard function $H_k(t)$ for the cause k can be expressed by the Nelson-Aalen estimator:

$$\hat{H}_k(t) = \sum_{j:t_j \leq t} \frac{d_{kj}}{n_j} \tag{2}$$

where d_{kj} is a number of events that have occurred due to the cause k, while n_j is a number of individuals at risk at the time t_j. The expression of the general survival function $S(t)$ is defined with the Kaplan-Meier estimator is given by:

$$\hat{S}(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{n_j} \right) \tag{3}$$

Having combined the above two estimators (2) and (3), we can estimate the function of the cumulative incidence due to the cause k (Marubini, Valsecchi, 1995) as:

$$\hat{CIF}_k(t) = \sum_{j:t_j \leq t} \hat{S}(t_{j-1}) \frac{d_{kj}}{n_j} \tag{4}$$

Therefore, the cumulative incidence function of a given event can be defined as the cumulative probability of the occurrence of the k-type event before or at the time t (Bryant,

Dignam, 2004). This event allows us to determine the patterns of the incidence of the event due to the cause k and assess to what extent every cause contributes to total failure.

Since $\sum_{k=1}^{K} d_{kj} = d_j$, then the following relation is true:

$$\sum_{k=1}^{K} C\hat{I}F_k(t) = 1 - \hat{S}(t) \qquad (5)$$

In the absence of competing-risk events, we have:

$$C\hat{I}F(t) = 1 - \hat{S}(t) \qquad (6)$$

When competing-risk events occur, a solution sometimes is used where the remaining endpoint events are considered censored observations. Note, however, that such a solution leads to the overestimation of *CIF* (Sherif, 2004). The competing risk can also be assessed by means of the multinomial logit model which is treated in the survival analysis as a discrete time hazard model. The parameters of this model help determine the relative possibility of competing-risk events (Landmesser, 2008; Bieszk-Stolorz, Markowicz, 2012).

3. Data used in the study

The study is based on the anonymous individual data of 22,078 jobseekers registered by the Poviat Labour Office (PUP) in Szczecin in 2013 and observed by the end of 2014. The analysed span of time commenced at the moment of registration and ended with deregistration due to a specified cause. Polish Labour Offices give dozens of causes for unemployed people leaving the register and only a few are related with finding a job. The causes have been categorised (Table 1). Please note that Employment in the strict sense means taking up a job, while Employment (office) includes government subsidised forms of employment.

Table 1. Causes for deregistration from the Labour Office

No.	Group	Cause for Deregistering
1	2	3
1	Employment (in the strict sense)	Finding a job or other form of employment
2	Employment (office)	Being hired for intervention or public work, a job: under an individual social employment scheme, created under a government loan scheme for employers or for the 50+ created under a government subsidy scheme for employers
3	Economic Activity	Starting non-agricultural economic activity, being granted single resources for taking up an economic activity, starting a business activity subsidised from resources of the State Fund for the Rehabilitation of the Disabled (PFRON)

1	2	3
4	Refusal	Refusal without a duly justified reason to accept a job offer or other paid work, intervention or public work or refusal to take up a training course, internship or on-the-job apprenticeship; failure to report to PUP in due time and to inform the Office about the justified reason to do so within 5 days; failure to submit a medical certificate attesting the jobseeker's incapacity to work; unjustified refusal to participate in The Activation and Integration programme (PAI), lack of notification of being available to work over at least 10 days; drop-out from a training or internship programme or from other forms of subsidised employment, PAI scheme implemented or commissioned by PUP; a jobseeker's application for cancellation of his/her registration by PUP
5	Permanent Disability Allowance, Social Benefit, Old-age Pension	Becoming entitled to: permanent disability allowance, rehabilitation benefit, permanent social benefit, retirement or disability insurance on account of permanent employment as a household member in an agricultural holding; the jobseeker receives an attendance allowance, guardian's allowance or a single parent allowance; becoming entitled to an early-retirement allowance
6	Taking up Residence in a Foreign Country	Residence abroad longer than 30 days
7	Other	Taking up residence outside the area of the local PUP's authority; incapacity to work due to a medical condition or addiction treatment in a closed rehabilitation establishment for the uninterrupted period of over 90 days; taking up a training programme implemented by an entity other than PUP; death; military service; taking up full time education

Source: own study.

