FORUM FOR PUBLIC HEALTH IN SOUTH EASTERN EUROPE

Programmes for Training and Research in Public Health

METHODS AND TOOLS IN PUBLIC HEALTH

A Handbook for Teachers, Researchers and Health Professionals

Editors:

Lijana Zaletel-Kragelj and Jadranka Božikov

MetaNET and FPH-SEE Project Coordinators:

Doris Bardehle, Luka Kovačić, Ulrich Laaser and Oliver Razum

Lage, 2010

Bibliographic information published by *Die Deutsche Bibliothek*

Die Deutsche Bibliothek lists this publication in the *Deutsche Nationalbibliografie*; detailed bibliographic data is available on the Internet at http://dnb.ddb.de

This publication has been supported by the German Academic Exchange Service (*Deutsche Academische Austauschdienst – DAAD*) with funds from the Stability Pact: MetaNET Project, including Academic Programmes for Training and Research in Public Health in South Eastern Europe (FPH-SEE).

Publisher: Hans Jacobs Publishing Company
Cover design: Alma Šimunec-Jović
Printed by: C.B. print, Samobor, Croatia
Number of copies: 200

ISBN 978-3-89918-176-0

CONTENTS

PREFACE

This is the sixth out of seven books planned to be published in a series as a support to teachers and trainers in teaching public health in South Eastern Europe. Originally planned to be on the internet platform only, the Forum for Public Health in South Eastern Europe (FPH-SEE) and the MetaNET project as its continuation together with the Hans Jacobs Publishing Company decided later to publish this training material also as hard copy books. The first four books were published with the support of FPH-SEE, and the last two with the support of MetaNET. Both projects are supported by the German Academic Exchange Service (DAAD - Deutsche Academic Austauschdienst) with funds from the Stability Pact for South Eastern Europe, provided by the German Ministry of Foreign Affairs.

We are proud that this book will be published on the 10th year of the Public Health Network in South Eastern Europe.

The book **Methods and Tools in Public Health** is a collection of 47 teaching modules in 5 chapters written by 53 authors from 11 countries. The teaching modules in this book cover areas of methods of studying population health, special epidemiological methods and methods of public health interventions, methods of planning and evaluation and modules as the supportive tools and technologies. Authors had autonomy in preparation the teaching modules, they were asked to present their own teaching/training materials with the idea to be as practical and lively as possible. The role of editors was to stimulate the authors in writing modules and to collaborate with them in editing the final version of the manuscripts in order to get them as much as possible to the planned format. By preparing and publishing this teaching/training modules authors and editors expect and wish to support and improve public health education and training of public health professionals.

The editors asked and encouraged authors to incorporate in their teaching modules exercises, tests, questionnaires and other practical forms of training. We will be thankful for any comments on use of them in everyday practice.

The next and the last book will be entitled "International Public Health".

You can find all volumes on the website of the Forum of Public Health: *http://www.snz.hr/ph-see/publications.htm*, and the volumes 4-6 on the open access Literature database of the University Bielefeld: *http://biecoll.ub.uni-bielefeld.de*.

Editors and Project coordinators:

Lijana Zaletel-Kragelj, Jadranka Božikov,
Doris Bardehle, Luka Kovačić, Ulrich Laaser and Oliver Razum

October 2010

Forum for Public Health in South Eastern Europe and Hans Jacobs Publishing Company in this series published the following books:

1. Vesna Bjegović and Doncho Donev (eds). HEALTH SYSTEM AND THEIR EVIDENCE BASED DEVELOPMENT. Lage: Hans Jacobs; 2004.
2. Silvia Gabriela Scîntee and Adriana Galan (eds). PUBLIC HEALTH STRATEGIES: A TOOL FOR REGIONAL DEVELOPMENT. Lage: Hans Jacobs; 2005.
3. Lidia Georgieva and Genc Burazeri (eds). HEALTH DETERMINANTS IN THE SCOPE OF NEW PUBLIC HEALTH. Lage: Hans Jacobs; 2005.
4. Doncho Donev, Gordana Pavleković and Lijana Zaletel-Kragelj (eds). HEALTH PROMOTION AND DISEASE PREVENTION. Lage: Hans Jacobs; 2007.
5. Luka Kovačić and Lijana Zaletel-Kragelj (eds). MANAGEMENT IN HEALTH CARE PRACTICE. Lage: Hans Jacobs; 2008.
6. Lijana Zaletel-Kragelj and Jadranka Božikov (eds). METHODS AND TOOLS IN PUBLIC HEALTH. Lage: Hans Jacobs; 2010.

All books can be found at: *http://www.snz.hr/ph-see/publications.htm*, and all modules included in volumes 4-6 on the open access Literature database of the University Bielefeld: *http://biecoll.ub.uni-bielefeld.de*.

In preparation:

7. INTERNATIONAL PUBLIC HEALTH to be published in 2011.

LIST OF AUTHORS

1. **Aleksandrova Silviya, MD, MB, Assistant professor**, Faculty of Public Health, Medical University of Pleven, Bulgaria
2. **Artnik Barbara, DMD, PhD, Teaching Assistant**, Chair of Public Health, Faculty of Medicine, University of Ljubljana, Slovenia
3. **Bardehle Doris, MD, PhD, Professor**, Faculty of Health Sciences, University of Bielefeld, Germany
4. **Bilban Marjan, MD, PhD, Professor**, Chair of Public Health, Faculty of Medicine, University of Ljubljana, Slovenia
5. **Bjegović Vesna, MD, PhD, Professor**, Institute of Social Medicine and Centre School of Public Health, School of Medicine, University of Belgrade, Serbia
6. **Božikov Jadranka, PhD, Professor**, Andrija Štampar School of Public Health, School of Medicine, University of Zagreb, Croatia
7. **Bulat Petar, MD, PhD, Professor**, University of Belgrade, School of Medicine, Occupational Health Department, Belgrade, Serbia
8. **Burazeri Genc, MD, PhD**, Department of International Health, Faculty of Health, Medicine and Life Sciences, Maastricht University, the Netherlands
9. **Cucu Alexandra, MD, PhD, Lecturer**, Department of Public Health and Management, University of Medicine and Pharmacy "Carol Davila", Bucharest, Romania
10. **Donev Doncho, MD, PhD, Professor**, Institute of Social Medicine, Medical Faculty, University of Skopje, Republic of Macedonia
11. **Dunn William N, PhD, Professor,** Graduate School of Public and International Affairs, University of Pittsburgh, PA, USA
12. **Dyakova Mariana, MD, PhD, Associate Professor**, Faculty of Public Health, Medical University Sofia, Bulgaria
13. **Erlih Saša, BSc, MSc**, Institute for Comprehensive Development Solutions, Slovenia
14. **Eržen Ivan, MD, PhD, Assistant Professor**, Chair of Public Health, Faculty of Medicine, University of Ljubljana, Slovenia
15. **Fink Rok, BSc, MSc candidate, Teaching Assistant,** Department of Sanitary Engineering, Faculty of Health Studies, University of Ljubljana, Slovenia
16. **Galan Adriana,** Public Health and Health Management consultant, Institute of Public Health, Bucharest, Romania
17. **Gjorgjev Dragan, MD, PhD, Professor**, Republic Institute for Health Protection and Faculty of Medicine, University Ss Cyril and Methodius, Skopje, Republic of Macedonia
18. **Grancharova Gena, MD, PhD, Associate Professor**, Faculty of Public Health, Medical University of Pleven, Bulgaria
19. **Gudeva Nikovska Dance, MD, MPH, Teaching Assistant**, Ministry of Health, Republic of Macedonia

