Attributable fractions used for partitioning risk and evaluating disease prevention

Concepts, estimation, interpretation and presentation

Geir Egil Eide

Dissertation for the degree doctor philosophiae (dr. philos.)

at the University of Bergen, Bergen, Norway

2008

Scientific environment

Most of the work with this thesis has been performed in my full-time position as Biostatistician at the Centre for Clinical Research, Haukeland University Hospital, Bergen, from 1999. The centre has provided all administrative and office facilities, computer resources, and expenses for software, literature and publication as well as most of my travel support. Part of the work has been carried out in my Associate Professor II position in the same period at the Section for Epidemiology and Medical Statistics, Department of Public Health and Primary Health Care, University of Bergen. The University of Bergen has also given financial support for the presentation of results at scientific meetings. The first article was written while I was Associate Professor at the Department of Mathematics and Statistics, Norwegian School of Economics and Business Administration, Bergen (1988-1999) and was also supported by the Lung Epidemiological Research Group, Department of Thoracic Medicine, Haukeland University Hospital (1992-1999). I am also indebted to the Department of Statistics, University of California at Berkeley, USA where I spent a sabbatical year 1995/96.

Acknowledgements

Completing a doctoral project may in many ways be a lonely ride. However, without continuing support and regular "drinking holes" manned by enthusiastic colleagues and friends the lone rider would probably perish underway. Their cheering and interest at various stages have been instrumental to the fulfilling of this thesis and I am deeply indebted to all.

An important background for the work herein is my collaboration with Professor Amund Gulsvik and his pulmonary epidemiology research group since 1984. At that time I was at the Section for Medical Informatics and Statistics (SMIS) at the University of Bergen (UoB) and my interest in attributable fractions was first evoked in 1988 by a note from his student, now professor, Per S. Bakke (Appendix I) about how to estimate attributable risks of lung symptoms in the Hordaland Study of Obstructive Lung Disease (HSOLD). Although I changed working place in 1988 to the Norwegian School of Economics and Business Administration (NHH), Amund and his disciples have been the most prominent reason for my continuing involvement in clinical-epidemiological research at the Haukeland University Hospital (HUH) and UoB. Amund was also instrumental in establishing the Centre for Clinical Research (CCR) where I became the first employed person in august 1999. His never-fading faith and encouragement has been crucial for my own belief in fulfilling this thesis, and the data from the first cross-sectional sample of the HSOLD has become the common basis of the examples in four of the five articles included.

Also crucial to the completion of this first article was a young and ambitious lad named Olaf Gefeller whose interest in my work at a conference in Brussels (1991) saved it from falling into oblivion. This was the beginning of our long-lasting co-work and friendship including also his cohabitant and statistician Annette Pfahlberg. Olaf is now a professor in biometry and epidemiology at the University of Nuremberg, Erlangen and the theme of attributable fractions is only one of his many

specialties. Nevertheless, through the years he has continuously involved me in his activities concerning attributable fraction methodology and has been my most important international co-worker. His more or less anonymous marks can be found on most of the articles.

At NHH I am much indebted to the liberal and methodologically inspiring milieu at the small, now defunct, Institute of Mathematics and Statistics. Teaching statistics and mathematics for large and tiny classes was learning-by-doing the hard way. It was in this period the first article was produced, as I wanted to prove to myself that I could be something more than a statistical consultant and co-author.

As a declared doctoral project my work on attributable fractions was not started until 1999 when I started in the position as biostatistician at the CCR and an associate professor II position at the Section for Epidemiology and Medical Statistics (SEMS, previously SMIS) at the Department of Public Health and Primary Health Care, UoB. I am grateful to the leaderships of both units for providing the legitimacy for giving priority to work with own methodological research besides serving the statistical needs of the medical researchers in the Western Health Trust (Helse Vest) and at the university.

Also, I am extremely thankful for the effort of Professor Ivar Heuch, my sparring partner and supervisor for Articles II-IV, whose meticulous reading and pertinent work with the many unfinished manuscripts since 1999 has improved the quality immensely. Maybe surprisingly, our cooperation has been almost solely by email, ensuring that every problem posed has been commented and resolved, often in late-night emails at weekends and holidays. Ivar is impressing focussed, on mathematical and linguistic details as well as on the whole.

For the cooperation on the final paper I am in debt to Associate Professor Sven Ove Samuelsen who came up with the proposal of a joint project on survival data in 2001 which suited perfectly to my own ideas of further methodological developments. This has been an interesting and stimulating project pointing into new directions for the

Last, but not the least, I thank my family for giving my life other dimensions, sharing their love, sticking up with my heavy work load and absenteeism both physically and mentally, giving me a home to come back to, and sharing adventures and experiences of less scientific value but nevertheless invaluable for coping with real life itself.