Table 2. Types of endpoint events and the corresponding censored observations

Endpoint event		Censored observation	
Cause for deregistration	Number	Cause for censoring	Number
Type I			
Any cause	20,222	Event had not happened before the end of 2014	1,856
Type II			
Employment (in the strict sense)	7,807	Event had not happened before the end of 2014 or deregistering was due to a reason other than taking up employment	14,271
Type III			
Employment (in the strict sense)	7,807	Event had not happened before the end of 2014	1,856
Employment (office)	929		
Economic Activity	897		
Refusal	8,965		
Permanent Disability Allowance, Social Benefit, Old-age Pension	685		
Taking up Residence in a Foreign Country	445		
Other	494		

Source: own study.

Beside the analysis of a specified phenomenon, i.e. the duration of unemployment, the paper points to the relations between the *CIF* estimators. Therefore, three types of endpoint events and the corresponding censored observations were selected (Table 2). The most common cause for deregistering was refusal followed by Employment (in the strict sense).

4. Analysis of duration of unemployment

The analysis consisted in determining the cumulative incidence function (*CIF*) for three types of endpoint events. Figure 1 shows the *CIF* estimators for endpoint events of type I and II. The type I endpoint event was the act of deregistering from the Poviat Labour Office (PUP) due to any cause. In this instance, the censored data were observations that had not been discontinued before the end of 2014. The type II endpoint event was employment, while censoring covered the observations that had not been discontinued before the end of 2014 or ended with an event different than employment. The *CIF* estimator shows the probability of deregistering from the Labour Office before or at the time *t*. The probability of deregistering from the Labour Office due to any cause was larger than due to employment, e.g. in a period of time shorter than 6 months this probability was 0.61 and 0.32, respectively. After 12 months it grew to 0.83 and 0.49.
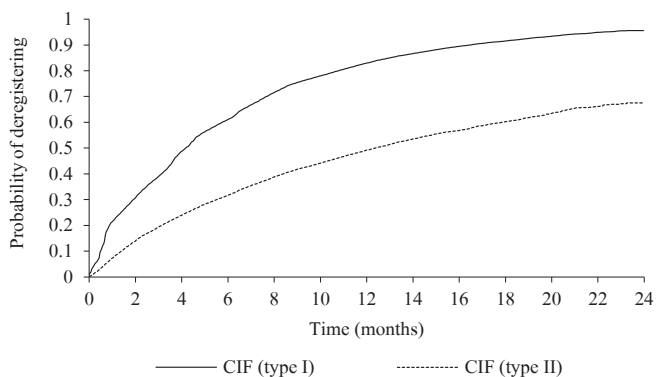

Figure 1. *CIF* estimators for endpoint events of type I and II

Source: own study.

Large difference between the curves means that the probability of deregistration may have been under the influence of causes other than employment. In order to examine this influence (or this competing risk) CIF_k ($k = 1, 2, \ldots, 7$) was determined for seven groups of causes for deregistering from the Labour Office (Table 1). Table 3 shows selected values of CIF_k and the

values of the *CIF* estimator for type I endpoint events with a view to demonstrate the equation (5). Significantly bigger values were attributed to the cases of deregistration due to Employment and Refusal. Other causes were of marginal importance and did not exceed 0.05.