20. **Imamović Enida, MD, DDM**, Public Health Institute of Federation B&H, Bosnia and Herzegovina
21. **Ivartnik Matej, MSc**, Regional Institute of Public Health Ravne na Koroškem, Slovenia
22. **Janković Slavenka, MD, PhD, Professor**, School of Medicine, University of Belgrade, Serbia
23. **Jevšnik Mojca, PhD, Senior Lecturer**, Department of Sanitary Engineering, Faculty of Health Studies, University of Ljubljana, Slovenia
24. **Jocić Neda, MD, PhD, Professor**, Faculty of medicine, Novi Sad, Serbia
25. **Karaslavova Emilia, MD, PhD, Assistant Professor**, Department of social medicine and health management, Medical University Plovdiv, Bulgaria
26. **Kasapinov Blaško, MD, Teaching Assistant**, Institute of Social Medicine, Faculty of Medicine Ss Cyril and Methodius University, Skopje, Republic of Macedonia
27. **Kendrovski Vladimir, MD, PhD, Assistant Professor**, Republic Institute for Health Protection and Faculty of Medicine, University Ss Cyril and Methodius, Skopje, Republic of Macedonia
28. **Kern Josipa, PhD, Professor**, Andrija Štampar School of Public Health, School of Medicine, University of Zagreb, Croatia
29. **Klavs Irena, MD, PhD, Assistant Professor**, Institute of Public Health of the Republic of Slovenia, Ljubljana, Slovenia
30. **Kragelj Anja, Student** of Andragogy at Department of Pedagogy and Andragogy, Faculty of Arts, University of Ljubljana, Slovenia
31. **Kukec Andreja, BSc, PhD candidate**, Teaching Assistant, Chair of Public Health, Faculty of Medicine, University of Ljubljana, Slovenia
32. **Lazarevik Vladimir, MD, MPH, Teaching Assistant**, Institute of Social Medicine, Faculty of Medicine Ss Cyril and Methodius University, Skopje, Republic of Macedonia
33. **Mašić Izet, MD, PhD, Professor**, Chair for Medical Informatics, Medical faculty, University of Sarajevo, Bosnia and Herzegovina
34. **Mateev Hristo, MD**, University Hospital "Lozenetz", Sofia, Bulgaria
35. **Milevska-Kostova Neda, MSc, MCPPM**, Centre for Regional Policy Research and Cooperation "Studiorum", Skopje, Republic of Macedonia
36. **Nikšić Dragana, MD, PhD**, Medical Faculty, University of Sarajevo, Bosnia and Herzegovina
37. **Nitescu Maria, MD, PhD, Lecturer**, Department of Hygiene and Ecology, University of Medicine and Pharmacy "Carol Davila", Bucharest, Romania
38. **Pavleković Gordana, MD, PhD, Professor**, Andrija Štampar School of Public Health, Medical School, University of Zagreb, Croatia
39. **Pekmezović Tatjana, MD, PhD, Professor**, Institute of Epidemiology, School of Medicine, University of Belgrade, Serbia

40. **Plesničar Andrej, MD, PhD, Teaching Assistant**, Faculty of Health Studies, University of Ljubljana, Slovenia
41. **Roshi Enver, MD, MPH**, Faculty of Medicine, University of Tirana, Albania
42. **Rotar Pavlič Danica, MD, PhD, Assistant Professor**, Department of Family Medicine, Faculty of Medicine, University of Ljubljana, Slovenia
43. **Savova Zaharina, PhD, Associate Professor**, Faculty of Public Health, Medical University, Sofia, Bulgaria
44. **Sidjimova Dobriana, PhD, Assistant Professor**, Faculty of Public Health, Medical University, Sofia, Bulgaria
45. **Sočan Maja, MD, PhD, Assistant Professor**, National Institute of Public Health, Ljubljana, Slovenia, Ljubljana, Slovenia
46. **Sović Slavica, MD, PhD candidate**, Andrija Štampar School of Public Health, School of Medicine, University of Zagreb, Croatia
47. **Stikova Elisaveta, MD, PhD, Professor**, National Public Health Institute and Faculty of Medicine, University "Ss. Cyril and Methodius", Skopje, Republic of Macedonia
48. **Škerget Nataša, Student** of Pedagogy at Department of Pedagogy and Andragogy, Faculty of Arts, University of Ljubljana, Slovenia
49. **Tozija Fimka, MD, PhD, Associate Professor**, Republic Institute for Health Protection and Faculty of Medicine, University Ss Cyril and Methodius-Skopje, Republic of Macedonia
50. **Vitcu Anca, PhD, Assistant Professor**, Faculty of Computer Science, University "Al. I. Cuza", Iasi, Romania
51. **Vračko Pia, MD, PhD candidate, Research Assistant, Teaching Assistant**, National Institute of Public Health, Ljubljana, Slovenia, Ljubljana, Slovenia
52. **Vuletić Mavrinac Gorka, PhD**, Andrija Štampar School of Public Health, School of Medicine, University of Zagreb, Croatia
53. **Zaletel-Kragelj Lijana, MD, PhD, Associate Professor**, Chair of Public Health, Faculty of Medicine, University of Ljubljana, Slovenia

