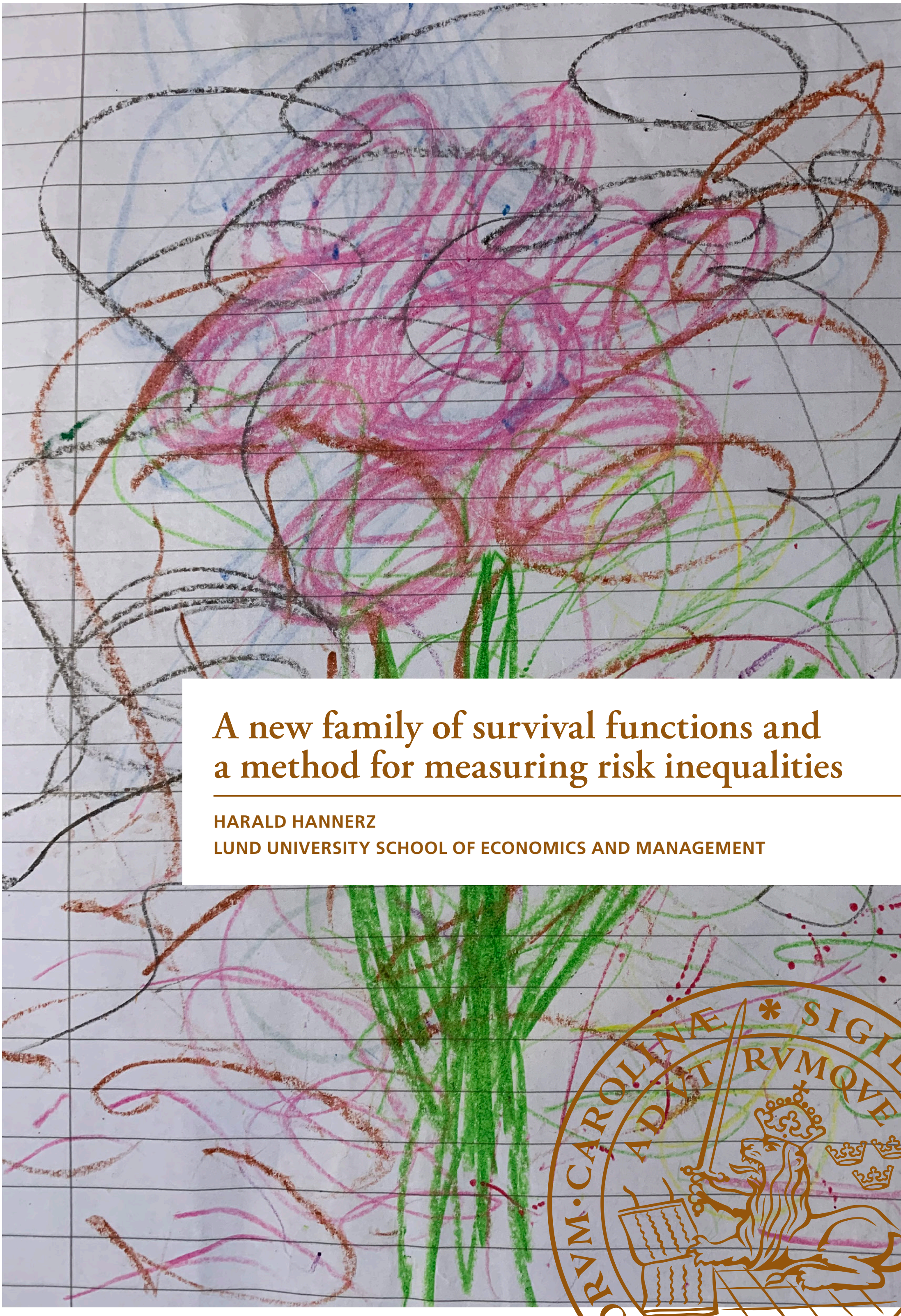
A new family of survival functions and a method for measuring risk inequalities

Hannerz, Harald

2023



A new family of survival functions and a method for measuring risk inequalities

HARALD HANNERZ
LUND UNIVERSITY SCHOOL OF ECONOMICS AND MANAGEMENT

A new family of survival functions and a method for measuring risk inequalities

Harald Hannerz

DOCTORAL DISSERTATION

by due permission of the Department of Statistics, School of Economics and Management, Lund University, Sweden. To be defended at room EC3:207, Holger Crafoord Centre, Lund on 31 May 2023, at 13:15.

Faculty opponent

Professor Anastasia Kostaki, Athens University of Economics and Business

Organization LUND UNIVERSITY

Document name Doctoral Dissertation

Date of issue 2023-05-21

Author(s): Harald Hannerz

Sponsoring organization

Title and subtitle: A new family of survival functions and a method for measuring risk inequalities.

Abstract

The present compilation thesis is divided into two sections, one for each of two separate methodological issues: reduction of random errors in mortality estimates and offsetting random variation related bias in data generated estimators of risk inequalities.

Summary of section A

In regard to the first issue a new family of survival functions is proposed. Its purpose is to provide valid and reliable age-specific estimates of death probabilities and life expectancies for all ages in the entire human life span. In Paper I, I introduced a five-parameter survival function intended to model mortality in modern female populations. It is shown that (i) the complement of the proposed survival function is a bona fide cumulative distribution function, and (ii) that the expected value of a random variable with such a distribution exists and is finite.

In Paper II, I showed that the age pattern of mortality among Swedish males differed significantly from the age pattern among Swedish females and that some extra parameters were needed to accommodate an added risk of fatal injuries among males in the early adulthood. To address this shortcoming, I introduced an eight-parameter survival function intended to model mortality among males as a solution to the problem associated with the gender difference in mortality patterns.

In Paper III, I addressed the use of collateral data as a means of improving the statistical precision of mortality estimates. A brief description of three main approaches that actuaries and demographers use to accomplish such improvements, namely, mortality laws, model life tables, and relational methods was given. I thereafter introduced a novel regression model that incorporates several beneficial principles from each of these approaches.

The survival functions introduced in [Hannerz, 1999] resulted in an, on average, five-fold decrease in the standard error of estimated sex and age-specific one-year death probabilities, compared with frequency substitution estimates, when applied to mortality in the total population of Sweden 1982.

In papers IV – VI, I applied the methods delineated in Papers I – III to estimate age, sex and diagnosis-specific life expectancies among individuals with a history of psychiatric and cerebrovascular disorders, respectively.

Summary of section B

The second methodological issue studied resulted in a Monte Carlo simulation procedure, which can be used to estimate excess fractions in the absence of a natural reference group. The procedure is based on the assumptions that the number of events in each group is Poisson distributed and that the true risk rates in the groups increases geometrically with their rank order. The methodological aspects of the procedure are described and discussed in Paper VII.

In paper VIII the procedure is applied to industrial inequalities in rates of disability retirement and in paper IX to hospital contact due to mood disorders, in both studies among economically active people in Denmark.

The Monte Carlo simulation procedure is designed to estimate excess fractions in situations where no natural reference group exists. Simpler methods are available when a reference group does exist. An overview of such measures is included in the thesis, and examples of excess fractions in relation to pre-specified reference groups are given in Papers X and XI.

Key words: Mortality law; survival analysis; risk inequalities; excess fraction

Classification system and/or index terms (if any)

Supplementary bibliographical information		**Language**
ISSN and key title		**ISBN** 978-91-8039-719-3
Recipient's notes	**Number of pages** 234	Price
	Security classification	

Signature Date 2023-04-24

A new family of survival functions and a method for measuring risk inequalities

Harald Hannerz

LUND UNIVERSITY

Cover pictures by Anton Hannerz (front) and Andreas Hannerz (back).

School of Economics and Management
Department of Statistics

ISBN 978-91-8039-719-3 (print)
978-91-8039-720-9 (pdf)

Printed in Sweden by Media-Tryck, Lund University
Lund 2023

Media-Tryck is an environmentally certified and ISO 14001:2015 certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

For my grandchildren

Table of Contents

Preface

The present dissertation contains eleven research papers that were published in between 2000 and 2020, at which time I was employed as a statistician at the National Research Centre for the Working Environment (NRCWE) in Denmark. Six of the papers were derived from a monograph, which I defended as a Licentiate thesis at Lund University in January 2000.

Permissions to reproduce the papers have been granted by each of the concerned copyright owners.

I thank each of the following researchers for having contributed as a co-author to one or more of the attached research papers: Per Borgå, Marianne Borritz, Martin Lindhardt Nielsen, Karen Albertsen, Ida Elisabeth Huitfeldt Madsen, Andreas Holtermann, Helene Feveile, Kim Lyngby Mikkelsen, Ole Olsen, Finn Tüchsen, Søren Spangenberg, Betina Nørgaard, Johnny Dyreborg, Reiner Rugulies, Bryan Cleal, Kjeld Børge Poulsen, and Lars Louis Andersen.

I thank the following institutions for having contributed data to one or more of the research projects covered: The Swedish National Board of Health and Welfare, The Swedish Insurance Federation, Statistics Sweden, The Danish Health Data Authority, The Danish National Institute for Health Data and Disease Control, NRCWE, and Statistics Denmark.

I thank all participants who responded to the questionnaire of the Danish Work Environment Cohort Study 2010, and thereby contributed to the last research paper of the dissertation.

I thank all the editorial staff and peer-reviewers that were involved in the publication of the attached research papers.

I thank the librarians Elizabeth Bengtsen and Rikke Sverker Nilsson (NRCWE), who have provided me with indispensable literature, throughout the above mentioned time period.

Last but not least, I extend my gratitude to Krzysztof Podgórski, Krzysztof Nowicki, Jakob Bergman and Björn Holmquist at the Department of Statistics, Lund University, for providing me with the supervision and instruction necessary to complete my doctoral programme.

Harald Hannerz

Copenhagen, April 2023

List of acronyms and abbreviations

ARIMA:	Autoregressive integrated moving average
Cf:	Confer
CI:	Confidence interval
EF:	Excess fraction
E.g.:	Exempli gratia
Et al.:	Et alia
Eq.:	Equation
GI:	Gini index
ID:	Index of dissimilarity
I.e.:	Id est
IHD:	Ischaemic heart disease
KMI:	Kunst-Mackenbach index
No.:	Number
PAR:	Population attributable fraction
RCI:	Relative concentration index
RII:	Relative index of inequality
SES:	Socioeconomic status
SII:	Slope index of inequality
SMR:	Standardised morbidity ratio
SRR:	Standardised rate ratio
US:	United States
Vs:	Versus
WHO:	World Health Organisation

The structure of the dissertation

The present dissertation consists of a set of introductory chapters that summarises a collection of 11 published papers. The following papers are included:

Methodology

Paper I

Hannerz H. Presentation and derivation of a five-parameter survival function intended to model mortality in modern female populations. Scandinavian Actuarial Journal 2001;101:176-187.

Paper II

Hannerz H. Manhood trials and the law of mortality. Demographic Research 2001;4:185-202.

Paper III

Hannerz H. An extension of relational methods in mortality estimation. Demographic Research 2001;4:337-367.

Applications

Paper IV

Hannerz H, Borgå P, Borritz M. Life expectancies for individuals with psychiatric diagnoses. Public Health 2001;115:328-337.

Paper V

Hannerz H, Nielsen ML. Life expectancies among survivors of acute cerebrovascular disease. Stroke 2001;32:1739-1744.

Paper VI

Hannerz H, Borgå P. Mortality among persons with a history as psychiatric inpatients with functional psychosis. Soc Psychiatry and Psychiatr Epidemiol 2000;35:380-387.

Methodology

Paper VII

Feveile H, Mikkelsen KL, **Hannerz H**, Olsen O. Quantifying inequality in health in the absence of a natural reference group. Sci Total Environ. 2006;367:112-22.

Applications Paper VIII

Hannerz H, Tüchsen F, Spangenberg S, Albertsen K. Industrial differences in disability retirement rates in Denmark 1996 - 2000. IJOMEH, 2004;17:465-471.

Paper IX

Hannerz H, Tüchsen F, Pedersen BH, Dyreborg J, Rugulies R, Albertsen K. Work-relatedness of mood disorders in Denmark. Scand J Work Environ Health. 2009;35(4):294-300.

Paper X

Cleal B, **Hannerz H**, Poulsen K, Andersen LL. Socio-economic status and incident diabetes mellitus among employees in Denmark: a prospective analysis with 10-year follow-up. Diabet Med. 2014 Dec;31(12):1559-62.

Paper XI

Hannerz H, Holtermann A, Madsen IEH. Musculoskeletal pain as a predictor for depression in the general working population of Denmark. Scand J Public Health. 2020 Jan 23:1403494819875337. doi: 10.1177/1403494819875337. [Epub ahead of print]

1. Summary

The present compilation thesis is divided into two sections, one for each of two separate methodological issues: A) reduction of random errors in mortality estimates and B) offsetting of random variation related bias in data generated estimators of risk inequalities. The papers of section A are addressed in chapter 2. The papers of section B are addressed in chapter 3. The concluding remarks of the thesis are given in chapter 4.

Summary of section A

In regard to the first issue a new family of survival functions is proposed. Its purpose is to provide valid and reliable age-specific estimates of death probabilities and life expectancies for the entire human life span. In Paper I, I introduced a five-parameter survival function intended to model mortality in modern female populations. It is shown that (i) the complement of the proposed survival function is a bona fide cumulative distribution function, and (ii) that the expected value of a random variable with such a distribution exists and is finite.

In Paper II, I showed that the age pattern of mortality among Swedish males differed significantly from the age pattern among Swedish females and that some extra parameters were needed to accommodate an added risk of fatal injuries among males in the early adulthood. To address this shortcoming, I introduced an eight-parameter survival function intended to model mortality among males as a solution to the problem associated with the gender difference in mortality patterns.

In Paper III, I addressed the use of collateral data as a means of improving the statistical precision of mortality estimates. Brief descriptions were given of three main approaches that actuaries and demographers use to accomplish such improvements, namely, mortality laws, model life tables, and relational methods. I thereafter introduced a novel regression model that incorporates several beneficial principles from each of these approaches.

The survival functions introduced in [Hannerz, 1999] resulted in an, on average, five-fold decrease in the standard error of estimated sex and age-specific one-year death probabilities, compared with frequency substitution estimates, when applied to mortality in the total population of Sweden 1982.

In papers IV – VI, I applied the methods delineated in Papers I – III to estimate age, sex and diagnosis-specific life expectancies among individuals with a history of psychiatric and cerebrovascular disorders, respectively.

Summary of section B

The second methodological issue studied resulted in a Monte Carlo simulation procedure, which can be used to estimate excess fractions in the absence of a natural reference group. The procedure is based on the assumptions that the number of events in each group is Poisson distributed and that the true risk rates in the groups increases geometrically with their rank order. The methodological aspects of the procedure are described and discussed in Paper VII.

In paper VIII the procedure is applied to industrial inequalities in rates of disability retirement and in paper IX to hospital contact due to mood disorders, in both studies among economically active people in Denmark.