Table 3. A part of the table containing *CIF* values for endpoint events of type I and III

Time (months)	Type I	Type III						
	Total	Employment	Employment (office)	Permanent Disability Allowance, Social Benefit, Old-age Pension	Economic Activity	Taking up Residence in a Foreign Country	Refusal	Other
0.00	0.010	0.003	0.000	0.003	0.001	0.000	0.002	0.001
0.03	0.015	0.003	0.004	0.003	0.001	0.000	0.003	0.001
0.07	0.018	0.004	0.005	0.003	0.001	0.000	0.004	0.001
0.10	0.028	0.006	0.012	0.003	0.001	0.000	0.004	0.001
0.13	0.034	0.007	0.016	0.003	0.001	0.000	0.005	0.001
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
12.26	0.835	0.325	0.040	0.028	0.035	0.019	0.367	0.020
12.30	0.836	0.326	0.040	0.028	0.035	0.019	0.367	0.021
12.33	0.837	0.326	0.040	0.028	0.035	0.019	0.367	0.021
12.36	0.838	0.326	0.040	0.028	0.035	0.019	0.368	0.021
12.39	0.839	0.327	0.040	0.028	0.035	0.019	0.369	0.021
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
23.77	0.957	0.367	0.043	0.033	0.044	0.021	0.426	0.023

Source: own study.

Figure 2 shows four CIF_k estimators for type III: Refusal and Employment reaching values considerably higher than the remaining causes, and two other causes related to employment: Employment (office) and Economic Activity. The analysis of the values of these functions shows that after 4 months from registration the probability of the jobseeker to be removed from the PUP register due to Refusal was bigger than due to the probability of Employment. Figure 3 shows the consequences of adopting the endpoint events of type II. The values of the thus determined *CIF* estimator (employment) are much bigger than the values of the CIF_k estimator (employment; endpoint events of type III). As it has been mentioned before, the adoption of the simplification (i.e. of the type II) results in the overestimation of value of the competing risk. In the present analysis this overestimation was very high – in the 24th month (the end of the observation) it was equal 84% (0.675 – II type, 0.367 – III type).
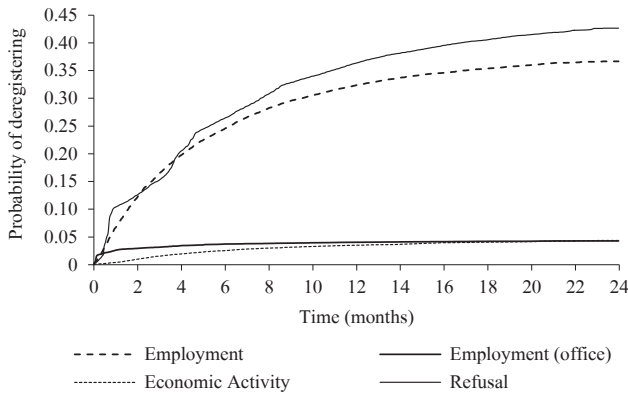
Figure 2. CIF_k estimators for endpoint events of type III

Source: own study.

Figure 3. CIF estimators for Employment and for endpoint events of type II and III

Source: own study.

Employment (office), Economic, Permanent Disability Allowance, Social Benefit, Old-age Pension as well as Taking up Residence in a Foreign Country are of marginal importance. Therefore, the three groups: Employment (in the strict sense), Employment (office) and Economic Activity can be integrated into one subgroup – Employment (in the general sense). Permanent Disability Allowance, Social Benefit, Old-age Pension as well as Taking up Residence in a Foreign Country can join the group Other. Consequently, the study results can be summarised by way of analysing the following subgroups: Employment (in the general sense), Refusal and Other (Table 4).