METHODS AND TOOLS IN PUBLIC HEALTH A Handbook for Teachers, Researchers and Health Professionals	
Title	**FREQUENCY MEASURES: ESTIMATING RISK**
Module: 1.2.6	**ECTS (suggested): 0.30**
Author(s), degrees, institution(s)	**Lijana Zaletel-Kragelj**, MD, PhD, Associate Professor Chair of Public Health, Faculty of Medicine, University of Ljubljana, Slovenia **Jadranka Božikov,** PhD, Professor Andrija Stampar School of Public Health, Medical School, University of Zagreb, Croatia
Address for correspondence	**Lijana Zaletel-Kragelj** Chair of Public Health, Faculty of Medicine, University of Ljubljana, Slovenia Zaloška 4, Ljubljana, Slovenia E-mail: lijana.kragelj@mf.uni-lj.si
Keywords	Risk, cumulative incidence, simple cumulative method, actuarial method, density method
Learning objectives	After completing this module students should be: • familiar with differences between four different methods for estimation of cumulative risk, being simple cumulative, actuarial, density, and Kaplan Meier method; • able to estimate cumulative risk measures of different level of accuracy independently.
Abstract	Risk is defined as the probability that a disease-free individual is developing a disease under observation over a specified period, conditional on that the same individual is not dying from any other disease during the period. In practice, risk is estimated by using different methods. The simple cumulative method is the easiest and most widely used. Risk cannot be accurately estimated by this method unless all subjects in the observed candidate population are followed for the entire follow-up period or are known to develop the disease during the period (no censoring). Because of serious limitations of this method, several methods more or less susceptible to censoring were proposed. Considering the censoring of the data in estimating cumulative risk requires the use of special analytic methods. These methods are actuarial, density, and Kaplan Meier method.
Teaching methods	An introductory lecture gives the students first insight in four methods for calculation of cumulative risk. The theoretical knowledge is illustrated by case studies. After introductory lectures students first carefully read the theoretical background of this module and complement their knowledge with recommended readings. Afterwards they on provided data set perform tasks on estimation of different types of measures. They are stimulated to compare results with other students and discuss the differences.
Specific recommendations for teachers	• work under teacher supervision/individual work proportion: 30%/70%; • facilities: a lecture room, a computer room; • equipment: computers (1 computer on 2-3 students), LCD projection, access to the Internet; • training materials: recommended readings or other related readings; • target audience: master degree students according to Bologna scheme.
Assessment of students	Written report on calculated measures in which detailed description of process of calculation is described.

FREQUENCY MEASURES: ESTIMATING RISK

Lijana Zaletel-Kragelj, Jadranka Božikov

THEORETICAL BACKGROUND

Introduction

In expressing relative incidence we are dealing with several measures. One of them is so called risk.

Risk is defined as the probability that a disease-free individual is developing a disease under observation over a specified period, conditional on that the same individual is not dying from any other disease during the period (1). Thus, risk is a conditional probability, with values varying between zero and one. It is dimensionless (1). It usually refers to the first occurrence of the disease for each initially disease-free individual, although it is possible to consider the risk of developing the disease under observation within a specified period more than once (1).

In practice, risk is estimated by using different methods. The simple cumulative method is the easiest and most widely used (1). For a cohort of subjects followed for a given period of time, risk is often estimated by calculating the proportion of candidate subjects who develop the disease during the observation period. This measure is usually referred as the cumulative incidence (CI) (1). Generally cumulative incidence is estimated only for first occurrence of the disease. If the durations of the individual follow-up periods for all non-cases are equal, the cumulative incidence is equivalent to the average risk for members of the cohort. This means thait under the condition of a fixed cohort cumulative incidence is good estimate of risk. This is the reason that cumulative incidence and risk are frequently equalized. But once again, because risk is, by its definition, a conditional probability, it cannot be accurately estimated by calculating cumulative incidence unless all subjects in the observed candidate population are followed for the entire follow-up period or are known to develop the disease (or other observed phenomenon) during the period (1).

The cumulative probability of the event during a given time interval is the proportion of new events during the interval in which the denominator is the initinal number of observed persons. The calculation of this measure is straightforward if no losses happen in the cohort during the interval (1-9). However, in real life the size of the cohort is more than likely to be decreased after a long period of follow-up as a result of different reasons. A situation in which the event and the time of individual is at risk for the event is unknown is usually called censoring (2,8-12).

There are usually three reasons why censoring occurs. The first is the termination of the observation because of the end of the study before the event occurs, the second is the termination because of some competing factors (death of other cause e.g. traffic accident), the third, the fourth simply the lost because of changing the domicile of the individual under observation, etc. In all cases the occurrence of observed phenomenon is unknown. The terms also used with this phenomenon are “withdrawals”, “losts-to-follow-up” and others (2,8-12). Considering the censoring of the data requires the use of special analytic methods.

The methods of risk estimation are the simple cumulative method, the actuarial method, the density method, and the Kaplan Meier product limit method (1,2,9-13).

Methods of risk estimation

Simple cumulative method

This method is the easiest for estimating risk (1,2,12). The risk calculated by this method is the most rough measure in this family of measures.

It is simply the proportion of new events during the interval in which the denominator is the initinal number of observed persons (Equation 1):

$$_{cum}R = \frac{N_{d+new\,cases(gp)}}{N_{all\ persons\,at\,risk\,(bgp)}} \qquad \textbf{Equation 1.}$$

$_{cum}R$ = *cumulative risk (risk of getting a disease during the entire period)*
$N_{d+\ new\ cases\ (gp)}$ = *number of new cases of the disease under observation during a given period*
$N_{all\ persons\ at\ risk\ (bgp)}$ = *number of all persons at risk for getting ill with the disease under observation at the beginning of a given period*

Usually it is estimated only for the first occurrence of the disease. This is the reason that the population at risk (the denominator in the equation) consists of disease-free individuals at the beginning of the observational period. The observation period has to be clearly stated since the value of the measure is increasing with the prolongation of period of observation. This period could be based upon a callendar time or not (e.g. first year after the exposure, first year after surgery etc.). It is good estimate of the risk only in the case of fixed cohorts in which there are no withdrawals from the follow-up (1,12).

Estimation of cumulative risk over entire 5-year observational period in practice is presented in in Case study 1.

For avoiding the drawbacks of this rough direct method of estimation of cumulative risk over longer period, we could split this longer period first to shorter periods (i.e. 1-year periods) and obtain cumulative risk indirectly through calculating risks for these periods (partial risks). When partial risk refers to 1-year period it is known as annual risk (Equation 2):

$$_{ann}R = \frac{N_{d+new\,cases(1\text{-}year\ period)}}{N_{all\ persons\,at\,risk\,(beginning\,of\ 1\text{-}year\ period)}} \qquad \textbf{Equation 2.}$$

$_{ann}R$ = *annual risk (risk of getting a disease during the 1-year period)*
$N_{d+\ new\ cases\ (1\text{-}year\ period)}$ = *number of new cases of the disease under observation during1-year period*
$N_{all\ persons\ at\ risk\ (beginning\ of\ 1\text{-}year\ period)}$ = *number of all persons at risk for getting ill with the disease under observation at the beginning of a given 1-year period*

The annual risk is annual probability of the event (12). The complement of this probability (the mirror image) is annual probability of survival without an event under observation (i.e. a breakout of a disease). Technically these probabilities are conditional probabilities. This means for example, that one has to survive through the first interval in order to be a part of the denominator for the calculation of the survival probability in the second interval. Similarly, the survival probability for the third interval is calculated only among those persons who survived first the first and then the second interval (12).

A cumulative probability of survival without a disease under observation over more than one interval (2-, 3-, 4-, 5-year interval, etc.) is obtained by multiplying the annual conditional survival probabilities over all intervals (12). Afterwards we calculate again complementary values (1 – cumulative survival) that are in fact cumulative risks over more than one interval.