The Monte Carlo simulation procedure is designed to estimate excess fractions in situations where no natural reference group exists. Simpler methods are available when a reference group does exist. An overview of such measures is included in the thesis, and examples of excess fractions in relation to pre-specified reference groups are given in Papers X and XI.

2. A new family of mortality laws

Figure 2.1. Photo by Veronika Valdova from Pexels [https://www.pexels.com/photo/d-day-cemetery-930711/]

Sometimes statisticians are confronted with problem in estimating age-specific life expectancies because data do not exist for all ages or the numbers of observations in single age groups are too small to allow meaningful age-stratified estimates by use of directly observed mortality rates. To obtain reasonable estimates in such situations, the information base of the estimators needs to be supplemented with collateral data.

Potential collateral data might be, for example, mortality among people in other age groups than those in the population of interest. To make use of this information we need a mortality law i.e. a mathematical expression that describes mortality as a function of age.

Collateral data may also be obtained from mortality observations in populations other than the one of interest. To make use of such information we would need a relational method i.e. a mathematical expression which relates mortality in one population to that in others.

Actuaries and demographers have a long tradition of using collateral data to improve mortality estimates, and a great many mortality laws have been proposed [Gompertz, 1825; Makeham, 1867; Thiele, 1872; Wittstein, 1883; Pearson, 1895; Perks, 1932; Brillinger, 1961; Heligman and Pollard, 1980; Petrioli, 1981; Mode and Busby, 1982; Siler, 1983; Anson, 1988; Kostaki, 1992; Hannerz, 1999; Boulougari et al., 2019]. Relational methods [Derrick, 1927; Brass, 1969, 1974; Zaba, 1979; Ewbank et al., 1983; Lee and Carter, 1992; Renshaw and Haberman, 2006; Neves et al, 2017] have also been developed.

Section A of the present thesis is devoted to some of my own contributions to these two research areas. Paper I introduces a five-parameter survival function intended to model mortality in modern female populations, Paper II introduces an eight-parameter survival function intended to model mortality among males, and Paper III introduces a reducible five-parameter relational method which is to be used in conjunction with the proposed survival functions. The methods described in the first three papers are then used in Paper IV – VI to estimate life expectancies among individuals with psychiatric and cerebrovascular diagnoses, respectively.

2.1 On mortality laws

Name	Year	Model	No. of parameters
De Moivre	1725	$l(x)=1-\frac{x}{w}$	1
Gompertz	1825	$\mu(x)=bc^x$	2
Makeham	1860	$\mu(x)=a+bc^x$	3
Thiele	1871	$\mu(x)=a_1\exp[-b_1x]+a_2\exp\left[-b_2\frac{(x-c)^2}{2}\right]+a_3\exp[b_3x]$	7
Wittstein	1883	$q_x=\frac{1}{m}a^{-(mx)^n}+a^{-(M-x)^n}$	4
Perks	1932	$\mu(x)=\frac{a+bc^x}{kc^{-x}+1+dc^x}$	5
Heligman and Pollard	1980	$\frac{q_x}{1-q_x}=A^{(x+B)^C}+De^{-E\left(\log\left(\frac{x}{F}\right)\right)^2}+GH^x$	8
Petrioli	1981	$\frac{1-l(x)}{l(x)}=kx^a(w-x)^b\exp[cx+dx^2]$	6
Siler	1983	$\mu(x)=a_1\exp[-b_1x]+a_2+a_3\exp[b_3x]$	5
Kostaki	1992	$\frac{q_x}{1-q_x}=\begin{cases}A^{(x+B)^C}+De^{-E_1\left(\log\left(\frac{x}{F}\right)\right)^2}+GH^x, \text{ for } x\le F\\ A^{(x+B)^C}+De^{-E_2\left(\log\left(\frac{x}{F}\right)\right)^2}+GH^x, \text{ for } x>F.\end{cases}$	9
Boulougari et al.	2019	$\mu(x)=\frac{c_1}{xe^{-c_2x}}+a_1(xe^{-a_2x})^{a_3}$	5

x = age, w = highest attainable age, M=w-1, $l(x)$ = probability of surviving to age x, $\mu(x)$ = hazard rate ate age x, q_x = probability of dying before age x+1 given survival to age x.

Figure 2.2. Mortality laws

The age pattern of mortality has been studied and documented for centuries. Billions of person years have been observed and several regression models have been developed [cf. reviews by Kostaki, 1992; Booth and Tickle, 2006; Pascariu, 2018]. In Figure 2.2, I give a list of some significant developments (other than my own) in this field. The models, which customarily are referred to as laws of mortality [Hartmann 1987], are listed in chronological order. The first one, by De Moivre in 1725, which was based on the rather naive assumption that death ages are uniformly distributed in a finite age interval, is of course not very useful. But the laws of Gompertz (1825), Makeham (1867) and Perks (1932) can be used to graduate mortality among adults, and the laws of Thiele (1872), Wittstein (1883), Heligman-Pollard (1980), Petrioli (1981), Siler (1983), Kostaki (1992) and Boulougari et al. (2019) are designed to graduate mortality in the entire human life span.

The rationale behind the various formulas in Figure 2.2 and the purpose of their parameters are detailed in the references cited above. The application of a law is, however, often independent of the interpretation of its parameters, which for most practical purposes may be viewed merely as intermediate values that need to be determined in order to obtain interpretable measures such as life expectancies, and probabilities for death as a function of age and risk period.

The mortality laws that are presented and used in section A of the present thesis [Paper I – VI] are given by the following equations:

$$F(x) = \frac{e^{G(x)}}{1+e^{G(x)}}$$

$$where \tag{2.1}$$

$$G(x) = a_0 - a_1 x^{-1} + \frac{a_2}{2} x^2 + \frac{a_3}{c} e^{cx}$$

$$F(x) = \alpha \frac{e^{G_1(x)}}{1+e^{G_1(x)}} + (1-\alpha) \frac{e^{G_2(x)}}{1+e^{G_2(x)}}$$

$$G_1(x) = a_0 - a_1 x^{-1} + \frac{a_2}{2} x^2 + \frac{a_3}{c} e^{cx} \tag{2.2}$$

$$G_2(x) = a_4 - a_5 x^{-1} + \frac{a_2}{2} x^2 + \frac{a_3}{c} e^{cx},$$

where $F(x)$ is the probability and $G(x)$ is the log-odds for death before age x. All of the model parameters are real-valued. The parameters a_0 and a_4 may be negative, while the rest of the parameters must be positive and α must be less than or equal to one. Equation 2.1 was developed to model the mortality of females, based on the assumption that the derivative of the log-odds, dG/dx, is inversely proportional to age squared in the early childhood, proportional to age among the labour force, and exponentially increasing in the age of senescence. The mixture distribution given by Equation 2.2 is an enlargement of Equation 2.1 to be used for males. The purpose of the enlargement is to handle an added mortality risk associated with the passage into manhood (the accident hump), and the factor (1-α) in Equation 2.2 may loosely be interpreted as the expected proportion of new-born boys that will die as a direct consequence of that added risk.

It has been shown that Equation 2.1 and 2.2 fit data well when applied to the female and male populations of Sweden, respectively, with the age divided into one-year intervals. It has however been pointed out that the equations fail to accurately describe the age pattern of mortality within the age interval (0,1) years. Although this problem is seldom of importance to actuaries, demographers and epidemiologists, I added a parameter in 1998 to rectify this shortcoming [Hannerz, 1999]. With the added parameter, the respective laws would be expressed by the following equations:

$$\begin{aligned} &F(x) = \frac{e^{G(x)}}{1+e^{G(x)}}, \\ &where \\ &G(x) = b_0 - b_1 x^{-1} + b_2 \log(x) + \frac{b_3}{2} x^2 + \frac{b_4}{c} e^{cx} \end{aligned} \tag{2.3}$$

$$\begin{aligned} &F(x) = \alpha \frac{e^{G_1(x)}}{1+e^{G_1(x)}} + (1-\alpha) \frac{e^{G_2(x)}}{1+e^{G_2(x)}} \\ &G_1(x) = b_0 - b_1 x^{-1} + b_2 \log(x) + \frac{b_3}{2} x^2 + \frac{b_4}{c} e^{cx} \\ &G_2(x) = b_5 - b_6 x^{-1} + \frac{b_3}{2} x^2 + \frac{b_4}{c} e^{cx} \end{aligned} \tag{2.4}$$

It has been shown [Hannerz, 1999] that Equation 2.3 agrees well with the law of Petrioli (1981) and that Equation 2.4 agrees well with the law of Heligman and Pollard (1980) in describing one-year death probabilities for integer x-values among Swedish females and males, respectively. However, Equation 2.3 and 2.4 are the only existing laws that also accurately describe the age pattern of mortality during the first year of life.

2.2. On relational methods

There are two dominant lines of proposed parametric relations between mortality schedules in different populations or time periods. One of the lines defines linear relationships between $log(\mu(x))$ values while the other defines linear relationships between $logit(F(x))$ values, where $\mu(x)$ is the hazard rate at age x and $logit(F(x)) = log\left(F(x)(1-F(x))^{-1}\right)$. A quick rundown of some significant developments in the respective lines is given below.

2.2.1 Models based on $log(\mu(x))$

Kermack et al. (1934) used the following relation to model mortality in England and Wales 1845 – 1925 and in Sweden 1755 – 1925 as a function of age and calendar year:

$$log(\mu_{x,t}) = \gamma_{t-x} + log(\mu_{x,s}) \,, \qquad (2.5)$$

where $\mu_{x,t}$ is the hazard rate at age x in calendar year t, γ_{t-x} is a birth year specific parameter and $\mu_{x,s}$ is the hazard rate at age x in a standard cohort life table.

Lee and Carter (1992) used the following relation to forecast mortality in the United States from 1990 to 2065:

$$log(\mu_{x,t}) = a_x + b_x k_t \,, \qquad (2.6)$$

where $\mu_{x,t}$ is the hazard rate at age x in calendar year t, a_x is the average of the $log(\mu_x)$ estimates among the period life tables that are included in the estimation of the parameters, k_t is a time varying index of the level of mortality, and $b_x k_t$ is the expected age-specific difference between $log(\mu_{x,t})$

and a_x. In order to obtain unique parameter estimates, the following constraints were imposed:

$$\sum_x b_x = 1,\ \sum_t k_t = 0.$$

The parameters were fitted to period life tables from 1900 – 1989. The forecasts were obtained by modelling the time varying index k_t as a random walk with a drift.

The Lee-Carter method was extended by Renshaw and Haberman (2006), through the following equation:

$$log(\mu_{x,t}) = a_x + b_x k_t + \gamma_{t-x}\ , \qquad (2.7)$$

where a_x and k_t are defined as above while γ_{t-x} is a birth year specific parameter aimed at controlling potential cohort effects. In order to obtain unique parameter estimates, the following constraints were imposed:

$$\sum_x b_x = 1,\ \sum_t k_t = 0\ ,\ \sum_{t-x} \gamma_{t-x} = 0.$$

Another interesting relation was proposed by Plat (2009). It is defined by the following equation:

$$log(\mu_{x,t}) = a_x + k_{1_t} + k_{2_t}(\bar{x} - x) + k_{3_t}(\bar{x} - x)^+ + \gamma_{t-x}\ , (2.8)$$

where a_x and γ_{t-x} are defined as above, $(\bar{x} - x)^+ = max(0, \bar{x} - x)$ and k_{1_t}, k_{2_t} and k_{3_t} are parameters aimed at modelling the level and age pattern of mortality in calendar year *t*. To obtain unique parameter estimates, the following constraints were imposed:

$$\sum_{t-x} \gamma_{t-x} = 0\ ,\ \sum_{t-x}(t - x)\gamma_{t-x} = 0\ ,\ \sum_t k_{3_t} = 0.$$

2.2.2. Models based on $logit(F(x))$

The usefulness of the logit transformation was demonstrated by Brass (1971), who showed that a life table for an arbitrarily chosen population *i* often can be reasonably well approximated by that in a selected standard population *j* by use of the relation

$$logit\left(F_i(x)\right) = \alpha_i + \beta_i logit\left(F_j(x)\right), \qquad (2.9)$$

where the parameter α_i modifies the level of mortality while β_i modifies the relationship between childhood and old age mortality.

Equation 2.9 has been extensively used to estimate death rates and life expectancies in populations with limited or defective mortality data [cf. Brass et al., 1968; Udjo, 2014]. The relation is, moreover, used to forecast mortality in populations where life tables exist for different periods (or birth cohorts). Such forecasts are usually obtained by estimating a series of time specific parameters α_t and β_t and then extrapolating these series into the future by use of suitably selected ARIMA processes [cf. Edinger, 2011; Tran, 2019].

In an examination of mortality patterns in countries with a long series of life tables, Brass noted that α_t tended to decrease steadily with time, often in a linear fashion, and that β_t, which appeared to be independent of α_t, had a tendency to fluctuate around a central value that was close to 1.0 [Brass, 1974]. By setting β to 1.0, we obtain the one-parameter relation:

$$logit\left(F_i(x)\right) = \alpha_i + logit\left(F_j(x)\right). \qquad (2.10)$$

Hartmann and Strandell (2006) conducted a study aimed at exploring the possibility to use Equation 2.10 as a survival model in stochastic population projections. The model was tested on data from annual life tables of Sweden in the period 1921 – 2002 and Denmark in the period 1922 – 1951. It was concluded that the one-parameter relation given by Equation 2.10 could serve as a suitable survival model for use with stochastic population projections.

Based on a factor analysis of life tables in various countries and time periods, Bourgois-Pichat (1962) proposed that the variation among the studied life tables could be explained by five factors, which he interpreted as (i) a factor governing the level of mortality; (ii) a factor governing the relationship between mortality in youth and adult life; and (iii) a factor governing mortality patterns at extreme ages (especially old age, 70+); (iv) a factor governing infant mortality; and (v) a factor associated with violence. He noted that the fifth factor mainly affected males, and recommended that any effects that this fifth factor might have on females should be deliberately ignored. To accommodate four-dimensional differences between life tables, extended versions of the logit(F) relation have been developed, by Zaba (1979), by Hannerz (2001) [Paper III] and by Murray et al. (2003). The contributions by Zaba and Murray require a model life table system which in addition to $logit(F(x))$ values uses two extra parameters for each age and sex category, and are therefore, in my opinion, not very practical.

My own contribution, which I address in paper III, is given by the equation

$$logit(F_i(x)) = logit\left(F_j(x)\right) + \theta_0 - \theta_1 x^{-1} + \frac{\theta_2}{2}x^2 + \frac{\theta_3}{c}e^{cx}, \quad (2.11)$$

where "the θ_k's are parameters that denote mortality differences and c is a constant held in common by the two mortality schedules, i and j. If the parameter θ_1 in the above expression is different from zero then the two mortality schedules would differ with regard to infant mortality; if $\theta_2 \neq 0$ then the schedules would differ in the relationship between mortality in youth and adult life; and if $\theta_3 \neq 0$ then the schedules would differ in the mortality pattern in old age. The parameter θ_0 would finally determine the difference in the level of mortality" [Paper III]. The simplest relation is obtained when all parameters except θ_0 are zero.