Table 4. Subgroups of causes for deregistering from labour office records

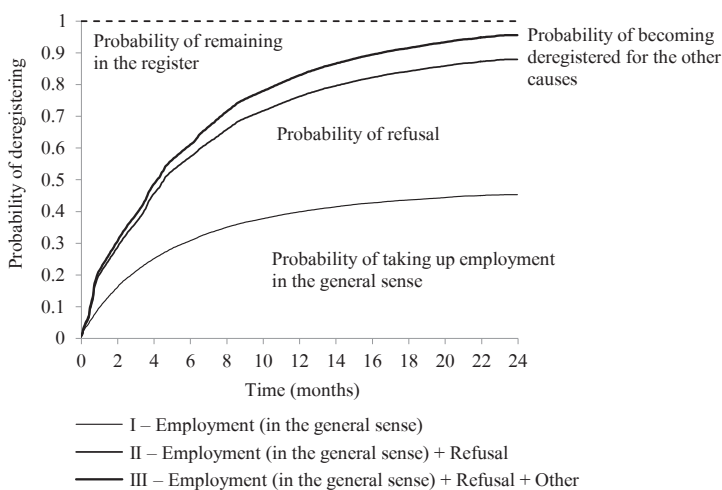| No	Subgroup	Group
1	Employment (in the general sense)	Employment (in the strict sense)
		Employment (office)
		Economic Activity
2	Refusal	Refusal
3	Other	Permanent Disability Allowance, Social Benefit, Old-age Pension
		Taking up Residence in a Foreign Country
		Other

Source: own study.

Figure 4. Cumulated *CIF* estimators for subgroups

Source: own study.

Figure 4 shows the summed *CIF* estimators for three subgroups. These curves provide a means of seeing dynamically the relationship between the competing risk probabilities. Distances between curves express the probability of deregistration due to a specified cause. The distance between the curves is measured by means of the metric:

$$d((x_1, y_1), (x_2, y_2)) = |y_2 - y_1|.$$

The distance between curve III and a line with a value of 1 is equal to the probability of remaining in the register. As seen from Figure 4, this probability was decreasing over time. After 12 months after registration it was at 0.17. At the same time, it denotes the probability of

entering prolonged unemployment. The probability of remaining registered by the labour office after 24 months (the endpoint) was at 0.04. The deregistration due to starting economic activity and taking up subsidised employment were of marginal importance, but they still influenced the increased possibility of deregistration due to employment in the general sense. After 12 months from the moment of registration was at 0.40, while after 24 months – 0.45. Likewise, the probability of refusal was equal 0.36 and 0.43, respectively.

Conclusions

A preliminary analysis of individual data obtained from the Labour Office in Szczecin allows for a conclusion that refusal was the most common cause of jobseekers' deregistration from the PUP records. The applied cumulative incidence function confirmed that this was the most probable cause of deregistration. Employment (in the strict sense) was in second place. The results implicate that the beneficiaries' will to find employment was not a principal reason for a registering decision. One of the main tasks of labour offices is helping unemployed citizens re-enter the job market. Subsidised employment is merely a support measure. It does not guarantee permanent employment, however, it is an important element of the economic activation of the unemployed, especially of those experiencing long-term unemployment. Other measures to prevent unemployment are programmes aimed at encouraging self-employment or financial support for start-ups offered by labour offices. Combining different types of employment into one group has changed the value of the estimate. Employment (in the general sense) was the most common cause of deregistration. The use of duration methods requires the use of individual data. This was a serious limitation in the study. The problem is access to the required data. In Poland only labour offices have such data.

An important subject-matter observation results from the study: when there are diverse endpoint events, then it is worthwhile to use models that allow us to estimate competing risk. From the point of view of labour market policies, it is important to analyse not only the job-finding events, but also the determinants of other routes of unemployment exit. In the case of the duration of registered unemployment, they help determine the probability of taking up a specified type of employment and compare them with other causes for deregistration. Recognising other endpoint events as censored can result in the considerable overestimation of value of the cumulative incidence function. Access to individual data is a serious limitation of the presented methods. These data in Poland are held only labour offices. Offices do not always want to share them.

References

Aly, E.A.A., Kochar, S., McKeague, E. (1994). Some Tests for Comparing Cumulative Incidence Functions and Cause-Specific Hazard Rates. *Journal of the American Statistical Association*, *89* (427), 994–999. DOI: 10.1080/01621459.1994.10476833.