By using this procedure the censoring is partially considered even when using simple method, as we need to define separately for every year the number of individuals under observation at risk, and all participants who terminated the observation because of extraneous factors (e.g. death because of traffic accident etc.) are not included.

Estimation of cumulative 5-year risk over observational period through calculation of annual risks is presented in in Case study 1.

Actuarial method

This is the first method in which the censoring is considered in calculation of risk estimate (1,8,11-13). It is tipically used to estimate the probability of death in survival analysis, but as mortality is a special case of incidence (12), it could be generalized to estimation of risk on general (2). It is refered also as interval-based life table or life table interval approach (12).

This method is working under the assumption that the censoring is occurring uniformly throughout the observed period (usually meaning that all withdrawals, i.e. censored observations, occur on average in the middle of the observational period) (1,2,11). If the periods are short (up to 1 year), or there is a small number of withdrawals this assumption does not affect the risk estimate seriously (1). However, one should be aware that this method still provides us more or less biased estimate of risk (1). The basic equation for calculating risk by using actuarial method directly is as follows (Equation 3):

$$ {}_{cum}R = \frac{N_{d+\,new\,cases(gp)}}{N_{all\ \ persons\,at\,risk\,(bgp)} - \frac{N_{w(gp)}}{2}} \qquad \textbf{Equation 3.} $$

${}_{cum}R$ = *cumulative risk (risk of getting a disease during the entire period)*
$N_{d+\,new\,cases\,(gp)}$ = *number of new cases of the disease under observation during a given period*
$N_{all\ persons\,at\,risk\,(bgp)}$ = *number of all persons at risk for getting ill with the disease under observation at the beginning of a given period*
$N_{w\,(gp)}$ = *number of withdrawals during a given period*

For avoiding the drawbacks of this method we could again split longer period first to shorter periods (i.e. 1-year periods) and calculate risks for these periods (i.e. annual risks). Only afterwards, on the basis of risks of shorter periods as intermediate elements, the cumulative risk is calculated indirectly. Annual risks could be calculated as follows (Equation 4):

$$ {}_{ann}R = \frac{N_{d+\ new\ cases\ (1\text{-}year\ period)}}{N_{all\ \ persons\ at\ risk\ (beginning\ of\ 1\text{-}year\ period)} - \frac{N_{w\ (1\text{-}year\ period)}}{2}} $$ **Equation 4.**

${}_{ann}R$ = *annual risk (risk of getting a disease during the 1-year period)*
$N_{d+\ new\ cases\ (1\text{-}year\ period)}$ = *number of new cases of the disease under observation during the 1-year period*
$N_{all\ persons\ at\ risk\ (beginning\ of\ 1\text{-}year\ period)}$ = *number of all persons at risk for getting ill with the disease under observation at the beginning of the 1-year period*
$N_{w\ (1\text{-}year\ period)}$ = *number of withdrawals during the 1-year period*

Estimation of this measure in practice is presented in Case study 2.

Again, a cumulative probability of survival without a disease under observation over more than one interval (2-, 3-, 4-, 5-year interval, etc.) is obtained by multiplying the partial conditional survival probabilities over all intervals (Equation 5) (12):

$$ {}_{cum}R = 1 - \left[\left(1 - {}_{ann}R_{(year1)}\right) \times \left(1 - {}_{ann}R_{(year2)}\right) \times \ldots \times \left(1 - {}_{ann}R_{(yearn)}\right)\right] $$ **Equation 5.**

${}_{cum}R$ = *cumulative risk (risk of getting a disease during the entire period of observation)*
${}_{ann}R_{(year\ 1)}$ = *annual risk (risk of getting a disease) during the 1st year*
${}_{ann}R_{(year\ 2)}$ = *annual risk (risk of getting a disease) during the 2nd year*
${}_{ann}R_{(year\ n)}$ = *annual risk (risk of getting a disease) during the nth year*

Estimation of this measure in practice is presented in Case study 2.

Because of serious limitations of this method, other methods were proposed (1).

Density method

Actuarial method is working under the assumption that all withdrawals occur on average in the middle of the observational period (1,2,11). If the periods are short, or there is a small number of withdrawals this assumption does not affect the risk estimate seriously (1). However, it is better to consider exact times of being at risk of developing a disease under observation. Another interval-based method based on the estimation of average incidence rates (person-time rate or incidence density) was proposed (1,3,4,11,12). This method depends on the functional relationship between a risk and an incidence rate (estimated through incidence density) (1).

Risk depends on incidence density and on the duration of the period of observation. Under the assumption that the cohort under observation is fixed (with no censored observations), and that the incidence density is constant over the period of observation, the risk estimate could be directly calculated as follows (Equation 6) (1,3):

$$_{cum}R = 1 - e^{\left(-ID \times t_{(gp)}\right)} \qquad \textbf{Equation 6.}$$

$_{cum}R$ = *cumulative risk (risk of getting a disease during the entire period)*
ID = *incidence density*
$t_{(gp)}$ = *duration of the given period of observation (period at risk)*

Incidence density, used in this equation was introduced in separate module in this book. It is the rate between the number of new cases which occur during the period under observation, and the quantity known under the term person-time (PT). It is calculated as (Equation 7):

$$ID = \frac{N_{d+new cases(gp)}}{PT} \qquad \textbf{Equation 7.}$$

ID = *incidence density*
$N_{d+ new cases (gp)}$ = *number of new cases of the disease under observation during a given period*
PT = *person-time*

However, usually the incidence density (as an estimate of incidence rate) does not remain constant during the entire follow-up period. Like in actuarial method, cumulative risk over a longer period also in this method is not calculated directly. We split this longer period first to shorter periods (i.e. 1-year periods) and calculate risks for these periods (partial risks), i.e. annual risks. They could be calculated as follows (Equation 8):

$$_{ann}R = 1 - e^{\left(-_{ann}ID \times 1\right)} \qquad \textbf{Equation 8.}$$

$_{ann}R$ = *annual risk (risk of getting a disease during the 1-year period)*
$_{ann}ID$ = *annual incidence density*

We can see that annual incidence densities need to be calculated prior calculation of annual risks (Equation 9):

$$_{ann}ID = \frac{N_{d+newcases(1-year period)}}{PT} \qquad \textbf{Equation 9.}$$

$_{ann}ID$ = *annual incidence density*
$N_{d+ new cases (gp)}$ = *number of new cases of the disease under observation during a 1-year period*
PT = *person-time*

Estimation of annual incidence densities and annual risks estimated by usig density method in practice is presented in Case study 3.