The relation given in Equation 2.11 is primarily intended to be used in conjunction with the mortality law given in Equation 2.1 or 2.3. The potential usefulness of Equation 2.11 is especially high in situations where it is relevant and possible to graduate mortality in several populations simultaneously (which I demonstrate in Paper III). Let us say, for example, that we want to construct county-specific life tables in a nation with 20 counties, and that the level of mortality differs between counties while the shape (age pattern) of $logit(F(x))$ is independent of county. Then, by use of the mortality law given in Equation 2.3 in conjunction with the relational method given by Equation 2.11 with $\theta_1 = \theta_2 = \theta_3 = 0$, we would only need to estimate a total of 25 parameters to obtain 20 complete county-specific life tables, which is quite awesome.

A similar relational method, intended to be used in conjunction with Equation 2.2 or 2.4, is presented in Paper III.

2.3. A final note

Finally, it should be noted that a lot of water has passed under the bridge since the development of Equation 2.3 – 2.4. With time, the ever-increasing performance of computers has facilitated the use of computer-intensive data generated models as an alternative or complement to traditional parametric models in the estimation and forecasting of mortality. It has, for example been shown that support vector machine regression may serve as an alternative to parametric mortality laws in the graduation of the age pattern of mortality [Kostaki et al., 2011; 2016]. It has, moreover, been shown that tree-based statistical learning techniques as well as artificial neural networks may serve as tools to improve the parameter estimates and ARIMA processes of the Lee-Carter model [cf. Levantesi and Pizzorusso, 2019; Nigri et al., 2019; Richman and Wüthrich, 2018; Hong et al., 2021]. Further research is needed to investigate possible pros and cons of using the mortality laws instead of computer-intensive data generated models.

3. On the measurement of risk inequalities

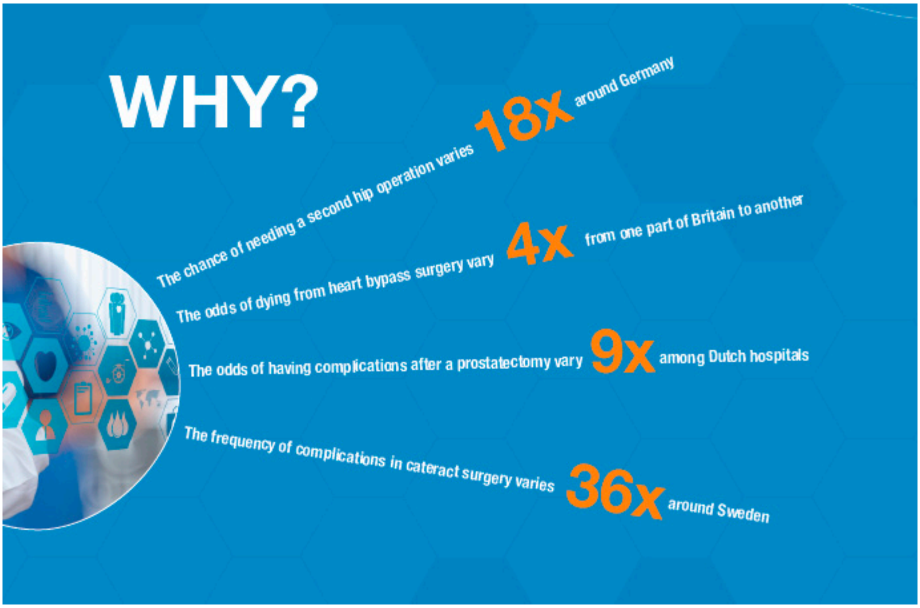


Figure 3.1. A snapshot from the front cover of "Finnegan G, Moran N, Kelly É, Hudson RL. Healthy measures. Brussels: Science Business Publishing Ltd, 2018."

Since the 1980s there has been an increasing emphasis on reduction of health inequalities, in international as well as national health policies. In 1985, the European office of the World Health Organisation (WHO) published 38 health targets [WHO, 1985], of which the first one read, "*By the year 2000, the actual differences in health status between countries and between groups within countries should be reduced by at least 25 percent, by improving the health of disadvantaged nations and groups.*". Towards the end of the 1990s, new targets were set to combat health issues in the 21st century [WHO, 1999]. The target for reduction of inequalities now read, "*By the year 2020, the health gap between socioeconomic groups within countries should be reduced by at least one fourth in all member states, by substantially improving the level of health of disadvantaged groups.*"

Crombie et al. (2005) reviewed public health policies in 13 developed countries (Australia, Canada, Denmark, England, Finland, Ireland, New Zealand, Northern Ireland, Norway, Scotland, Sweden, the United States and Wales), and found that the reduction of health inequalities was an overarching aim of all public health policies. The wording of the national objectives differed, however, among countries as well as among time periods within countries. For example, in the United States the objective towards the year 2000 was to "*reduce disparities in health status among different populations*" [US Department of Health and Human Services, 1990] while the objective towards the year 2010 was to "*eliminate health disparities between different segments of the population, including those relating to gender, race and ethnicity, education, income, disability, living in rural localities, and sexual orientation*" [cf. Davis, 2000]. Another example: The objective towards the year 2000 in Finland was to "*reduce health disparities between population groups, i.e. smaller health differences between genders, socioeconomic categories and people living in different regions*" while the objective towards the year 2015 was "*to reduce mortality differences between the genders, groups with different educational backgrounds, and different vocational groupings by a fifth*" [Ministry of Social Affairs and Health, Finland, 2001].

The widespread interest in health inequalities is, moreover, evident from a recent bibliometric analysis [Cash-Gibson et al., 2018], which found that authors from 159 countries had contributed to research papers on health inequalities in the time period 1965 – 2015, according to the Scopus database.

There is, moreover, an abundance of literature on inequalities in the quality of health care [e.g. Wright et al., 2006; de Vos et ai., 2009; Cooper et al., 2011; Lovaglio et al., 2012; Finnegan et al., 2018]

To determine to what extent health inequalities of a certain type have increased or decreased with time, the inequalities must be quantified and preferably summarised into a single inequality index. Such inequality indices are usually called summary measures of health inequality and are defined as follows: "Summary measures of health inequality draw from disaggregated data in two or more subgroups to yield a single number that represents inequality and are useful to make comparisons between health indicators and over time" [Hosseinpoor et al., 2018]

The present chapter deals with summary measures of inequality. Section 3.1 briefly describes some commonly used summary measures for ordinal group variables. In section 3.2 I briefly describe some commonly used summary measures for nominal group variables without a natural reference group. I will, moreover, show that these measures are subject to bias. In section 3.3, I will tell about my own experience of inequality measurements and the problems which prompted the work given in Papers VII – IX. I will, moreover, present a SAS program, which (based on the methods described in Paper VII) can be used to evaluate Poisson distributed risk inequalities in the absence of a natural reference group.

3.1. A brief rundown of some commonly used summary measures of inequality for ordinal group variables

The following summary measures are applied when a population has been partitioned into n groups that are ranked from 1 = most advantaged (e.g. richest or most educated subgroup) to n = most disadvantaged (e.g. poorest or least educated subgroup). For each group, as well as for the whole population we can then calculate an average rate of illness or death, $\bar{r}_1, \dots \bar{r}_{\mathrm{n}}$ and $\bar{r}_{pop}$, respectively.

3.1.1. Simple measures

There are three commonly used measures of inequality; the rate ratio $\frac{\bar{r}_{\mathrm{n}}}{\bar{r}_1}$, the difference in absolute rates $\bar{r}_n - \bar{r}_1$ and the difference in relative rates $\frac{\bar{r}_{\mathrm{n}} - \bar{r}_1}{\bar{r}_1}$ between the most disadvantaged and most advantaged subgroup [cf. WHO, 2018].

Another commonly used summary measure is the excess fraction (EF), defined as

$$EF = 1 - \frac{\bar{r}_1}{\bar{r}_{pop}} \qquad (3.1)$$

where $\bar{r}_1$ is the mean rate of illness in the most advantaged group while $\bar{r}_{pop}$ is the mean rate of illness in the total population [cf. Paper VII]. The excess fraction is also known as the population attributable fraction (PAF) [cf. Greenland and Robins, 1988; Zapata-Diomedi et al., 2018; WHO, 2018] and can be interpreted as "the proportion of the cases of ill-health that would not have occurred if the rate of illness in the total population had been the same as in the most advantaged sub-group".

A typical example of the usage of EF as a summary measure of risk inequalities is given in Paper X, where inequalities in the risk of developing diabetes among five socio-occupational groups were estimated, with professionals being pre-specified as the most advantaged group. Another example is given in Paper XI, where a simple excess fraction was used to estimate pain-related inequalities in the incidence of depression among employees in Denmark. Here, the pre-specified most advantaged group consisted of employees without musculoskeletal pain as baseline.

3.1.2. Regression based measures

Let x_j be the ridit score [Bross, 1958] also known as the relative rank [cf. Wagstaff et al., 1991; Harper et al. 2008; Harper et al 2010; Hosseinpoor et al., 2016; Skaftun et al., 2018] for the group with rank order j, defined as

$$x_j = \sum_{i=1}^{j} p_i - \frac{p_j}{2}, \tag{3.2}$$

where p_i is the proportion of the population that belongs to group i.

Let $f(x)$ be the rate of illness as a function of the ridit score and assume that this function has been determined through regression analysis. Then the following summary measures of inequality may be calculated:

The slope index of inequality [cf. Mackenbach and Kunst, 1997]

$$SII = f(1) - f(0), \tag{3.3}$$

the relative index of inequality [cf. Hosseinpoor et al, 2016; Pamuk, 1985; Kakwani et al., 1997; Wagstaff et al., 1991]

$$RII = \frac{f(1) - f(0)}{\bar{r}_{pop}}, \tag{3.4}$$

the Kunst-Mackenbach index [cf. Hosseinpoor et al, 2016]

$$KMI = \frac{f(1)}{f(0)}, \tag{3.5}$$

the regression based excess fraction

$$EF = \frac{\sum_{i=1}^{n} p_i\big(f(x_i) - f(0)\big)}{f(0) + \sum_{i=1}^{n} p_i\big(f(x_i) - f(0)\big)}. \tag{3.6}$$

It should be noted that the Kunst-Mackenbach index (KMI) has also been referred to as the relative inequality index (RII) [cf. Mackenbach and Kunst, 1997; Sergeant and Firth, 2006; WHO, 2018; Hoebel et al., 2018]. It should also be noted that the above equations hold when the groups are ordered from the most advantaged to the most disadvantaged. If we were to reverse the ordering so that 1 = the most disadvantaged group and n = most advantaged group [cf. Sergeant and Firth, 2006; WHO, 2018], then $f(1)$ and $f(0)$ would need to shift places in Eq. 3.3 – 3.6, and $\bar{r}_1$ would need to be replaced by $\bar{r}_n$ in Eq. 3.1.

3.1.3. The relative concentration index

The relative concentration index (RCI) [cf. Kakwani et al., 1997; Harper et al. 2008; Harper et al. 2010; Hosseinpoor et al., 2016] is another summary measure that relies on ridit scores. It is defined as

$$RCI = \frac{2}{\bar{r}_{pop}} \sum_{i=1}^{n} p_i\, x_i \bar{r}_i - 1, \tag{3.7}$$

where p_i, x_i, $\bar{r}_i$ and $\bar{r}_{pop}$ are defined as above.

The RCI can take on values in the interval -1 to 1. The expected value under a null-hypothesis of no inequality is zero. The RCI will be positive if the rate of illness increases with the ridit score and negative if the rate decreases with the ridit score.

3.2. A brief rundown of some commonly used summary measures of inequality for nominal group variables without a natural reference group

3.2.1. Simple measures

According to WHO (2018), when a dimension of inequality does not have a natural ordering, the simple measures that are described in section 3.1.1 may be calculated by use of the subgroups with the highest and lowest observed rates, instead of the most disadvantaged and most advantaged subgroup.

If the estimated group averages are free from random errors and systematic bias, i.e. if the observed rates in each of the studied subgroups are exactly the same as the true unobservable expected risk rates, then the difference or ratio between the subgroups with the highest versus lowest observed rates would yield an unbiased measure of the difference or ratio between the true unobservable expected risk rates in the subgroup with the highest versus lowest risk. The excess fraction of (Eq. 3.1), with $\bar{r}_1$ replaced by the mean rate of illness in the group lowest observed rate, could be interpreted as "the proportion of the cases of ill-health that would not have occurred if the rate of illness in the total population had been as low as it was in the sub-group with the lowest risk".

If the group averages, on the other hand, are outcomes of a random process, then the expected value of such comparisons will reflect the combined effect of disparities in risk and disparities that are expected to occur by chance. From theorems on order statistics [cf. Larsen and Marx, 1986], it follows that differences between the highest and lowest estimated risk may be large even if the true risk is exactly the same in all of the included subgroups.

3.2.2. The index of dissimilarity

The index of dissimilarity (ID) is another measure that has been recommended for use in situations where no natural ordering and no natural reference group exist [cf. Mackenbach, 1993]. It is defined by the equation:

$$ID = \frac{1}{2}\sum_{i=1}^{n} p_i \left|\frac{\bar{r}_i}{\bar{r}_{pop}} - 1\right|, \qquad (3.8)$$

where p_i is the proportion of the population that belong to group i, $\bar{r}_i$ is the mean rate of illness in group i and $\bar{r}_{pop}$ is the mean rate of illness in the total population.

From Equation 3.8, it follows that ID is independent of the ordering of the groups. If the estimated group averages are free from random errors and systematic bias, then ID can be interpreted as the proportion of the expected cases of the concerned illness that need to be redistributed between the subgroups in order to make the expected rate in each group equal to that in the total population [cf. Mackenbach, 1993].

If the estimated group averages are free from random errors and systematic bias, then ID equals zero under the null-hypothesis of no inequalities in risk. We note, however, that if X is a normal random variable with mean zero and standard deviation σ then the expected absolute value of X equals $\sigma\sqrt{2/\pi}$ [cf. Leone et al., 1961]. Hence, it follows from the central limit theorem [Larsen and Marx, 1986] that, whenever the observed risk rates are subject to random variation, the expected value of ID will be larger than zero, even if the true risk is exactly the same in all of the included subgroups.

3.2.3. The Gini index

The Gini index (GI), also known as the Gini coefficient, was developed in the beginning of the 20th century by the Italian Statistician Corrado Gini as a measure of income inequality [cf. Giorgi and Gigliarano, 2017]. The index has also been used to measure geographic inequalities in mortality [cf. Skaftun et al., 2018]. GI requires the sub-groups to be ranked in ascending order on the basis of the observed rates but is otherwise defined in the same way as the relative concentration index (Equation 3.7). I.e.

$$GI = \frac{2}{\bar{r}_{pop}} \sum_{i=1}^{n} p_i\, x_i \bar{r}_i - 1 \qquad (3.9)$$

where p_i, x_i, $\bar{r}_i$ and $\bar{r}_{pop}$ are defined as above.