Bieszk-Stolorz, B., Markowicz, I. (2012). Wykorzystanie wielomianowego modelu logitowego do oceny szansy podjęcia pracy przez bezrobotnych. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, *242*, *Taksonomia*, *19*, 628–636.

Bieszk-Stolorz, B., Markowicz, I. (2015). Influence of Unemployment Benefit on Duration of Registered Unemployment Spells. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *10* (3), 167–183. DOI: 10.12775/EQUIL.2015.031.

Bryant, J., Dignam, J.J. (2004). Semiparametric models for cumulative incidence functions. *Biometrics*, *60* (1), 182–190. DOI: 10.1111/j.0006-341X.2004.00149.x.

Crowder, M. (1994). Identifiability Crises in Competing Risks. *International Statistical Review*, *62* (3), 379–391. DOI: 10.2307/1403768.

Crowder, M. (1996). On assessing independence of competing risks when failure times are discrete. *Lifetime Data Analysis*, *2* (2), 195–209. DOI: 10.1007/BF00128575.

Crowder, M. (1997). A test for independence of competing risks with discrete failure times. *Lifetime Data Analysis*, *3* (3), 215–223. DOI: 10.1023/A:1009696830515.

Gooley, T.A., Leisenring, W., Crowley, J., Storer, B.E. (1999). Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine*, *18* (6), 695–706. DOI: 10.1002/(SICI)1097-0258(19990330)18:6<695::AID-SIM60>3.0.CO;2-O.

Kalbfleisch, J.D, Prentice, R.L (2002). *The Statistical Analysis of Failure Time Data, Second Edition*. Hoboken, NJ: John Wiley & Sons, Inc.

Klein, J.P., Moeschberger, M.L (2003). *Survival Analysis: Techniques for Censored and Truncated Data. Second Edition*. New York, NY: Springer-Verlag.

Kleinbaum, D.G., Klein, M. (2005). *Survival Analysis. A Self-Learning Text. Second Edition*. New York, NY: Springer.

Landmesser, J.M. (2008). Aktywność ekonomiczna ludności: klasyfikacja osób za pomocą wielomianowych modeli logitowych oraz jej związek z modelami hazardu dla czasów trwania. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, *7*, *Taksonomia*, *15*, 426–434.

Landmesser, J. (2009). The survey of economic activity of people in rural areas – the analysis using the econometric hazard models. *Acta Universitatis Lodziensis, Folia Oeconomica*, *228*, 385–392.

Markowicz, I. (2013). Hazard Function as a Tool to Diagnose Business Liquidation. *Folia Oeconomica Stetinensia*, *13* (2), 23–36. DOI: 10.2478/foli-2013-0019.

Marubini, E., Valsecchi, M. (1995). *Analysing Survival Data from Clinical Trials and Observational Studies*. Chichester, England: JohnWiley & Sons.

Pepe, M.S. (1991). Inference for Events With Dependent Risks in Multiple Endpoint Studies. *Journal of the American Statistical Association*, *86* (415), 770–778. DOI: 10.1080/01621459.1991.10475108.

Sączewska-Piotrowska, A. (2015a). Badanie ubóstwa z zastosowaniem nieparametrycznej funkcji przeżycia dla zdarzeń powtarzających się. *Przegląd Statystyczny*, *62* (1*)*, 29–51.

Sączewska-Piotrowska, A. (2015b). Poverty duration of hauseholds of the self-employed. *Econometrics*, *1* (47), 44–55. DOI: 10.15611/ekt.2015.1.03.

Sherif, Bintu, N. (2008). *A comparison of Kaplan-Meier and cumulative incidence estimate in the presence or absence of competing risks in breast cancer data.* Master's Thesis, University of Pittsburgh.

Wycinka, E. (2015). Time to Default Analysis in Personal Credit Scoring. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławi*u, *381*, 527–536. DOI: 10.15611/pn.2015.381.38.