Only afterwards, on the basis of annual risks as intermediate elements, the cumulative risk is calculated as follows (Equation 10):

$$_{cum}R = 1 - e^{\left[\left(-_{ann}ID_{(year1)}\times 1\right)+\left(-_{ann}ID_{(year2)}\times 1\right)+\ldots+\left(-_{ann}ID_{(yearn)}\times 1\right)\right]} \qquad \textbf{Equation 10.}$$

$_{cum}R$ = *cumulative risk (risk of getting a disease during the entire period)*
$_{ann}ID_{(year\ 1)}$ = *annual incidence density during the 1st year*
$_{ann}ID_{(year\ 2)}$ = *annual incidence density during the 2nd year*
$_{ann}ID_{(year\ n)}$ = *annual incidence density during the nth year*

Estimation of this measure in practice is presented in Case study 3.

Kaplan Meier product limit method

Kaplan Meier product limit method (8,11,12) combines calculated probabilities of survival and estimates to allow censored observations, which are assumed to occur randomly. The intervals are defined as ending each time an event (i.e. disease, death, withdrawal) occurs and are therefore unequal (2,12). Again, these probabilities are conditional – they are conditioned on being at risk (present in the study without a disease under observation or censored) at each event time. The formula for calculation of conditional probability is simply (Equation 11):

$$p = \frac{N_{d+i}}{N_{persons\ at\ risk\ i}} \qquad \textbf{Equation 11.}$$

p = *conditional probability for an event in time i*
N_{d+i} = *number of events (new cases of a disease or death) occurring at time i*
$N_{persons\ at\ riski\ i}$ = *number of individuals still under observation (still at risk of the event under observation) at time i*

When time i is measured exactly, the number of events is usually 1.

The complement of this conditional probability of an event is probability of survival without an event under observation (i.e. a breakout of a disease) (12). A cumulative probability of survival without a disease under observation over more than one interval (2-, 3-, 4-, 5-year interval, etc.) is obtained by multiplying the annual conditional survival probabilities over all intervals (12).

Estimation of cumulative 5-year risk over observational period through calculation of conditional probabilities is presented in Case study 4.

CASE STUDIES

Data set

For the illustration of differences between the simple, the actuarial, the density, and the Kaplan Meier product limit method of calculation of cumulative risk an imaginary data-set is used. A cohort of 20 individuals initially without a disease under observation, were followed up for 5 years (Figure 1).

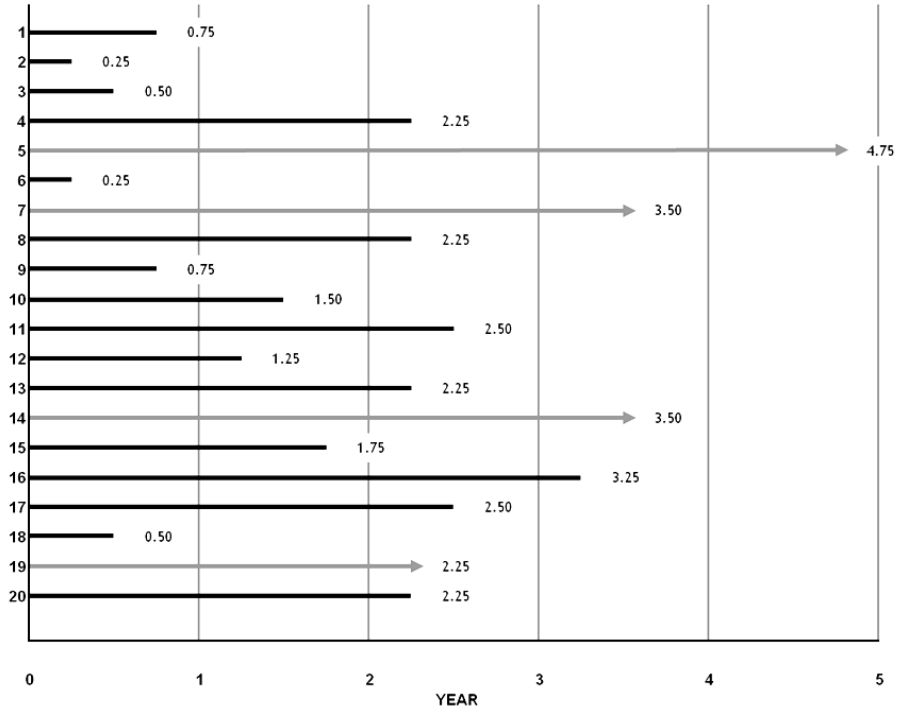

Figure 1. Graphic presentation of events in a cohort of 20 people. LEGEND: ▬ the period of exposure to the effect of the noxious agent (being at risk of developing a disease under observation before an event occurred) in individiuals that developed the disease under observation; ▬ the period of exposure to the effect of the noxious agent (being at risk of developing a disease under observation before censoring occurred) in individiuals that were lost to follow-up (voluntarily withdrawal from the study or change of domicile).

In this period, 16 individuals got a disease under observation (an event under observation) (Figure 1, persons with black lines of follow-up time), while 4 of them were lost to follow-up because of voluntarily withdrawal from the study or change of domicile (persons No. 5, 7, 14 and 19) (Figure 1, persons with gray lines of follow-up time). The lines with arrows indicate that individuals were alive at the time of the lost of follow-up.

In Figure 1 the members of a cohort are presented in order as they were numbered at the time of the entry into the study, while in Figure 2, the members are rearranged in

rank order regarding the time of an event or withdrawal. This presentation is useful in determination of times of being at risk fot the event under observation.