If we let $SRR_i = \bar{r}_i / \bar{r}_{pop}$ be the standardised rate ratio of group *i,* then GI may be interpreted as half the average absolute difference between any two standardised rate ratios [cf. Skaftun et al., 2018]. If the estimated group averages are free from random errors, then GI will be a measure of risk inequalities. If the group averages, on the other hand, are outcomes of a random process, then it follows from the arguments that were given in the last paragraph of section 3.2.2 that the expected value of GI will reflect the combined effect of disparities in risk and disparities that are expected to occur by chance.

3.2.4. Demonstration of bias from random variation in traditional summary measures of inequality for nominal group variables without a natural reference group

To summarise, the inequality measures listed in sections 3.2.1 – 3.2.3 are biased whenever the observed risk rates are subject to random variation, and the size of the bias will, all else being equal, depend on the standard errors of the estimated group averages. It follows, however, from the law of large numbers [Grimmet and Stirzaker, 1992] that the measures are asymptotically unbiased, since the standard errors of the group averages will approach zero when the number of observations within each group approaches infinity.

To give an idea of the importance of such bias, I have used Monte Carlo simulation to compare expected values of some summary measures of inequalities in observed risk rates with the values that would have been obtained in the absence of random errors. The simulations are based on a hypothetical population that has been partitioned into 50 equally sized subgroups. The rate of illness in the respective groups is governed by a Poisson distribution, and the true risk rates in the groups increase geometrically with their rank order. The expected value of i) the ratio between the highest and lowest observed risk, (ii) the observed excess fraction, (iii) the observed index of dissimilarity and (iv) the observed Gini-index are estimated as a function of the expected number of cases in the total population (500; 1000; 5000; 50,000; 500,000; 5,000,000; infinite) and the ratio between the highest and lowest true risk (1.0; 1.5; 3.0).

The simulations were performed by use of the software SAS (version 9.4). The program that was used is given in Appendix 1. The results of the simulations are given in Table 3.1.

The table indicates that the examined summary measures will provide satisfactory approximations when there are 1000 or more expected cases per group (50,000 or more in the total population), especially if the true risk inequalities are large. However, the table indicates that the summary measures are severely biased when they are applied to data sets with 100 or less expected cases per group, especially if the true risk inequalities are small.

Table 3.1 Illustration of bias from random variation in data generated naive estimates of (i) the risk ratio between the group with the highest vs. lowest risk, (ii) the excess fraction (EF), (iii) the index of dissimilarity (ID) and (iv) the Gini-index (GI), among fifty equally large groups.

Highest / lowest true risk	Total number of cases	Expected value of Highest / lowest observed risk	Expected value of Observed EF	Expected value of Observed ID	Expected value of Observed GI
1.0	500	5.39	0.631	0.125	0.174
	1000	2.95	0.464	0.088	0.123
	5000	1.59	0.225	0.040	0.055
	50,000	1.15	0.071	0.012	0.017
	500,000	1.05	0.022	0.004	0.006
	5,000,000	1.01	0.007	0.001	0.002
	Infinite	1.00	0.000	0.000	0.000
1.5	500	6.15	0.659	0.134	0.186
	1000	3.43	0.510	0.101	0.140
	5000	1.97	0.311	0.063	0.087
	50,000	1.58	0.213	0.053	0.071
	500,000	1.51	0.192	0.052	0.069
	5,000,000	1.50	0.189	0.052	0.069
	Infinite	1.50	0.189	0.052	0.069
3.0	500	11.05	0.773	0.181	0.249
	1000	6.28	0.664	0.160	0.218
	5000	3.68	0.524	0.142	0.190
	50,000	3.09	0.463	0.138	0.184
	500,000	3.00	0.452	0.138	0.183
	5,000,000	3.00	0.452	0.138	0.183
	Infinite	3.00	0.452	0.138	0.183

3.3. Presentation of a Monte Carlo simulation procedure to estimate excess fractions

3.3.1. Background

I carried out several statistical studies on behalf of the Swedish National Board of Health and Welfare to investigate the possibility of using national patient registers to monitor and compare the quality of treatment in public emergency hospitals. Of particular interest in the present dissertation are two studies which focus on age and gender standardised case fatality rates among patients admitted for acute myocardial infarction. The studies were carried out at more than 90 hospitals, the first for the time period 1987 – 1991 [Hannerz, 1996a] and the second for the time period 1992 – 1994 [Hannerz, 1996b]. The findings revealed that inequalities between hospitals in case fatality rates were significantly larger than what could be expected to occur by chance alone.

The inequalities were tested for statistical significance by use of the classic Pearson statistic

$$\chi^2 = \sum_i \frac{(x_i - m_i)^2}{m_i}, \qquad (3.10)$$

where the $x_i's$ are the hospital specific observed numbers of deaths and the $m_i's$ are the age and gender standardised expected numbers (under the null hypothesis of no inequalities between hospitals).

An important quantity to study in these type of comparisons is the excess fraction, which in the above example would be defined as the proportion of the deaths that would not have occurred if the risks in all hospitals had been as low as they were in the hospital with the lowest risk. The main difficulty lay in obtaining an unbiased estimate of the excess fraction because no natural reference group exists. Clearly, using the lowest observed rate as a proxy for the lowest true risk rate would have yielded an upwards biased estimate of the excess fraction and could thereby exaggerate the prevention potential [cf. section 3.2].

In a series of papers [Paper VII, Paper VIII, Paper IX] I solved this problem by using Monte Carlo simulation to estimate the expected value of Pearson's chi-square as a function of the true excess fraction in a hypothetical population. I let the hypothetical population be partitioned in the same way, and with the same distribution and numbers of expected cases as the observed population (under the null hypothesis of no inequalities in risk). Then using the results of the simulations, I was able to determine the value of the excess fraction that corresponded to the Pearson chi-square of the observed population. A detailed description of the procedure was published in 2006 [Paper VII].

In Paper VIII the estimated excess fractions were used as summary measures of industrial inequalities in disability retirement rates. The simulation procedure was based on the following assumptions:

- The number of events in each industry follows a Poisson distribution.
- The industries' true risk rates increase geometrically with their rank order.

The same procedure was subsequently used to estimate industrial inequalities in mood disorders [Paper IX] and accidental injuries [Kines et al., 2007; Pedersen et al., 2010].

3.3.2. The Monte Carlo simulation procedure

In the present section, I will demonstrate the use of the simulation procedure described in Paper VII but without going into mathematical and technical details.

The demonstration will revolve around the data given in Table 3.3, which lists industry specific standardised morbidity ratios (SMR) for ischaemic heart disease 1994 – 1999 among people in Denmark who were 20-59 years old and economically active on January 1st, 1994. The data were obtained through calculations based on a record linkage between the Central Person Register (with information on gender and dates of births, deaths and migration) [cf. Pedersen, 2011], the Employment Classification Module (with information on employment status, industry and socioeconomic status (SES)) [cf. Petersson et al., 2011] and the National Patient Register (with information on dates and diagnoses produced during hospital contacts) [cf. Lynge et al., 2011]. The industries are classified into 57 groups in accordance with the classification AT49X [Appendix 4], which is an aggregation of the Danish Industrial Classification of All Economic Activities 1993 [Statistics Denmark, 1996].

Table 3.3 lists the observed numbers of cases, the expected numbers (under the null hypothesis of no inequalities in risk) and the SMRs which equal the observed divided by the expected numbers for each of the 57 industries. The expected numbers are standardised, firstly for gender and age (5-year classes) and secondly, for gender, age and SES (legislators, senior officials, and managers; professionals; technicians and associate professionals; workers in occupations that require skills at a basic level; workers in elementary occupations; gainfully employed people with an unknown occupation) [cf. Statistics Denmark, 1997].

To estimate the industry-related excess fraction of ischaemic heart disease, I used the SAS program given in Appendix 2

The input data consist of the observed and the age and gender standardised expected number of observations [Table 3.3]. The program thereafter calculates the observed Pearson chi-square sum, performs the simulations, and outputs the data given in Table 3.2, thus indicating that the observed chi-square sum is approximately equal to the expected chi-square sum at an excess fraction of 0.26.

Table 3.2 Output from the SAS-program given in Appendix 2

Excess fraction	Expected chi-square sum	Observed chi-square sum
0.20	473	803
0.21	520	803
0.22	569	803
0.23	621	803
0.24	676	803
0.25	734	803
0.26	796	803
0.27	862	803
0.28	929	803
0.29	1002	803
0.30	1076	803
0.31	1156	803
0.32	1238	803
0.33	1323	803
0.34	1413	803
0.35	1507	803
0.36	1607	803
0.37	1710	803
0.38	1819	803
0.39	1932	803
0.40	2048	803

Next, I estimated a 95% confidence interval of the simulated chi-square sums at the excess fraction of 0.26 using the SAS-program in Appendix 3. The upper and lower limits of the confidence interval of the chi-square sum are estimated to be 517 and 1182, respectively. Now, looking back at Table 3.2, I note that (i) a chi-square sum of 517 corresponds to an excess fraction of approximately 0.21 and (ii) a chi-square sum of 1182 corresponds to an excess fraction of approximately 0.31. Hence, the upper and lower limits of a 95% confidence interval of the excess fraction is given by 0.21 and 0.31, respectively.

After having established an estimate and confidence interval for the age and gender standardised excess fraction, the same steps were carried out, instead using the age, gender and SES standardised expected numbers. Here I estimated the age, gender and SES standardised excess fraction at 0.20 (95% CI: 0.16 – 0.24) and thereby conclude that approximately 23% (100 x (0.26 – 0.20) / 0.26) of the age and gender standardised industrial inequalities in risk of ischaemic heart disease could be attributed to differences in the industries' socioeconomic compositions.

Table 3.3. Standardised morbidity ratio (SMR) with 95 % confidence interval (CI) for ischaemic heart disease 1994 – 1999 among people who were economically active in Denmark 1st January 1994

			Standardised for age and gender			Standardised for age, gender and socio-economic status		
Industrial group	Persons	Observed No. of cases	Expected No. of cases	SMR	95% CI	Expected No. of cases	SMR	95% CI
010 Metal and steelworks, and foundries	5819	72	81.5	0.88	0.70 - 1.11	88.0	0.82	0.65 - 1.03
020 Manufacture of transport equipment	11 399	195	154.6	1.26	1.10 - 1.45	158.6	1.23	1.07 - 1.41
030 Shipyards	11 754	213	187.6	1.14	0.99 - 1.30	196.4	1.08	0.95 - 1.24
040 Electricity and heat supply	13 443	264	262.7	1.01	0.89 - 1.13	268.0	0.99	0.87 - 1.11
050 Iron and metal industries	44 260	610	566.4	1.08	0.99 - 1.17	595.2	1.02	0.95 - 1.11
060 Engineering industry	64 605	845	816.7	1.03	0.97 - 1.11	831.0	1.02	0.95 - 1.09
070 Electricity and electronics industry	30 833	315	313.7	1.00	0.90 - 1.12	318.2	0.99	0.89 - 1.11
080 Car industry	22 578	312	298.2	1.05	0.94 - 1.17	291.8	1.07	0.96 - 1.19
090 Navvy and road contractors	48 777	726	738.4	0.98	0.91 - 1.06	777.9	0.93	0.87 - 1.00
100 Bricklayer, joiner, and carpentry work	35 147	464	543.2	0.85	0.78 - 0.94	548.9	0.85	0.77 - 0.93
110 Finishing	15 904	233	230.0	1.01	0.89 - 1.15	229.7	1.01	0.89 - 1.15
120 Insulation and installation businesses	35 916	415	426.8	0.97	0.88 - 1.07	426.2	0.97	0.88 - 1.07
130 Printing works and publishing	32 010	396	423.5	0.93	0.85 - 1.03	415.6	0.95	0.86 - 1.05
140 Paper, cardboard and bookbinding industries	9392	127	124.3	1.02	0.86 - 1.22	131.9	0.96	0.81 - 1.15
150 Wholesale trade	136 776	1677	1704.2	0.98	0.94 - 1.03	1675.3	1.00	0.95 - 1.05
160 Transport of goods	106 294	1803	1505.1	1.20	1.14 - 1.25	1629.8	1.11	1.06 - 1.16
170 Transport of passengers	53 120	1180	848.2	1.39	1.31 - 1.47	899.9	1.31	1.24 - 1.39
180 Fire service, lighthouse and salvage corps	8043	144	122.3	1.18	1.00 - 1.39	126.5	1.14	0.97 - 1.34
190 Textile, clothing, and leather industry	24 360	243	212.1	1.15	1.01 - 1.30	238.3	1.02	0.90 - 1.16
200 Manufacture of wood and wood products	33 561	424	399.5	1.06	0.97 - 1.17	426.2	0.99	0.90- 1.09

Industrial group	Persons	Observed No. of cases	Standardised for age and gender			Standardised for age, gender and socio-economic status		
			Expected No. of cases	SMR	95% CI	Expected No. of cases	SMR	95% CI
210 Mineral, oil, rubber and plastic products	22 141	296	272.5	1.09	0.97 - 1.22	288.1	1.03	0.92 - 1.15
220 Stone-works, pottery, and glass industry	13 535	228	204.3	1.12	0.98 - 1.27	221.6	1.03	0.90 - 1.17
230 Medical equipment/toys/cameras/etc.	14 642	144	137.3	1.05	0.89 - 1.23	143.4	1.00	0.85 - 1.18
240 Manufacture of industrial chemicals	14 339	186	187.4	0.99	0.86 - 1.15	192.9	0.96	0.84 - 1.11
250 Heavy raw material and semi-manufacture	7188	148	114.3	1.30	1.1 0- 1.52	122.5	1.21	1.03 - 1.42
260 Pharmaceutical industry	9515	73	93.0	0.78	0.62 - 0.99	92.2	0.79	0.63 - 1.00
271 Office & adm. (transport & wholesale)	15 016	164	150.4	1.09	0.94 - 1.27	150.2	1.09	0.94 - 1.27
272 Office & adm. (service)	19 881	428	376.1	1.14	1.04 - 1.25	381.6	1.12	1.02 - 1.23
273 Finance/ Public office & adm.	190 987	1703	1931.8	0.88	0.84 - 0.92	1779.8	0.96	0.91 - 1.00
274 Private office & adm.	142 013	1368	1534.1	0.89	0.85 - 0.94	1423.4	0.96	0.91 - 1.01
281 Car dealers	17 391	171	189.5	0.90	0.78 - 1.05	188.0	0.91	0.78 - 1.06
282 Garage	5660	61	47.6	1.28	1.10 - 1.65	47.4	1.29	1.00 - 1.65
283 Shops	69 402	613	620.0	0.99	0.91 - 1.07	617.4	0.99	0.92 - 1.07
290 Supermarkets, department stores etc.	59 299	402	406.9	0.99	0.90 - 1.09	404.8	0.99	0.90 - 1.10
300 Sewers, water- and gas supply	6666	155	122.7	1.26	1.08 - 1.48	125.5	1.24	1.06 - 1.45
310 Personal care and other services	19 349	175	179.4	0.98	0.84 - 1.13	182.9	0.96	0.83 - 1.11
320 Cleaning, laundries, and dry cleaners	29 345	264	237.4	1.11	0.99 - 1.25	265.2	1.00	0.88 - 1.12
330 Telecommunication	15 081	210	205.6	1.02	0.89 - 1.17	203.8	1.03	0.90 - 1.18
340 Surveillance, armed forces, police etc.	61 805	731	681.6	1.07	1.00 - 1.15	667.6	1.10	1.02 - 1.18
350 Hotels and restaurants	50 373	456	354.4	1.29	1.17 - 1.41	368.2	1.24	1.13 - 1.36
361 Photographers/film & videoproduction	4206	22	38.6	0.57	0.38 - 0.87	38.1	0.58	0.38 - 0.88
362 Entertainment, culture and sport	25 878	280	295.4	0.95	0.84 - 1.07	291.0	0.96	0.86 - 1.08