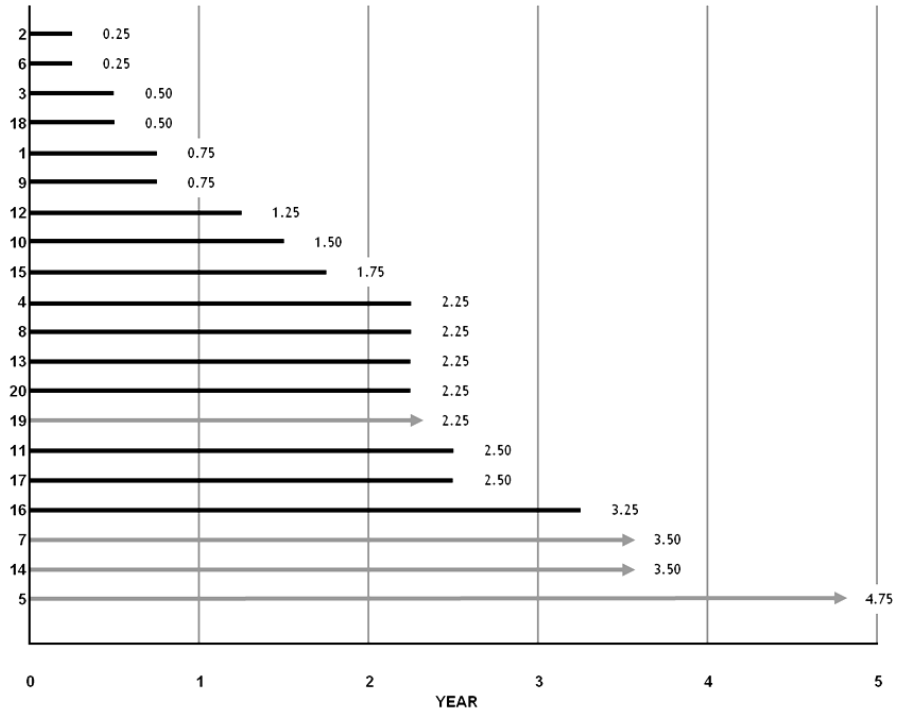

Figure 2. Ordered time of being at risk of developing a disease under observation in a cohort of 20 people from Figure 1. LEGEND: ▬ the period of exposure to the effect of the noxious agent (being at risk of developing a disease under observation before an event occurred) in individiuals that developed the disease under observation; ▬ the period of exposure to the effect of the noxious agent (being at risk of developing a disease under observation before censoring occurred) in individiuals that were lost to follow-up (voluntarily withdrawal from the study or change of domicile).

Case study 1: Estimation of cumulative risk using simple cumulative method

Results of counting of cases of observed disease which broke out during the entire 5-year time of observation (Figure 1) show that the cumulative 5-year risk estimated by the simple cumulative method according to Equation 1 is (Equation 12):

$$_{cum}R = \frac{16}{20} = 0.8000 \qquad \textbf{Equation 12.}$$

But this estimate is unreliable as there are censorings in 4/20 individuals under observation (No. 5, 7, 14 and 19) (Figure 1). In these individuals the occurrence of the event of interst is uncertain because of the termination of the observation before the event occurred. To diminish the drawbacks of this method we can split 5-year interval to 5 1-year intervals, and for each 1-year interval we calculate the annual risk by following next steps:

- define the number of persons entered in the interval (Table 1, column 1), number of persons with the disease at the end of interval (Table 1, column 2), and the number of losts (withdrawals) (Table 1, column 3),
- by using Equation 2 calculate annual risks (Table 1, column 4).

From the Table 1 it could be seen that in case of calculation of annual risks, the censoring is partially considered even when using simple cumulative method, as we need to define separately for every year the number of individuals at risk, and all participants who terminated the observation because of extraneous factors (e.g. death because of traffic accident etc.) are not included.

Table 1. Elements for calculation and calculation of annual risks using simple cumulative method.

Year of observation	**1**	**2**	**3**	**4**
	Entered in the interval (N)	**With the disaese at the end of interval (d+)**	**Lost**	**d+/N (annual risk) ($_{ann}$R)**
1st	20	6	0	0.3000
2nd	14	3	0	0.2143
3rd	11	6	1	0.5455
4th	4	1	2	0.2500
5th	1	0	1	0.0000

The annual risk (Table 1, column 4) is annual probability of the event (12). The complement of this probability is annual probability of survival without an event under observation (i.e. a breakout of a disease) (Table 2, column 5). Technically these probabilities are annual conditional probabilities. A cumulative probability of survival without a disease under observation over more than one interval (2-, 3-, 4-, and 5-year interval) is obtained by multiplying the annual conditional survival probabilities over all intervals (Table 2, column 6) (12).

Table 2. Calculation of cumulative 5-year risk from annual risks using simple cumulative method.

Year of observation	**4**	**5**	**6**	**7**
	d+/N (annual risk) ($_{ann}$R)	**1 - $_{ann}$R**	**product (1 - $_{ann}$R) (Π)**	**1 – Π (cumulative risk) ($_{cum}$R)**
1st	0.3000	0.7000	0.7000	0.3000
2nd	0.2143	0.7857	0.5500	0.4500
3rd	0.5455	0.4545	0.2500	0.7500
4th	0.2500	0.7500	0.1875	0.8125
5th	0.0000	1.0000	0.1875	0.8125

The cumulative probability of having an event is the complement of joint probability of survival through every of five years of observation (Table 2, column 7) (12).

Case study 2: Estimation of cumulative risk using actuarial method

Simple cumulative method assumes no withdrawals during the period of observation. Since in our case (Figures 1 and 2) there were four individuals lost to observation, this must be considered. Their limited participation need to be considered in the denominator of the cumulative probability of an event. Actuarial method considers censored observations most roughly (Equation 3). Since we have at the end of the 5-year interval 16 individuals with a disease out of 20 persons at the beginning of the observation, and 4 persons were lost to follow up, we calculate cumulative 5-year risk directly as (Equation 13):

$$_{cum}R = \frac{16}{20 - \frac{4}{2}} = 0.8889 \qquad \textbf{Equation 13.}$$

Again, we can split 5-year interval first into five 1-year intervals and calculate first the annual risks and afterwards cumulative 5-year risk. For each 1-year interval we:

- define the number of persons entered in the interval (Table 3, column 1), number of persons with the disease at the end of interval (Table 3, column 2), and the number of withdrawals (Table 3, column 3),
- calculate the adjusted number of withdrawals (1,12),
- by using Equation 4 calculate annual actuarial risks (Table 3, column 6).

Table 3. Elements for calculation and calculation of annual risks using actuarial method.

Year of observation	1	2	3	4	5	6
	Entered in the interval (N)	With the disaese at the end of interval (d+)	Withdrawals (W)	W/2	N – (W/2)	d+/N–(W/2) (annual risk) ($_{ann}$R)
1st	20	6	0	0	20	0.3000
2nd	14	3	0	0	14	0.2143
3rd	11	6	1	0.5	10.5	0.5714
4th	4	1	2	1	3	0.3333
5th	1	0	1	0.5	0.5	0.0000

After annual risks are calculated we follow exactly the same priciples for calculation of 2-, 3-, 4- and 5-year cumulative risks as discussed in simple method. The results are presented in Table 4. Results of calculating the cumulative 5-year risk estimated by using the actuarial method (Table 4, column 9) show that its value is 0.8428, what is much higher than estimated by using the simple method.

Table 4. Calculation of cumulative 5-year risk from annual risks using actuarial method.

Year of observation	6	7	8	9
	d+/N – (W/2) (annual risk) ($_{ann}$R)	1 – $_{ann}$R	product (1 - $_{ann}$R) (II)	1 – II (cumulative risk) ($_{cum}$R)
1st	0.3000	0.7000	0.7000	0.3000
2nd	0.2143	0.7857	0.5500	0.4500
3rd	0.5714	0.4286	0.2357	0.7643
4th	0.3333	0.6667	0.1572	0.8428
5th	0.0000	1.0000	0.1572	0.8428

Case study 3: Estimation of cumulative risk using density method

The first method that consider exact times of being at risk of developing a disease under observation is density method.

Table 5. Data for calculation of person-years.

Id. number	Time of being at risk* (Years)	Status at the end of observation (1=with the disease, 0=cesored (cause of censoring))
2	0.25	1
6	0.25	1
3	0.50	1
18	0.50	1
1	0.75	1
9	0.75	1
12	1.25	1
10	1.50	1
15	1.75	1
4	2.25	1
8	2.25	1
13	2.25	1
20	2.25	1
19	2.25	0 – free of disease, change of domicile
11	2.50	1
17	2.50	1
16	3.25	1
7	3.50	0 – free of disease, voluntarily withdrawal
14	3.50	0 – free of disease, change of domicile
5	4.75	0 – free of disease, change of domicile
Total	38.75	Diseased = 16, Lost-to-follow-up = 4

* time in which an individual under observation is exposed to effect of noxious agent (is at risk of getting an event under observation)

In order to perform the procedure (Equation 6) we need first to calculate the person-years (PY) since we need this quantity in calculation of the incidence density. We use the information given in Figure 2. In Table 5 data for calculation of PY for the entire 5-year period are presented.

The incidence density for 5-year period could be now calculated using the Equation 7. The results are presented in following equation (Equation 14):

$$ID = \frac{16}{38.75} = 0.4129 \qquad \textbf{Equation 14.}$$

This quantity afterwards enters the equation for calculating the 5-year cumulative risk using the Equation 6. The results are presented in following equation (Equation 15):

$$_{cum}R = 1 - e^{(-0.4129 \times 5)} = 0.8731 \qquad \textbf{Equation 15.}$$

Again, we can split 5-year interval first into five 1-year intervals and calculate first the annual risk using the density method and afterwards cumulative 5-year risk. The steps are as follows

- first we summarize the events in each of 1-year intervals which are five as the duration of the longest observation is 4.75 let: entered in the interval (Table 6, column 1), with the disease at the end of interval (Table 6, column 2), lost to follow-up (Table 6, column 3), and present at the end of the period without a disease (Table 6, column 4),

Table 6. Summary of the events in each of 1-year intervals.

Year of observation	**1**	**2**	**3**	**4**
	Entered in the interval (N)	**With the disaese at the end of interval (d+)**	**Lost to follow-up**	**Present at the end of the period**
1st	20	6	0	14
2nd	14	3	0	11
3rd	11	6	1	4
4th	4	1	2	1
5th	1	0	1	0
Total		16	4	

- in following step we calculate the person-years (PY) for for each of 1-year periods (Table 7). We use the information given in Figure 2,

Table 7. The summary of calculation of person-years in each of 5 1-year intervals.

Year of observation	Contribution to person-years (PY) at the end of 1-year interval	PY Total
1st	(0.25 × 2) + (0.50 × 2) + (0.75 × 2) + (1.00 × 14)	17.00
2nd	(0.25 × 1) + (0.50 × 1) + (0.75 × 1) + (1.00 × 11)	12.50
3rd	(0.25 × 5) + (0.50 × 2) + (1.00 × 4)	6.25
4th	(0.25 × 1) + (0.50 × 2) + (1.00 × 1)	2.25
5th	(0.75 × 1)	0.75

- in following step the annual incidence density is calculated (Table 8). As the incidence density is not constant over 5-year period (the highest is in the third year of observation) this has to be considered in the calculation of cumulative risk,
- at the final step from incidence density the risk is calculated (Table 9).

Table 8. Calculation of incidence density in each of 5 1-year intervals.

Year of observation	2	5	6
	With the disaese at the end of interval (d+)	Annual person-years (PY)	Annual incidence density (d+/PY) ($_{ann}$ID)
1st	6	17.00	0.3529
2nd	3	12.50	0.2400
3rd	6	6.25	0.9600
4th	1	2.25	0.4444
5th	0	0.75	0.0000

Table 9. Calculation of the annual risk in each of 5 1-year intervals.

Year of observation	6	7	8
	Annual incidence density (d+/PY) ($_{ann}$ID)	$e^{(-IDann\times1)}$	$1-e^{(-annID\times1)}$ (annual risk) ($_{ann}$R)
1st	0.3529	0,.7027	0.2974
2nd	0.2400	0.7866	0.2134
3rd	0.9600	0.3829	0.6171
4th	0.4444	0.6412	0.3588
5th	0.0000	1.0000	0.0000

Results of calculating the cumulative 5-year risk estimated by using the density method (Figure 1) show that its value is 0.8643, what is much higher than estimated using the simple method, and also higher than estimated using the actuarial method. The elements for calculation and its results are presented in Table 10.

Table 10. Elements for calculation of cumulative risk using the density method.

	6	9	10	11
Year of observation	**Annual incidence density (d+/PY) ($_{ann}$ID)**	**$\sum(-_{ann}ID\times1)$**	**$e^{\sum(-annID\times1)}$**	**$1-e^{\sum(-annID\times1)}$ (cumulative risk) ($_{cum}$R)**
1st	0.3529	-0.3529	0.7026	0.2974
2nd	0.2400	-0.5929	0.5527	0.4473
3rd	0.9600	-1.5529	0.2116	0.7884
4th	0.4444	-1.9973	0.1357	0.8643
5th	0.0000	-1.9973	0.1357	0.8643

Case study 4: Estimation of cumulative risk using Kaplan Meier product limit method

This method also considers exact times of being at risk of developing a disease under observation (2,12). The intervals are defined as ending each time an event (i.e. disease, death, withdrawal) occurs. The procedure is as follows:

- first we determine the times when events or censoring occurred. We use the information given in Figure 2,
- define the number of persons entered in the interval (Table 11, column 1), number of persons with the event (occurrence of the disease or death) at time I (Table 11, column 2), and the number of censored cases (Table 11, column 3) at time i,

Table 11. Elements for calculation of cumulative risk by using the Kaplan Meier product limit method.

Time of the events/ censoring (years)	**1**	**2**	**3**	**4**	**5**	**6**	**7**
	Entered in the interval (N)	**Occurrence of the event (d+)**	**Censored**	**d+/N (conditional probability of the event) (p)**	**1 – p (survival) (S)**	**Product (S) (Cumulative survival) ($_{cum}$S)**	**$1 - S_{cum}$ (cumulative conditional probability of an event) ($_{cum}$R)**
0.25	20	2	0	0.1000	0.9000	0.9000	0.1000
0.50	18	2	0	0.1111	0.8889	0.8000	0.2000
0.75	16	2	0	0.1250	0.8750	0.7000	0.3000
1.25	14	1	0	0.0714	0.9286	0.6500	0.3500
1.50	13	1	0	0.0769	0.9231	0.6000	0.4000
1.75	12	1	0	0.0833	0.9167	0.5500	0.4500
2.25	11	4	1	0.3636	0.6364	0.3500	0.6500
2.50	6	2	0	0.3333	0.6667	0.2333	0.7667
3.25	4	1	0	0.2500	0.7500	0.1750	0.8250
3.50	2	0	2	0.0000	1.0000	0.1750	0.8250
4.75	1	0	1	0.0000	1.0000	0.1750	0.8250

- by using Equation 11 calculate conditional probalities (Table 11, column 4),
- calculate the complement of conditional probalities of the event at every time of occurrence of the events or censoring – the conditional probability of survival without an event under observation up to the time i (Table 11, column 5),
- calculate cumulative probability of survival over more than one interval by multiplying the conditional survival probabilities over all intervals (Table 11, column 6),
- calculate the complement of cumulative probabilities of survival over more than one interval (Table 11, column 7).