Industrial group	Persons	Observed No. of cases	Standardised for age and gender			Standardised for age, gender and socio-economic status		
			Expected No. of cases	SMR	95% CI	Expected No. of cases	SMR	95% CI
363 Libraries and archives	8534	61	91.8	0.66	0.52 - 0.85	86.1	0.71	0.55 - 0.91
370 Slaughterhouse industry	20 300	308	250.5	1.23	1.10 - 1.37	276.8	1.11	1.00 - 1.24
380 Poultry slaughtering and fish products	11 601	142	93.0	1.53	1.30 - 1.80	106.9	1.33	1.13 - 1.57
390 Beverage industry	14 010	222	210.8	1.05	0.92 - 1.20	225.7	0.98	0.86 - 1.12
400 Manufacture of bread, chocolate, tobacco etc.	23 274	269	218.5	1.23	1.09 - 1.39	230.7	1.17	1.03 - 1.31
410 Manufacture of dairy products	11 891	148	138.8	1.07	0.91 - 1.25	151.9	0.97	0.83 - 1.14
420 Agriculture	71 151	869	1267.1	0.69	0.64 - 0.73	1232.6	0.71	0.66 - 0.75
430 Horticulture and forestry	18 739	234	228.5	1.02	0.90 - 1.16	246.3	0.95	0.84 - 1.08
440 Hospitals	98 079	724	736.4	0.98	0.91 - 1.06	680.8	1.06	0.99 - 1.14
450 Nursing homes, home care, etc.	152 660	1420	1102.8	1.29	1.22 - 1.36	1181.5	1.20	1.14 - 1.27
460 Child care etc.	109 998	530	548.4	0.97	0.89 - 1.05	557.2	0.95	0.87 - 1.04
471 General practitioners, dentists etc.	23 278	163	244.9	0.67	0.57 - 0.78	228.2	0.71	0.61 - 0.83
472 Health care, not elsewhere classified	32 144	287	332.2	0.86	0.77 - 0.97	311.0	0.92	0.82 - 1.04
480 Education and research	172 362	1654	2173.9	0.76	0.73 - 0.80	1984.4	0.83	0.79 - 0.87
490 Fishing	6166	117	107.6	1.09	0.91 - 1.30	115.5	1.01	0.85 - 1.21

4. Concluding remarks

The series of mortality laws presented in this dissertation resulted in the finding that the standard error of estimated sex and age-specific one-year death probabilities, were, on average five times smaller than for the life-tables system of Statistics Sweden, 1982 [cf. Hannerz, 1999]

Moreover, the mortality laws provided a remarkable goodness-of-fit compared with many other life table systems when tested on mortality in different calendar years of the population of Sweden [Hannerz, 1999]. In Paper II, it was shown that the laws are able to accommodate many different shapes and levels of mortality as a function of age, which suggests that they would also fit well with respect to all-cause mortality in many other populations. The drawback of the mortality laws is their non-linear parametrisation, which makes them difficult to work with. It should also be noted that the laws ordinarily provide more accuracy than needed in most public health research endeavours.

A Monte Carlo simulation procedure was designed to offset bias from random variation in the estimation of excess fractions in the absence of a natural reference group [cf. Paper VII]. The procedure was designed to estimate inequalities under the assumption that the true risk rates increase geometrically with their rank order and the outcome variables are Poisson distributed. It is clear that a similar procedure could be constructed for outcome variables with other distributions and also for other assumptions of the true risk as a function of the true underlying rank order.

It can be objected that the assumptions about the relation between the true risks and the true rank orders cannot be formally tested, since the true rank order is unknown and unobservable. The procedure is also difficult to explain to people not versed in statistics, such as many of the stakeholders of health inequality evaluations.

Finally, it should be noted that the naïve summary measures given in section 3.2 will work quite well when the expected numbers of events in each group are large. Table 3.1 suggests, for example, that, for most practical purposes, the bias in such naïve summary measures of inequality will be negligible if we are looking at inequalities among 50 groups with at least 1000 expected events in each group.

References

Anson J. (1988). “The parameters of death: a consideration of the quantity of information in a life table using a polynomial representation of the survivorship curve.” Statistics in Medicine, 7:895-912.

Booth H, Tickle L. (2008). Mortality Modelling and Forecasting: A Review of Methods. Annals of Actuarial Science, 3(1-2), 3-43.

Boulougari A, Lundengård K, Rančić M, Silvestrov S, Suleiman S, Strass B. (2019). Application of a power-exponential function-based model to mortality rates forecasting, Communications in Statistics: Case Studies, Data Analysis and Applications, 5:1, 3-10, DOI: 10.1080/23737484.2019.1578705.

Bourgois-Pichat J. (1962). Factor analysis of sex-age specific death rates, Population Bulletin of the United Nations, No. 6.

Brass W. (1968). The Demography of Tropical Africa. Princeton: Princeton University Press.

Brass W. (1969). A generation method for projecting death rates. In: Bechhofer F. editor, Population growth and the brain drain. Techniques and Methods of study. Birmingham: Edinburgh University Press: 75-91.

Brass W. (1971). On the scale of mortality. In: Brass W, editor. Biological aspects of mortality. Symposia of the society for the study of human biology. Volume X. London: Taylor & Francis Ltd.: 69-110.

Brass W. (1974). “Mortality models and their uses in demography.” Transactions of the Faculty of Actuaries, 33:122-133.

Brillinger DR. (1961). “A justification of some common laws of mortality.” Trans Soc Actuaries, 13:116-119.

Bross IDJ. How to use ridit analysis. Biometrics 1958;14:18-38.

Cash-Gibson L, Rojas-Gualdrón DF, Pericàs JM, Benach J. Inequalities in global health inequalities research: A 50-year bibliometric analysis (1966-2015). PLoS One. 2018 Jan 31;13(1):e0191901.

Cooper Z, Gibbons S, Jones S, McGuire A. Does Hospital Competition Save Lives? Evidence from the English NHS Patient Choice Reforms. Econ J (London). 2011 Aug;121(554):F228-F260.

Crombie IK, Irvine L, Elliott L, Wallace H. Closing the Health Inequalities Gap: An International Perspective (WHO Regional Office for Europe) [Internet]. 2005. Available: http://www.euro.who.int/__data/assets/pdf_file/0005/124529/E87934.pdf. Accessed July 28, 2019

Davis RM. Healthy People 2010: objectives for the United States. Impressive, but unwieldy. BMJ. 2000 Mar 25;320(7238):818-9.]

Derrick VPA (1927): "Observation on (1) Errors of Age in the Population Statistics of England and Wales, and (2) the changes in Mortality indicated by the national records." Journal of the Institute of Actuaries, 58:117-146.

de Vos M, Graafmans W, Kooistra M, Meijboom B, Van Der Voort P, Westert G. Using quality indicators to improve hospital care: a review of the literature. Int J Qual Health Care. 2009 Apr;21(2):119-29.

Edinger BJ. (2011). A Stochastic Population Projection for Canada's First Nations. Kingston, Ontario: Department of Economics. Queen's University, Kingston, Ontario, Canada.

Ewbank DC, Gomez de Leon JC, Stoto MA. (1983). "A reducible four-parameter system of model life tables. "Population Studies, 37:105-127.

Finnegan G, Moran N, Kelly É, Hudson RL. Healthy measures. Brussels: Science Business Publishing Ltd, 2018.

Giorgi GM, Gigliarano C. The Gini concentration index: a review of the inference literature. Journal of Economic Surveys . 2017; 31:1130–1148.

Gompertz B. (1825). "On the nature of the function expressive of the law of human mortality." Philosophical Transactions, 27:513-519.

Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. Am J Epidemiol 1988;128: 1185–97.

Grimmet GR, Stirzaker DR. (1992). Probability and Random Processes. Oxford: Claredon Press.

Hannerz H. Korttidsdödlighet bland hjärtinfarktpatienter vid svenska akutsjukhus 1987–91. Stockholm: Socialstyrelsen; 1996a.

Hannerz H. Korttidsdödlighet bland hjärtinfarktpatienter vid svenska akutsjukhus 1992–94. Stockholm: Socialstyrelsen; 1996b.

Hannerz H. (1999). Methodology and applications of a new law of mortality. Lund: Department of Statistics. University of Lund, Sweden.

Harper S, Lynch J, Meersman SC, Breen N, Davis WW, Reichman ME. An overview of methods for monitoring social disparities in cancer with an example using trends in lung cancer incidence by area- socioeconomic position and race-ethnicity, 1992-2004. Am J Epidemiol. 2008 Apr 15;167(8):889-99.

Harper S, King NB, Meersman SC, Reichman ME, Breen N, Lynch J. Implicit value judgments in the measurement of health inequalities. Milbank Q. 2010 Mar;88(1):4-29.

Hartmann M. (1987). “Past and recent attempts to model mortality at all ages.” Journal of official Statistics, 3:19-36.

Hartmann M, Strandell G. (2006). Stochastic Population Projections for Sweden. Research and Development – Methodology reports from Statistics Sweden 2006:2

Heligman M, Pollard JH. (1980). “The age pattern of mortality.” Journal of the Institute of Actuaries 107: 49-80.

Hoebel J, Kroll LE, Fiebig J, Lampert T, Katalinic A, Barnes B, Kraywinkel K. Socioeconomic Inequalities in Total and Site-Specific Cancer Incidence in Germany: A Population-Based Registry Study. Front Oncol. 2018 Sep 25;8:402. doi: 10.3389/fonc.2018.00402. eCollection 2018.

Hong WH, Yap JH, Selvachandran G. Thong PH, Son LH. (2021). Forecasting mortality rates using hybrid Lee–Carter model, artificial neural network and random forest. Complex Intell. Syst. 7, 163–189.

Hosseinpoor AR, Bergen N, Schlotheuber A, Grove J. Measuring health inequalities in the context of sustainable development goals. Bull World Health Organ. 2018 Sep 1;96(9):654-659.

Hosseinpoor AR, Nambiar D, Schlotheuber A, Reidpath D, Ross Z. Health Equity Assessment Toolkit (HEAT): software for exploring and comparing health inequalities in countries. BMC Med Res Methodol. 2016 Oct 19;16(1):141.

Kakwani N, Wagstaff A, van Doorslaer E. Socioeconomic inequalities in health: measurement, computation, and statistical inference. Journal of Econometrics. 1997;77: 87-103.

Kermack WO, McKendrick AG, McKinlay PL. Death-rates in Great Britain and Sweden: Expression of Specific Mortality Rates as Products of Two Factors, and some Consequences thereof. J Hyg (Lond). 1934 Dec;34(4):433-57.

Kines P, Hannerz H, Mikkelsen KL, Tüchsen F. Industrial sectors with high risk of women's hospital-treated injuries. Am J Ind Med. 2007 Jan;50(1):13-21.

Kostaki A. (1992). Methodology and applications of the Heligman-Pollard formula. Lund: Department of Statistics. University of Lund, Sweden.

Kostaki A, Moguerza JM, Olivares A, Psarakis S. Support Vector Machines as tools for mortality graduation. Canadian Studies in Population 38, No. 3–4 (Fall/Winter 2011):37–58.

Kostaki A, Moguerza JM, Olivares A, Psarakis S. (2016). Nonparametric graduation techniques as a common framework for the description of demographic patterns. International Journal of Population Studies, vol.2(1): 1–20.

Larsen RJ, Marx ML. (1986). An introduction to mathematical statistics and its applications. New Jersey: Prentice-Hall, Englewood Cliffs.

Lee RD, Carter LR. (1992). Modelling and forecasting U.S. mortality. Journal of the American Statistical Association, 87(419), 659-671.

Leone FC, Nelson LS, Nottingham RB. (1961) The Folded Normal Distribution, Technometrics. 1961;3:543-550.

Levantesi S, Pizzorusso V. (2019). Application of Machine Learning to Mortality Modeling and Forecasting. Risks, MDPI, Open Access Journal, vol. 7(1), pages 1-19.

Lovaglio PG. Benchmarking strategies for measuring the quality of healthcare: problems and prospects. Scientific World Journal. 2012;2012:606154. doi:10.1100/2012/606154.

Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. Scand J Public Health 2011 Jul;39(7 Suppl):30-3.

Mackenbach JP. Inequalities in health in The Netherlands according to age, gender, marital status, level of education, degree of urbanization, and region. Eur J Public Health. 1993;3:112-118.

Mackenbach JP, Kunst AE. Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe. Soc Sci Med. 1997 Mar;44(6):757-71.

Makeham WM. (1867). “On the law of mortality.” Journal of the Institute of Actuaries, 13:335-340.

Ministry of Social Affairs and Health, Finland. Government Resolution on the Health 2015 public health programme. Helsinki, 2001 (http://1nj5ms2lli5hdggbe3mm7ms5.wpengine.netdna-cdn.com/files/2010/03/pnsfin.pdf. Accessed 27 July 2019).]

Mode C, Busby R. (1982).” An eight-parameter model of human mortality— the single decrement case.” Bulletin of Mathematical Biology, 44:647-659.

Murray CJL, Ferguson BD, Lopez AD, Guillot M, Salomon JA, Ahmad O. (2003) Modified logit life table system: principles, empirical validation, and application, Population Studies, 57:2, 165-182.

Neves C, Fernandes C, Hoeltgebaum H. (2017). Five different distributions for the Lee–Carter model of mortality forecasting: A comparison using GAS models. Insurance: Mathematics and Economics, 75, 48-57.

Nigri A, Levantesi S, Marino M Scognamiglio S, Perla F. (2019). A Deep Learning Integrated Lee–Carter Model. Risks, MDPI, Open Access Journal, vol. 7(1), pages 1-16.

Pamuk ER. Social class inequality in mortality from 1921 to 1972 in England and Wales. Popul Stud (Camb). 1985 Mar;39(1):17-31.

Pascariu MD. (2018). Modelling and forecasting mortality. Odense: Syddansk Universitet. Det Sundhedsvidenskabelige Fakultet. Research output: Thesis › Ph.D. thesis

Pearson K. (1895). “Contributions to the mathematical theory of evolution. II. Skew variation in homogenous material.” Philosophical Transactions of the Royal Society of London, 186:343-414.