Conclusion

In table 12 the summary over results of all four methods of estimation of cumulative risk is presented.

Table 12. Summary over results of estimating cumulative risk over 5-year period using four different methods of estimation.

Method	Direct 5-year cumulative risk	Indirect 5-year cumulative risk
Simple	0.8000	0.8125
Actuarial	0.8889	0.8429
Density	0.8731	0.8643
Kaplan Meier		0.8250

Since the most accurate measure is Kaplan Meier method we could compare all other results to this result. We could conclude that in this case study, the closest results to Kaplan Meier method are obtained by indirect simple method, and by actuarial indirect method, while the most far away were results obtained by direct actuarial method. One should be aware that this is not always so. The results depend on number of events and number of censored cases. When the events are rare and there is no censoring, the discrepancy tends to be smaller (12).

EXERCISE

Data set

In Figure 3, another imaginary data-set is presented. Again, a cohort of 20 individuals initially without a disease under observation, were followed up for 5 years.

Task 1

For the data set presented in Figure 3, calculate cumulative risk using simple method:

- directly,
- indirectly by calculating annual risks first.

Task 2

For the data set presented in Figure 3, calculate cumulative risk using actuarial method:

- directly,
- indirectly by calculating annual risks first.

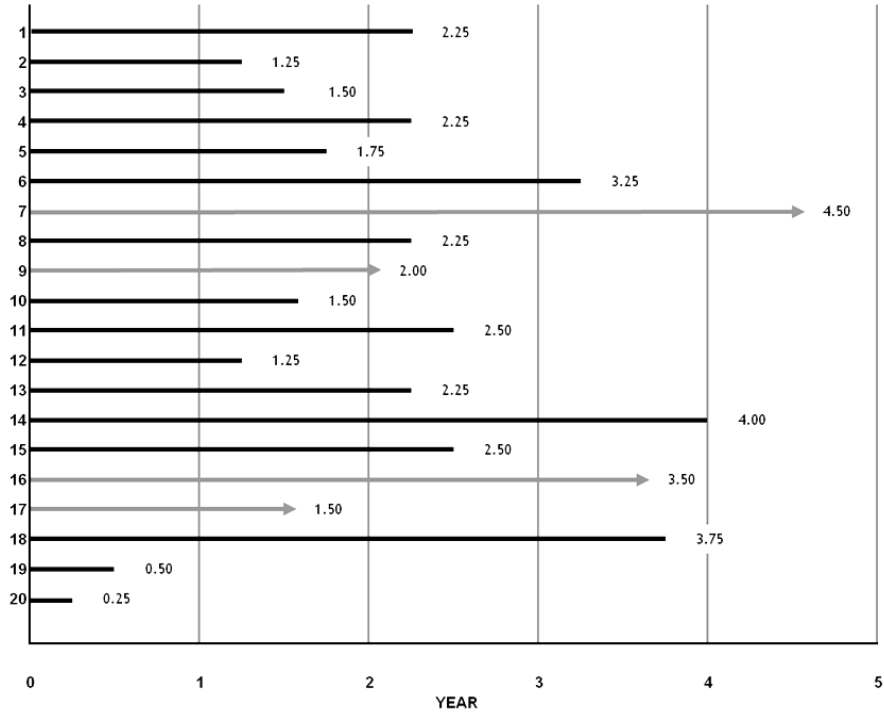

Figure 3. Graphic presentation of events in a cohort of 20 people. LEGEND: ▬ the period of exposure to the effect of the noxious agent (being at risk of developing a disease under observation before an event occurred) in individuals that developed the disease under observation; ▬ the period of exposure to the effect of the noxious agent (being at risk of developing a disease under observation before censoring occurred) in individuals that were lost to follow-up (voluntarily withdrawal from the study or change of domicile).

Task 3

For the data set presented in Figure 3, calculate cumulative risk using density method:

- directly,
- indirectly by calculating annual risks first.

Task 4

For the data set presented in Figure 3, calculate cumulative risk using Kaplan Meier method.

REFERENCES

1. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. New York: John Wiley&Sonc, cop.; 1982.
2. Last JM, editor. A dictionary of epidemiology. Oxford: Oxford University Press; 2001.
3. Dos Santos Silva I. Cancer epidemiology: principles and methods. Lyon: IARC; 1999. Available from URL: http://www.iarc.fr/en/publications/pdfs-online/epi/cancerepi/CancerEpi.pdf. Accessed: June 24, 2009.
4. Jekel JF, Katz DL, Elmore JG. Epidemiology, biostatistics, and preventive medicine. Philadelphia: W.B. Saunders Company; 2001.
5. Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Company; 1987.
6. Friis RH, Sellers TA. Epidemiology for public health practice. Fourth edition. Sudbury, MA: Jones&Bartlett Publishers; 2009.
7. Kue Youtng T. Population health: concepts and methods. Second edition. Oxford: Oxford University Press; 2005.
8. Kirkwood BR, Sterne JAC. Essential medical statistics. 2nd ed. Malden MA: Blackwell Science, Inc.; 2005.
9. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Third edition. London: Lippincott, Williams&Wilkins; 2008.
10. Kleinbaum DG, Klein M. Survival analysis. A Self-Learning Text. Second edition. New York: Springer; 2005.
11. Parmar M, Machin D. Survival analysis. A practical approach. Chichester: John Wiley&Sons; 1995.
12. Szklo M, Nieto FJ. Epidemiology: Beyond the basics. Gaithersburg, MA: Aspen Publishers; 2005.
13. Schoenbach VJ, Rosamond WD. Understanding the fundamentals of epidemiology - an evolving text [Internet]. Chapel Hill: University of North Carolina, Epidemiology.net; 2000. Available from URL: http://www.epidemiolog.net/evolving/TableOfContents.htm. Accessed: June 25, 2009.

RECOMMENDED READINGS

1. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. New York: John Wiley&Sonc, cop.; 1982.
2. Parmar M, Machin D. Survival analysis. A practical approach. Chichester: John Wiley&Sons; 1995.
3. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Third edition. London: Lippincott, Williams&Wilkins; 2008.
4. Szklo M, Nieto FJ. Epidemiology: Beyond the basics. Gaithersburg, MA: Aspen Publishers; 2005.