Pedersen BH, Hannerz H, Tüchsen F, Mikkelsen KL, Dyreborg J. Industry and injury related hospital contacts: a follow-up study of injuries among working men in Denmark. J Occup Health. 2010;52(3):147-54.

Pedersen CB. The Danish Civil Registration System. Scand J Public Health 2011 Jul;39(7 Suppl):22-25.

Perks W. (1932). “On some experiments in the graduation of mortality statistics.” Journal of the Institute of Actuaries, 63:12-57.

Petersson F, Baadsgaard M, Thygesen LC. Danish registers on personal labour market affiliation. Scand J Public Health 2011 Jul;39(7 Suppl):95-98.

Petrioli L. (1981). A new set of models of mortality. Paper presented at IUSSP seminar on methodology and data collection, Dakar, Senegal, reprinted by Universita degli Studi di Siena, Italy.

Plat R. (2009). On stochastic mortality modeling. Insurance: Mathematics and Economics 45: 393–404.

Renshaw AE, Haberman S. (2006). A cohort-based extension of the Lee-Carter model for mortality reduction factors. Insurance: Mathematics and Economics, 38, 556-570.

Richman R, Wüthrich MV. (2018). A Neural Network Extension of the Lee-Carter Model to Multiple Populations. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3270877.

Sergeant JC, Firth D. Relative index of inequality: definition, estimation, and inference. Biostatistics. 2006 Apr;7(2):213-24.

Siler W. (1983). “Parameters of mortality in human populations with widely varying life spans.” Statistics in Medicine, 2:373-380.

Skaftun EK, Verguet S, Norheim OF, Johansson KA. Geographic health inequalities in Norway: a Gini analysis of cross-county differences in mortality from 1980 to 2014. Int J Equity Health. 2018 May 24;17(1):64.

Statistics Denmark. Dansk Branchekode 1993, udgave 2. 1996. URL: http://www.dst.dk/da/Statistik/Publikationer/VisPub?cid=4829 [accessed 2016-06-10]

Statistics Denmark. SOCIO Danmarks Statistiks Socioøkonomiske Klassifikation. Copenhagen: Statistics Denmark; 1997.

Thiele TN. (1872). “On a mathematical formula to express the rate of mortality throughout life.” Journal of the Institute of Actuaries, 16:313-329.

US Department of Health and Human Services. Healthy people 2000: national health promotion and disease prevention objectives. Washington, DC: US Department of Health and Human Services, Public Health Service, 1990; DHHS publication no. (PHS)90-50212.

Wagstaff A, Paci P, van Doorslaer E. On the measurement of inequalities in health. Soc Sci Med. 1991;33(5):545-57.

WHO. Targets for Health for All. Targets in Support of the European Regional Strategy for Health for All. Copenhagen: WHO Regional Office for Europe, 1985.

WHO. Health 21. The health for all policy framework for the WHO European Region. ed. Copenhagen, World Health Organization, 1999.

WHO. Health Equity Assessment Toolkit (HEAT): Software for exploring and comparing health inequalities in countries. Built-in database edition. Version 2.1. Geneva, World Health Organization, 2018. [cited 2019 June 8]. Available from: http://origin.who.int/gho/health_equity/heat_technical_notes_vFeb2018.pdf

Wittstein J. (1883). "The mathematical law of mortality." Journal of the Institute of Actuaries, 24:152-173.

Wright J, Dugdale B, Hammond I, Jarman B, Neary M, Newton D, Patterson C, Russon L, Stanley P, Stephens R, Warren E. Learning from death: a hospital mortality reduction programme. J R Soc Med. 2006 Jun;99(6):303-8.

Zaba B. (1979). "The four-parameter logit life table system." Population Studies, 33:79-100.

Zapata-Diomedi B, Barendregt JJ, Veerman JL. Population attributable fraction: names, types and issues with incorrect interpretation of relative risks. Br J Sports Med. 2018 Feb;52(4):212-213.

Appendix 1

SAS-program to demonstrate bias from random variation in some traditional summary measures of inequality

```
/* This SAS-program uses Monte-Carlo simulation to
demonstrate bias from random variation in data generated
naive estimates of (i) the risk ratio between the group
with the highest vs. lowest risk (Ymax / Ymin), (ii) the
excess fraction (EF), (iii) the index of dissimilarity
(ID) and (iv) the Gini-index.

The simulations are guided by the following rules: (i) The
included groups are ranked in relation to their true risk
rates,(ii) the true risk rate of the groups increases
geometrically with their rank order, (iii) all groups are
equally large and (iv) the observed number of cases follows
a Poisson distribution.

The output variables that are prefixed with ”True_” contain
the values that would have been obtained in the absence of
random variation. The variables that are prefixed with
”O_” represent the values that are expected to occur due
to a combination of true risk inequalities and random
variation. */

data test (keep = Cases True_max_by_min O_max_by_min
True_ID O_ID True_EF O_EF True_Gini O_Gini);

  groups = 50; /* Number of groups */
```

```
  Cases = 5000; /* Total number of expected cases */

  True_max_by_min = 1.5; /* True Ymax / Ymin */

  array x(50); /* Observed number of cases, by group */

  array True_srr(50); /* True group rate divided by true
mean rate in the total population */

  array O_srr(50);  /*  Observed  group  rate  divided  by
observed mean rate in the total population */

  do k = 1 to 100000;

    Sum = 0;

    do i = 1 to groups;

      sum  =  sum  +  True_max_by_min  **  ((i  -  1)/(groups  -
1));

    end;

    Sum_x = 0;

    do i = 1 to groups;

      True_srr(i)  =  True_max_by_min  **  ((i  -  1)/(groups  -
1)) * groups / sum;
      x(i) = RAND('POISSON',True_srr(i) * cases / groups);
      sum_x = sum_x + x(i);

    end;

    do i = 1 to groups;

      O_srr(i) = x(i) * groups / sum_x;
```

```
    end;

    do i = 1 to groups;
      do j = 1 to groups - i;
        if O_srr(j) > O_srr(j + 1) then do;
          Temp = O_srr(j);
          O_srr(j)= O_srr(j + 1);
          O_srr(j + 1) = Temp;
        end;
      end;
    end;

    True_ID = 0; O_ID = 0; True_Gini = 0; O_Gini = 0;

    do i = 1 to groups;

      True_ID = True_ID + abs(True_srr(i) - 1);
      O_ID = O_ID + abs(O_srr(i) - 1);

      True_Gini = True_Gini + True_srr(i) * (i - 0.5) /
groups ** 2;
      O_Gini = O_Gini + O_srr(i) * (i - 0.5) / groups **
2;

    end;

    O_max_by_min = O_srr(50) / O_srr(1);

    True_ID = True_ID / groups / 2;
    O_ID = O_ID / groups / 2;

    True_Gini = 2 * True_Gini - 1;
    O_Gini = 2 * O_Gini - 1;

    True_EF = 1 - True_srr(1);
    O_EF = 1 - O_srr(1);

    output;
```

```
  end;

run;

proc means data = test; var Cases True_max_by_min
O_max_by_min True_ID O_ID True_EF O_EF True_Gini O_Gini;
run;
```

Appendix 2

SAS-program to estimate an excess fraction

```
/* The input to the this SAS-program comes from a data set
named "ihd", which contains the information given in table
3.3. In the first series of data steps and procedures, the
observed number of cases (cases) and the age and gender
standardised expected number of cases (expect1), for each
of the 57 industrial groups that are included in table
3.3, are read, transposed and merged into a new data set
named "observed", which contains one observation and 114
variables, m1 - m57 (the expected numbers of cases) and x1
- x57 (the observed numbers of cases).*/

proc sort
  data = ihd;
  by smr1; /* smr1 = Age and gender standardised morbidity
ratio */
run;

data cases (keep = cases);
  set ihd;
run;

proc transpose
  data = cases
  out = cases
  prefix = x
  ;
run;
```

```
data cases;
  set cases;
  drop _name_;
run;

data expect1 (keep = expect1);
  set ihd;
run;

proc transpose
  data = expect1
  out = expect1
  prefix = m
  ;
run;

data expect1;
  set expect1;
  drop _name_;
run;

data observed;
  merge cases expect1;
run;

/* The Pearson chi-square statistic (obs_chi2) is
calculated and an array with rank orders is added to the
data set, which now contains 172 variables, m1 - m57, x1
- x57, order1 - order57 and obs_chi2. */

data observed;
  set observed;

  array m(*) m1 - m57;
  array x(*) x1 - x57;
  array order(57);

  do j = 1 to dim(m);
    order(j) = j;
```

```
  end;

  sum_m = 0;
  sum_x = 0;
  do j = 1 to dim(m);
    sum_m = sum_m + m(j);
    sum_x = sum_x + x(j);
  end;

  obs_chi2 = 0;
  do j = 1 to dim(m);
    m(j) = m(j) * sum_x / sum_m;
    obs_chi2 = obs_chi2 + (x(j) - m(j))**2 / m(j);
  end;

  drop sum_x sum_m j;
run;

/* The data are replicated 100,000 times. The result is a
data set containing 100,000 identical observations and 172
variables. */

data observed;
  set observed;
  do j=1 to 100000;
    output;
  end;
run;

/* An empty dataset is prepared for the simulations. */

data simulations;
run;

/* The macro is defined and executed 99 times: For excess
fractions from 0.01 to 0.99 by step of size 0.01 (Division
by 100 is done later). Output from the macro is suppressed.
*/
```

```
options mprint;

%macro simul2;

  %do ef = 1 %to 99 %by 1;

    data temp1;
      set observed;
      array m(*) m1 - m57;
      array x(*) x1 - x57;
      array order(57);

/* A random ordering of the 57 industrial groups is obtained
by a call to the ranperm function. */

      seed = 0;
      call ranperm(seed, of order1 - order57);

/* The constant "a" (the ratio between the risk in the
group with rank order j + 1 and the group with rank order
j) that corresponds to the current excess fraction and
random rank ordering is found by solving the equation f(a)
= 0, by use of the Newton-Raphson method (50 iterations).
The function f(a) is given by the variable "fa" and its
derivative is given by the variable "dfa".*/

      ef = &ef / 100;
      a = 2;
      do i = 1 to 50;
        fa = 0;
        dfa = 0;
        do j = 1 to dim(m);
          fa = fa + m(j) * ((1 - ef) * a ** (order(j) - 1)
- 1);
          dfa = dfa + m(j) * (1 - ef) * (order(j) - 1) *
a ** (order(j) - 2);
        end;
        a = a - fa / dfa;
      end;
```

```
/* Observed numbers corresponding to the distribution
given by the excess fraction, the constant ratio and the
random orders are simulated. The expected numbers
corresponding to the simulated observed values are
calculated
(Notice the simple form). Finally the Pearson chi-square
statistic is calculated from the simulated data. */

      sum_m = 0;
      sum_x = 0;
      chi2 = 0;
      do j = 1 to dim(m);
        x(j) = rand('poisson',(1 - ef) * m(j) * a **
(order(j) - 1));
        sum_m = sum_m + m(j);
        sum_x = sum_x + x(j);
      end;
      do j = 1 to dim(m);
        m(j) = m(j) * sum_x / sum_m;
        chi2 = chi2 + (x(j) - m(j)) ** 2 / m(j);
      end;
      drop sum_x sum_m j;
    run;

/*A line containing the excess fraction (ef), the mean of
the simulated chi-square sums (chi2), the mean of the
constant ratio (a), and the observed chi-square sum
(obs_chi2)is appended to the data set 'simulations'.*/

    proc means data = temp1 noprint;
      var
        ef
        a
        chi2
        obs_chi2
      ;
      output out = temp2;
    run;
```

```
    data temp2;
      set temp2;
      where _stat_ = "MEAN";
      drop
        _type_
        _freq_
        _stat_
      ;
    run;

    data simulations;
      set simulations temp2;
    run;

  %end;

%mend simul2;

%simul2

proc print data = simulations;
  var
    ef
    chi2
    obs_chi2
  ;
run;
```

Appendix 3

SAS-program to estimate a confidence interval of Pearson's chi-square sum

```
/* If the excess fraction has been estimated at e.g. 0.26
then change the following line of the SAS-program given in
Appendix 2

from

"%do ef = 1 %to 99 %by 1;"

to

"%do ef = 26 %to 26 %by 1;"

and then rerun that program.

After that, the following program will provide a 95%
confidence interval of Pearson's chi-square sum, which in
turn can be used to estimate a 95% confidence interval of
the excess fraction. */

proc univariate data = temp1 noprint;
  var chi2;
  output
    out = CI
    pctlpts = 2.5 97.5
    pctlpre = pop_
  ;
```

```
run;

proc print data = CI; run;
```

Appendix 4

Industrial group classification

AT49X is an industrial classification, which contains 57 different industrial groups. The classification is an aggregation of the Danish Industrial Classification of All Economic Activities 1993 (DB93), which is a national version of the European Industrial Classification of All Economic Activities, 1993. The classification is used in Paper VII, VIII and IX as well as in chapter three of the present thesis.

Table A.1. The industrial groups of AT49X in relation to DB93

AT49X	*DB93*	
010. Metal and steelworks, and foundries	271000	Manufacture of basic iron and steel
	272100	Manufacture of cast iron tubes
	273100	Cold drawing
	273200	Cold rolling of narrow strips
	273300	Cold forming or folding
	273500	Other first processing of iron and steel n.e.c.
	274100	Precious metals production
	274200	Aluminium production
	274300	Lead, zink and tin production
	274400	Copper production
	274500	Other non-ferrous metal production
	275100	Casting of iron
	275200	Casting of steel
	275300	Casting of light metals
	275400	Casting of other non-ferrous metals
	314000	Manufacture of accumulators, primary cells and batteries
	371000	Recycling of metal waste and scrap
020. Manufacture of transport equipment	291110	Manufacture of ship engines
	291120	Repair of ship engines
	341000	Manufacture of motor vehicles

	342000	Manufacture of coachwork for motor vehicles
	343000	Manufacture of parts for motor vehicles
	351200	Building and repairing of pleasure and sporting boats
	352000	Manufacture of locomotives and rolling stock
	353000	Manufacture of aircraft
	354200	Manufacture of bicycles
	354300	Manufacture of invalid carriages
	355000	Manufacture of other transport equipment n.e.c.
030. Shipyards	351100	Building and repairing of ships
040. Electricity and heat supply	401000	Production and distribution of electricity
	403000	Steam and hot water supply
050. Iron and metal industries	272200	Manufacture of steel tubes
	281100	Manufacture of metal structures
	281200	Manufacture of metal building material
	282100	Manufacture of metal tanks and containers
	282200	Manufacture of central heating radiators and boilers
	283000	Manufacture of steam generators
	284000	Forging, pressing, stamping and rollforming of metal
	285100	Treatment and coating of metals
	286100	Manufacture of cutlery
	286310	Manufacture of locks
	286320	Manufacture of metal fittings
	287100	Manufacture of steel drums and similar containers
	287200	Manufacture of light metal packaging
	287300	Manufacture of wire products
	287400	Manufacture of fasteners, chains, springs etc.
	287510	Manufacture of metal sign plates
	287520	Manufacture of metal household appliances
	287590	Manufacture of other fabricated metal products n.e.c.
	291300	Manufacture of taps and valves
	296000	Manufacture of weapons and ammunition
	362100	Striking of coins and medals
	362210	Gold- and silverware factories
	362220	Goldsmiths
060. Engineering industry	285200	General mechanical engineering (contractors)
	286200	Manufacture of tools
	291190	Manufacture of other motors and turbines
	291210	Manufacture of air pumps and compressors
	291220	Manufacture of pumps for liquids

291230	Manufacture of fluid and pneumatic power engines
291400	Manufacture of bearings, gears etc.
292100	Manufacture of furnaces and furnace burners
292210	Manufacture of conveyor belts and elevators
292220	Manufacture of cranes and other lifting devices
292230	Manufacture of work trucks
292290	Manufacture of other handling equipment
292310	Manufacture of non-domestic cooling equipment
292320	Manufacture of non-domestic ventilation equipment
292410	Manufacture of weighing machinery
292420	Manufacture of packing and wrapping machinery
292430	Manufacture of fire extinguishers, sandblasters etc.
292490	Manufacture of other general purpose machinery n.e.c.
293100	Manufacture of agricultural tractors
293210	Manufacture of harvesting machinery etc.
293220	Manufacture of agricultural machinery for soil preparation etc.
293230	Manufacture of other agricultural and forestry macinery n.e.c.
293240	Repair of agricultural and forestry machinery
294000	Manufacture of machine tools
295100	Manufacture of machinery for metallurgy
295210	Manufacture of concrete and mortar mixers
295290	Manufacture of other machinery for mining, quarrying and construction
295310	Manufacture of machinery for the dairy industry
295320	Manufacture of machinery for the grain milling industry
295330	Manufacture of machinery for the bakery industry
295340	Manufacture of machinery for the fish and meat industry
295390	Manufacture of other machinery for food beverage and tobacco
295400	Manufacture of machinery for textile, apparel and leather
295500	Manufacture of machinery for paper and paperboard production
295610	Manufacture of moulds
295620	Manufacture of drying systems
295690	Manufacture of other special purpose machinery
297110	Manufacture of domestic refrigerators and freezers
297130	Manufacture of domestic washing and drying machines
297200	Manufacture of non-electric domestic appliances
300100	Manufacture of office machinery
300200	Manufacture of IT equipment
311040	Manufacture of windmills

070. Electricity and electronics industry	297120	Manufacture of domestic electric ovens
	297190	Manufacture of other electric domestic appliances
	311010	Manufacture of electric motors and generators
	311020	Manufacture of electric generator sets
	311030	Manufacture of electrical transformers
	312010	Manufacture of electric control or distribution boards
	312090	Manufacture of switches, fuses, etc.
	313000	Manufacture of insulated wire and cable
	315000	Manufacture of lighting equipment
	316100	Manufact. of electrical equipment for engines and vehicles n.e.c.
	316210	Manufacture of electrical traffic regulation equipment
	316220	Electromechanical workshops
	316290	Manufacture of electrical equipment n.e.c.
	321010	Manufacture of printed circuits
	321090	Manufacture of other electronic components
	322010	Manufacture of apparatus for radio-telephony or radio-telegraphy
	322020	Manufacture of other telephony apparatus etc.
	323010	Manufacture of television and radio receivers
	323020	Manufacture of loud speakers etc.
	323030	Manufacture of antennas etc.
	332010	Manufacture of navigation equipment
	332020	Manufact. measuring instruments for gas and liquid pressure
	332030	Manufacture of measuring instruments for electrical quantities
	332040	Manufacture of apparatus for physical and chemical analysis
	332090	Manufacture of measuring and control instruments n.e.c.
	333000	Manufacture of industrial process control equipment
	335000	Manufacture of clocks and watches
080. Car industry	501010	Wholesale, cars
	502010	Maintenance and repair of motor vehicles
	502020	Body work of motor vehicles
	502030	Auto electricians
	502040	Undersealing of cars
	502050	Car lacquers
	502060	Tyre and tube repair
	502090	Other servicing of motor vehicles
	527210	Repair, electrical houshold machines
	527220	Repair, radio and television
	527410	Repair, bicycle
	725000	Maintenance and repair of office machinery

	743020	Testing activities, cars, ships etc.
090. Road contractors etc.	451100	Demolition of buildings; earth moving
	451200	Test drilling and boring
	452100	General construction of buldings and civil engineering work
	452520	Paving
	452530	Sewer contractors
	452540	Scaffold suppliers
	452590	Other construction work involving special trades
	455000	Renting of construction equipment
100. Bricklayer, joiner, and carpentry work	452200	Erection of roof coverings and frames
	452510	Brick laying
	454200	Joinery installation
110. Finishing	454100	Plastering
	454310	Floor and wall covering
	454320	Planing of floor boards
	454410	Painting firms
	454420	Glazing firms
	454500	Other building completion
120. Insulation and installation businesses	453100	Installation of electrical wiring and fittings
	453200	Insulation work activities
	453300	Plumbing
130. Printing works and publishing	221110	Publishing of books etc. with publishing house
	221120	Publishing of books etc. without publishing house
	221210	Publishing of newspapers, with publishing house
	221220	Publishing of newspapers, without publishing house
	221310	Publishing of journals, with publishing house
	221320	Publishing of journals, without publishing house
	221330	Publishing of advertisment sheets with publishing house
	221340	Publishing of advertisment sheets, without publishing house
	221400	Publishing of sound recordings
	221500	Other publishing
	222100	Printing of newspapers
	222210	Book and offset printing
	222230	Serography
	222290	Printing n.e.c.

	222410	Reproduction
	222420	Composing
	222500	Other activities related to printing
	223100	Reproduction of sound recording
	223200	Reproduction of video recording
	223300	Reproduction of computer media
140. Paper, cardboard and bookbinding industries	212100	manufacture of corrugated paper and paperboard
	212200	Manufacture of household and toilet requisites of paper
	212310	Manufacture of writing paper
	212390	Manufacture of office articles in paper
	212500	Manufacture of other paper articles n.e.c.
	222300	Bookbinding and finishing
150. Wholesale trade	157110	Manufacture of prepared feeds for farm animals
	157120	Manufacture of prepared feeds for fish
	158600	Processing of tea and coffee
	159300	Manufacture of wine from fresh grapes
	159400	Manufacture of other fruit wines
	503010	Wholesale, motor vehicle parts
	511710	Agents, fishing auctions
	511790	Agents, other food, beverages and tobacco
	512100	Wholesale, grain, seeds, animal feeds
	512200	Wholesale, flowers and plants
	512300	Wholesale, live animals
	512400	Wholesale, hides, skins and leather
	513100	Wholesale, fruit and vegetables
	513200	Wholesale, meat and meat products
	513300	Wholesale, dairy produce, eggs, edible oil&fats
	513410	Wholesale, beer and soft drinks
	513420	Wholesale, wine and spirits
	513490	Wholesale, other beverages n.e.c.
	513500	Wholesale, tobacco
	513600	Wholesale, sugar, chocolate, confectionary
	513700	Wholesale, coffe, tea, cocoa, spices
	513810	Wholesale, fish and fish products
	513820	Wholesale, bread and biscuits
	513830	Wholesale,health foods
	513890	Other specialized wholesale, food, beverage n.e.c.
	513900	Non-specialized wholesale, beverage, food and tobacco
	514100	Wholesale, textiles
	514210	Wholesale, clothing

	514220	Wholesale, footwear
	514310	Wholesale, white goods
	514320	Wholesale, radio and television goods
	514330	Wholesale, gremmophone records, video tapes etc.
	514340	Wholesale, electrical household appliances
	514410	Wholesale, china and glassware
	514420	Wholesale, cleaning materials
	514500	Wholesale, perfume and cosmetics
	514610	Wholsale, pharmaceutical goods
	514620	Wholesale, medical appliances
	514705	Wholesale, furnitures
	514710	Wholesale, floor coverings
	514715	Wholesale, jewellery
	514720	Wholesale, clocks and optical goods
	514725	Wholesale, photographic goods
	514730	Wholesale, bicycles
	514735	Wholesale, sport articles etc.
	514740	Wholesale, games and toys
	514745	Wholesale, books and paper products
	514750	Wholesale, luggage, leather products etc.
	514790	Wholesale, houshold articles n.e.c.
	515100	Wholesale, fuels and related products
	515200	Wholesale, metal and metal ores
	515310	Wholesale, construction materials
	515320	Wholesale, paint, varnish, wallpaper
	515400	Wholesale, hardware, plumbing and heating equipment
	515500	Wholesale, chemical products
	515610	Wholesale,packing and wrapping articles
	515690	Wholesale, other intermediate products
	515700	Wholesale, waste and scrap
	516100	Wholesale, machine tools for working metal and wood
	516200	Wholesale, construction machinery
	516300	Wholesale, machinery for the textile industry
	516410	Wholesale, office machinery
	516420	Wholesale, office furniture and appliances
	516510	Wholesale, machinery for electrical installation
	516520	Wholesale, electronic components
	516590	Wholesale, other machinery for use in industry etc.
	516600	Wholesale, agrcultural machinery
	517000	Other wholesale
	524890	Retail, fuel for domestic use
160. Transport of goods	602410	Haulage contractors
	602420	Furniture removers

	611010	Shipping, goods
	622020	Non-scheduled air transport of goods
	631100	Cargo handling
	631200	Storage and warehousing
	632110	Freight terminals
	632210	Harbours, industrial
	641100	National post activities
	641200	Other courier activities
	900020	Refuse disposal
170. Transport of passengers	601000	Transport via railways
	602100	Other scheduled land transport
	602200	taxi operation
	602300	Other non-scheduled road passenger transport
	611020	Shipping, passengers
	612000	Inland water transport
	621000	Scheduled air transport
	622010	Chartered air transport of passengers
	622030	Air taxi operation
	632120	Parking houses and garages
	632220	Harbours, pleasure boats
	632300	Airway terminals etc.
180. Fire service, lighthouse and salvage corps	632230	Lighthouse activities and pilotage
	632240	Salvage activities
	752500	Fire service activities
190. Textile, clothing, and leather industry	171000	Preparation of spinning of textile fibers
	172000	Textile weaving
	173000	Finishing of textiles
	174010	Manufacture of sails, flags and tents
	174020	Manufacture of furnishing fabrics
	174090	Manufacture of other made-up textile articles
	175100	Manufacture of carpets and rugs
	175210	Manufacture of ropes
	175220	Manufacture of fish nets
	175400	Manufacture of other textiles n.e.c.
	176000	Manufacture of knitted and crocheted fabrics
	177100	Manufacture of knitted and crocheted hosiery
	177200	Manufacture of knitted and crocheted pullovers, cardigans etc.
	181000	Manufacture of leather clothes

	182100	Manufacture of workwear
	182210	Manufacture of outerwear (men and women)
	182220	Manufacture of dresses and trousers (women and girls)
	182230	Manufacture of coates and trousers (men and boys)
	182310	Manufacture of shirts
	182390	Manufacture of underwear n.e.c.
	182410	Manufacture of baby clothes
	182490	Manufacture of other clothes n.e.c.
	183000	Dressing and dyeing of fur; manufacture of articles of fur
	191000	Tanning and dressing of leather
	192000	Manufacture of luggage, handbags etc.
	193010	Manufacture of shoes
	193020	manufacture of wooden shoes
	372000	Recycling of non-metal waste and scrap
200. Manufacture of wood and wood products	201010	Sawmills
	201020	Impregnation of wood
	202000	Manufacture of veneer sheets etc.
	203010	Manufacture of frames and furniture mouldings
	203020	Manufacture of wooden parts for buildings
	203030	Manufacture of wooden bulding units
	204000	Manufacture of wooden containers
	205110	Woodturners
	205120	Manufacture of coffins
	205190	Manufacture of other wooden products n.e.c.
	205200	manufacture of articles of cork, straw and plaiting
	361110	Manufacture of chairs and seats
	361120	Upholstery of chair and seats
	361200	Manufacture of other office and shop furniture
	361300	Manufacture of kitchen furniture
	361410	Manifacture of other furniture
	361490	Furniture lacquers
	361500	Manifacture of mattresses
210. Mineral, oil, rubber and plastic products	232000	Manufacture of refined petroleum products
	251100	Manufacture of rubber tyres and tubes
	251200	Retreading and rebuilding of rubber tyres
	251300	Manufacture of other rubber products
	252110	Manufacture of plastic plates and sheets
	252120	Manufacture of plastic tubes
	252130	Manufacture of plastic profiles
	252200	Manufacture of plastic packing goods

	252310	Manufacture of plastic sanitary appliances
	252390	Manufacture of other plastic building material
	252410	Manufacture of plastic office appliances
	252420	Manufacture of plastic kitchen appliances
	252490	Manufacture of other plastic products n.e.c.
	268210	Manufacture of asphalt and roofing felt
220. Stone-works, pottery, and glass industry	261100	Manufacture of flat glass
	261200	Shaping and processing of flat glass
	261300	Manufacture of hollow glass
	261400	Manufacture of glass fibres
	261500	Manufacture and processing of other glass n.e.c.
	262100	Manufacture of ceramic household articles
	262200	Manufacture of ceramic sanitary fixtures
	262300	Manufacture of ceramic insulators
	262400	Manufacture of other technical ceramic products
	262500	Manufacture of other ceramic products
	262600	Manufacture of refractory ceramic products
	263000	Manufacture of ceramic tiles and flags
	264000	Manufacture of bricks etc.
	266110	Manufacture of concrete products
	266120	Manufacture of building units in concrete
	266200	Manufacture of plaster products for buildings
	266500	manufacture of fibre cement
	266600	Manufacture of other articles of concrete, plaster and cement
	267000	Cutting, shaping and finishing of stone
	268220	Manufacture of rockwool
	268290	Manufacture of other non-metallic mineral products n.e.c.
230. Medical equipment, toys, cameras etc.	212400	Manufacture of wallpaper
	246500	Manufacture of prepared unrecorded media
	331010	Manufacture of syringes and hypodermic needles
	331020	Manufacture of hearing aids
	331030	Manufacture of electro-diagnostic devices
	331040	Manufacture of medical or dental furniture and equipment
	331090	Manufacture of other medical and surgical equipment
	334010	Manufacture of optical instruments
	334020	Manufacture of reproduction cameras
	334090	Manufacture of photographic and cinema equipment
	363000	Manufacture of musical instruments
	364000	Maanufacture of sport goods

	365000	Manufacture of games and toys
	366100	Manufacture of imitation jewellery
	366200	Manufacture of brooms and brushes
	366390	Other manufacturing n.e.c.
240. Manufacture of industrial chemicals	241100	Manufacture of industrial gases
	241200	Manufacture of dyes and pigments
	241300	Manufacture of other inorganic basic chemicals
	241400	Manufacture of other organic basic chemicals
	241500	Manufacture of fertilizers etc.
	241600	Manufacture of plastics in primary forms
	242000	Manufacture of agro-chemical products
	243000	Manufacture of paints, varnishes etc.
	245110	Manufacture of soap detergents
	245120	Manufacture of cleaning and polishing preparations
	245200	Manufacture of perfumes and toilet preparations
	246100	Manufacture of explosives
	246200	Manufacture of glues and gelatine
	246400	Manufacture of photographic chemical material
	246600	Manufacture of other chemical products n.e.c.
	247000	Manufacture of man-made fibres
	366310	Manufacture of stearin candles
250. Heavy raw material and semi-manufacture	103000	Extraction and agglomeration of peat
	111000	Extraction of crude petroleum and natural gas
	112000	Service activities incidental to oil and gas extraction
	120000	Mining of uranium and thorium ores
	141110	Quarrying of stone for construction
	141120	Stonefishing
	141200	Quarrying of limestone
	142100	Operation of gravel and sand pits
	142200	Mining of clays and kaolin
	143000	Mining of chemicaland fertilizer minerals
	144000	Production of salt
	145000	Other mining and quarrying n.e.c.
	211100	Manufacture of pulp
	211200	manufacture of paper and paperboard
	265100	Manufacture of cement
	265200	Manufacture of lime
	266300	Manufacture of ready-mixed concrete
	266400	Manufacture of mortars
	268100	Production of abrasive products

260. Pharmaceutical industry	244100	Manufacture of pharmaceutical products
	244200	Manufacture of pharmaceutical preparations
271. Office & adm. (transport & wholesale)	511100	Agents, agricultural raw material, animals, textile and semi-finished goods
	511200	Agents, fuels, ores, metals and industrial chemicals
	511300	Agents, timber and building material
	511400	Agents, machinery, industrial equipment, ships and aircrafts
	511500	Agents, furniture, household goods, hardware and iron
	511600	Agents, textiles, clothing, footwear, leather goods
	511800	Agents, particular products n.e.c.
	511900	Agents, a variety of goods
	634010	Shipbrokers
	634020	Shipping agents
	634030	Weighers and measurers
	634090	Activities of other transport agencies
	711000	Rental of cars
272. Office & adm. (service)	702010	Non-profit housing associations
	702020	Private housing cooperatives
	702030	Other rental of dwellings
	702040	Non-residential buildings
273. Finance/ Public office & adm	651100	Central banking
	651200	Other monetary intermediation
	652100	Financial leasing
	652210	Morgage credit institutes
	652295	Other credit granting
	652310	Investment in securities
	652395	Other financial intermediation n.e.c.
	660100	Life insurance
	660210	Pension funding
	660290	Other deferred annuity assurances
	660310	Accident insurance
	660320	Health insurance
	660390	Other insurance activities
	671100	Administration of financial markets
	671200	Security dealing activities
	671300	Activities auxiliary to financial intermediation n.e.c.
	672010	Insurance agencies
	672090	Other activities auxiliary to insurance and pension funding
	751100	General public service activities

	751200	Regulation, health care, education and social services
	751300	Regulation, business operations
	751400	Ancillary service for the government
	752100	Foreign affairs
	752310	Law courts
	753000	Compulsory social security activities
274. Private office & adm.	633010	Tourist agencies
	633020	Travel agencies, activities of tour operators
	633030	Travel agencies, sale of tickets
	633040	Tourist guide activities
	701100	Development and selling of real estate
	701200	Buying and selling of own or leased real estate
	703110	Real estate agency
	703120	Accomodation agency
	703130	Rental of holiday homes
	703220	Societies for people in owner-occupied dwellings
	712110	Rental of containers
	712190	Rental of lorries
	712200	Rental of ships
	712300	Rental of air crafts
	713100	rental of agricultural machinery
	713200	Rental of construction equipment
	713310	Rental of computers and related equipment
	713320	Rental of office machinery
	713400	Rental of other machinery n.e.c.
	721000	Hardware consultancy
	722000	Software consultancy and supply
	723000	Data processing
	724000	Data base activities
	726000	Other computer related activities
	741100	Legal activities
	741200	Accounting, book-keeping etc.
	741300	Market research
	741410	Consultant activities, agriculture
	741490	Consultant activities, other industries
	741500	Management holding companies
	742010	Consultant engineers, construction
	742020	Consultant engineers, machinery
	742030	Consultant engineers, industrial plant design
	742040	Architects
	742050	Landscape architects
	742060	Geological and prospecting activities
	742070	Chartered surveyors etc.
	742080	Patent agencies

	742090	Other technical consultancy
	743010	Testing activities, food hygiene
	743030	Testing activities, outdoor water and air quality etc.
	743090	Other testing and analysis activities
	744010	Advertising agencies
	744090	Other activities related to advertisement
	745010	Employment agencies
	745020	Temporary employment agencies
	745030	Personell procurement
	748200	Packaging activities
	748310	Envelope addressing
	748320	Translation and interpretation
	748390	Other secretarial services
	748410	Interior decorators
	748420	Designers
	748430	Evaluators of creditworthiness
	748440	Organisers of exhibitions etc.
	748490	Other buisiness activities n.e.c.
	853255	Public health associations
	853260	Charity funds
	911100	Business and employers organisations
	911200	Activities of professional organisations
	912000	Trade unions
	913200	Political organisations
	913310	Tenants' associations
	913320	Organisations for outdoor life
	913330	Other ideological organisations
	913340	Other cultural organisations
	913390	Social clubs etc.
	924000	News agency activities
	990000	Extra-territorial organisations
281. Car dealers	501020	Retail, cars
	504000	Sale and repair of motorcycles and accessories
282. Garage	505010	Retail, automotive fuel without kiosk sale
	505020	Retail, automotive fuel and kiosk sale
283. Shops	501030	Wholesale and retail, camping vehicles etc.
	503020	Retail, motor vehicle parts
	521120	Kiosks
	522100	Retail, fruit and vegetables
	522200	Retail, meat and meat products
	522300	Retail, fish, crustaceans and molluscs

522410	Retail, bread
522420	Retail, confectionary
522500	Retail, alcoholic and other beverages
522600	Retail, tobacco
522710	Retail, cheese
522730	Retail, health foods
522790	Retail, other specialized food or beverage products
523200	Retail, medical and orthopaedic goods
523310	Retail, perfume
523320	Retail, cosmetics
524100	Retail, textiles
524210	Retail, clothing for women
524220	Retail, clothing for men
524230	Retail, clothing for men and women
524240	Retail, clothing for children
524310	Retail, footwear
524320	Retail, leather goods
524430	Retail, domestic textiles
524440	Retail, kichen articles
524450	Retail, lighting equipment
524530	Retail, musical records, CD's etc.
524540	Retail, musical instruments
524610	Retail, hardware
524630	Retail, paint and wallpaper
524700	Retail, books, newspapers, stationary
524805	Retail, clocks
524810	Retail, clocks and jewellery
524815	Retail, jewellery
524820	Retail, optical equipment
524825	Retail, photography
524830	Retail, souvenirs
524835	Retail, art
524840	Retail, stamps and coins
524845	retail, sport articles
524850	Retail, games and toys
524855	Retail, pleasure boats and related equipment
524860	Retail, bicycles
524865	Retail, computers and software
524870	Retail, telecommunication
524875	Retail, flowers
524880	Retail, garden centers
524885	Retail, pets
524895	Retail, pornography
524899	Retail, other commodities
525010	Retail, second-hand books

	525020	Retail, antiquities
	525090	Retail, other second-hand goods
	526100	Retail, mail order houses
	526210	Retail, fruit and vegetable stalls
	526290	Other retail via stalls
	526300	Other non-store retail sale
	527300	Repair, clocks
	714010	Rental of video recordings
	900030	Rubbish dumps
290. Supermarkets, department stores etc.	521110	Grocer's
	521130	Supermarkets
	521210	Smaller department stores
	521220	Larger department stores
	524410	Retail, furniture
	524420	Retail, carpets
	524510	Retail, electric household appliances
	524520	Retail, radio and television
	524620	Retail, building material
300. Sewers, water- and gas supply	402000	Manufacture and distribution of gas
	410000	Collection, purification and distribution of water
	900010	Sewage disposal
310. Personal care and other services	527100	Repair of boots and shoes
	527420	Locksmiths
	527490	Repair n.e.c.
	703210	Management of real estate on a contract basis
	714090	Rental of goods for domestic and personal use
	747050	Desinfection and extermination activities
	930210	Hairdressing
	930220	Beuty treatment
	930310	Funeral undertakers
	930320	Burial authorities
	930400	Physical well-being activities
	930500	Other service activities n.e.c.
	950000	Private housholds with employees
320. Cleaning, laundries, and dry cleaners	747010	General cleaning activities
	747020	Specialised cleaning activities
	747030	Window cleaning
	747040	Chimney sweeping

	930110	Laundries
	930120	Management of self-service laundries
	930130	Dry-cleaning
330. Telecommunication	642000	Telecommunication
340. Surveillance, armed forces, police etc.	746000	Investigation and security activities
	752200	Defence activities
	752320	Prison administration
	752400	Police force
	802270	Police and defence schools
350. Hotels and restaurants	551110	Hotels with restaurant
	551120	Conference centres
	551200	Hotels without restaurant
	552100	Youth hostels
	552200	Camping sites
	552310	Holiday chalets and flats
	552390	Other facilities n.e.c.
	553010	Restaurants
	553020	Cafeterias, icecream bars and takeaways
	553090	Rooms for private parties
	554010	Public houses and wine bars
	554020	Discotheques and night clubs
	554090	Coffe bars etc.
	555100	Canteens
	555200	Catering
361. Photographers, film and videoproduction	748110	Photographers
	748120	Film processing
	921100	Motion picture and video production
	921200	Motion picture and video distribution
362. Entertainment, culture and sport	921300	Motion picture projection
	922000	Radio and television activities
	923110	Production of theatrical and concert presentations
	923120	Self-employed artists
	923200	Operation of art facilities
	923300	Amusement parks
	923400	Other entertainment activities n.e.c.
	925200	Museums activities

	925300	Botanical and zoological gardens
	926110	Swimming pools and stadiums
	926190	Other sport facilities
	926210	Athletic associations
	926290	Other sporting activities n.e.c.
	927100	Gambling and betting activities
	927200	Other recreational activities n.e.c.
363. Libraries and archives	925110	Public libraries
	925120	Research and university libraries
	925130	Archives
370. Slaughterhouse industry	151110	Slaughterhouse (swine)
	151120	Slaughterhouse (cattle)
	151130	Gut scraping
	151140	Processing of animal offal; flours and meals of meat
	151390	Other processing of meat
380. Poultry slaughtering and fish products	151200	Production of poultry meat
	152010	Production of fish fillets etc.
	152020	Smoke-curing and salting of fish
	152030	Manufacture of fish meal
	157200	Manufacture of prepared feeds for pets
390. Beverage industry	151310	Production ready-prepared dish
	153100	Processing and preserving of potatoes
	153200	Manufacture of fruit and vegetable juice
	153300	Processing and preservation of fruit and vegetables n.e.c.
	154100	Manufacture of crude oils and fat
	154200	Manufacture of refined oils and fat
	156200	Manufacture of starches and starch products
	158300	Manufacture of sugar
	158800	Manufacture of homogenised preparations and dietetic food
	159100	Manufacture of distilled potable alcoholic beverages
	159600	Manufacture of beer
	159700	Manufacture of malt
	159800	Production of mineral waters and soft drinks
400. Manufacture of bread, chocolate, tobacco etc.	156100	Manufacture of grain mill products

	158110	Bread factories
	158120	Bakeries
	158200	Manufacture of rusks and biscuits
	158400	Manufacture of chocolate and sugar confectionary
	158500	Manufacture of macaroni, noodles and similar products
	158700	manufacture of condiments and seasoning
	158900	Manufacture of other food products, n.e.c.
	160000	Manufacture of tobacco products
410. Manufacture of dairy products	154300	Manufacture of margarine and similar edible fats
	155110	Dairies and manufacture of cheese
	155120	Manufacture of condensed milk products
	155200	Manufacture of ice cream
420. Agriculture	11110	Growing of cereals
	11190	Growing of other crops
	12110	Dairy farming
	12190	Other cattle farming
	12210	Stud farm
	12220	Farming of sheep and goats
	12300	Farming of swine
	12400	Farming of poultry
	12510	Farming of furred animals
	12520	Bee-keeping
	12530	Kennels
	12590	Other farming of animals
	13000	Growing of crops combined with farming of animals
	14110	Agricultural machinery stations
	14190	Other agricultural service activities
	14200	Animal husbandry service activities
	50200	Fish hatcheries and fish farms
430. Horticulture and forestry	11210	Horticulture
	11220	Garden centres
	11300	Growing of fruit and berries
	14120	Landscape gardener
	15000	Hunting, trapping and game propagation
	20100	Forestry and logging
	20200	Forestry and logging related service activities
440. Hospitals	851100	Hospitals

450. Nursing homes, home care, etc.	801020	Schools for handicapped people
	853130	Institutions, alcohol and drug dependance
	853140	Residential homes for handicapped adults
	853150	Nursing homes and sheltered accomodations
	853160	Reception centres
	853190	Other care institutions n.e.c.
	853235	Home help
	853240	Day care for elderly
	853245	Revalidation institutons
	853250	Refugee centres
	853290	Other social care activities
460. Child care etc.	853110	Residential homes for children and young people
	853120	family care activities
	853205	Child-minders
	853210	Day nursery
	853215	Kindergarten
	853220	After-school centres
	853225	Institutions for all ages
	853230	Clubs for children
471. General practitioners, dentists etc.	851210	General practitioners
	851220	Special medical practice activities
	851300	
	851310	Practice activities, dentists
	851320	Practice activities, dental technicians
472. Health care n.e.c.	523100	Dispensing chemists
	851410	District nurse activities
	851420	Midwife activities
	851430	Physiotherapy
	851440	Psychologic counselling
	851450	Medical laboratories
	851460	Employees' health service
	851470	Chiropractors
	851480	Chiropodists
	851490	Sanatoria
	852000	Veterinary activities
	913100	Religous organisations
480. Education and research	731000	Research etc., natural sciences and engineering
	732000	Research etc.,social sciences and humanities

	801010	Primary and lower secondary school
	801030	Continuation schools
	802100	Upper secondary school
	802210	Schools with training in pedagogy
	802220	Business and office schools
	802230	Craftmanship schools
	802240	Agricultural schools
	802250	Transport schools
	802260	Health care schools
	803010	Universities
	803020	College, pedagigy
	803030	College, art and humanities
	803040	College, social sciences
	803050	College, technology
	803060	College, agriculture
	803070	College, health care
	804100	Driving schools
	804210	Professional training for adults
	804220	Folk high school
	804230	Other adult education
	804290	Education n.e.c.
490. Fishing	50100	Fishing
990. Unstated	980000	Unstated

Printed by Media-Tryck, Lund 2023 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY
School of Economics and Management
Department of Statistics

ISBN 978-91-8039-719-3