Bergen 9 October 2007

Abbreviations

AAF   Average attributable fraction

AAFE   Average attributable fraction in exposed

AF   Attributable fraction

AFE   Attributable fraction in exposed

AHF   Attributable hazard fraction

AFB   Attributable fraction before (time t)

AFW   Attributable fraction within study

AV   Attributable squared variation

CCR   Centre for Clinical Research

EAF   Extra attributable fraction

HSOLD   Hordaland Study of Obstructive Lung Disease

HUH   Haukeland University Hospital

MLE   Maximum likelihood estimator

NHH   Norges handelshøyskole (Norwegian School of Economics and
             Business Administration)

PMVD   Proportional marginal variance decomposition

RR   Relative risk (risk ratio)

R^2 Determination coefficient

SAF   Sequential attributable fraction

SEMS   Section for Epidemiology and Medical Statistics

SMIS   Section for Medical Informatics and Statistics (former SEMS)

SSR   Sum of squares due to regression

SSSC   Scaled sample space cube

SSSS   Scaled sample space square (scaled Venn-diagram)

SST   Sum of squares due to total variation

UoB   University of Bergen

2D   Two-dimensional

3D   Three-dimensional

Summary

Background

In medical research some fundamental tasks are to study potential harmful exposures that may give increased risk of getting some disease, potential beneficial treatments that may increase chance of recovering from a disease, or interventions that may reduce the extent or effect of a harmful exposure. In epidemiologic research these questions are studied by collecting individual data for representative samples of the population. For a specified disease (e.g. breast cancer) there will usually be many risk factors, some may be modifiable (e.g. life style factors like smoking habits, physical activity, dietary factors) and other factors not so easy to modify (like reproductive factors, aging, genetic factors). Provided that enough data for the individuals in the sample is collected on the occurrence of disease and the relevant risk factors, statistical models are identified to estimate the effects of the various risk factors on the prevalence or incidence of the disease in the population. Estimating the factual situation in the population and quantifying the uncertainty in the estimates are thus important aims of such statistical analyses. Having done so, a natural next question of importance is what kind of exposures can be avoided, or how many diseased cases can be prevented, if such exposure could be completely or partially eliminated. A statistical concept that can be used to quantify this is the attributable fraction. For a single disease caused by a single exposure the attributable fraction due to this factor is the proportion of diseased subjects that could have been prevented if the specified exposure had not been present. Or, in other words, one questions what would the proportionate reduction in diseased subjects in the population be if the exposure distribution had been different from what it actual is? For illustration, an Italian study estimated that 15.0 % of the breast cancer cases might have been avoided if the beta-carotene intake had been increased to at least 3366 μg/day for everyone while not changing the distribution of a number of other risk factors (low vitamin E intake, residence, alcohol habits, physical activity, age, educational level, calorie intake and menopausal status). Increasing also vitamin E intake (to at least 8.5 mg /day) for all

subjects gave a combined attributable fraction per cent of 21.5%. Sometimes eliminating a common exposure with a moderate increased risk of disease may have the same effect in the population at large as eliminating a rare exposure with a highly increased risk of disease. Thus, an attributable fraction depends both on the risk of disease if exposed and the extent of the exposure in the population studied.

In general, the attributable fraction quantifies the proportion of cases prevented if the factual exposure distribution were replaced with a hypothetical, so called counterfactual, exposure distribution. The attributable fraction can also be crudely defined as excess proportion of diseased in the population relative to the total proportion. The attributable fraction has also several other applications, e.g. to quantify the proportion of diseased that can be ascribed to one or more exposures (epidemiology), to predict the effect of planned preventive interventions (health policy) and to apportion the responsibility for the disease to various agents responsible for the exposure (liability law). It has been used in regional and national research, as well as in global studies like the Global Burden of Disease and Comparative Risk Assessment projects of the World Health Organization.

Results

With multiple risk factors attributable fractions can be defined in many ways depending on how the counterfactual situation is hypothesized. This thesis describes how attributable fractions can be defined, interpreted and estimated for various scenarios, e.g. one factor is eliminated while the rest is kept fixed; several factors are eliminated; and, multiple factors are removed sequentially may be in different orderings. It also describes convenient graphical methods to illustrate the potential impact on disease load in a population from interventions on one or more risk factors. The statistical and graphical methodology is potentially useful as tools in health policy discussions illustrating possible effects of different preventive strategies under evaluation and may ease the communication between researchers, decision takers and the public. Which strategy will have the largest effect in a public health perspective? Which factors should be given priority in a public health intervention or in

legislation? How much can be achieved by changing personal habits versus general prevention of environmental exposure locally, nationally or globally? Methodology for computerized, and possibly interactive, manipulations of different scenarios is developed to depict the estimates of possible consequences.

The statistical methodology for attributable fractions has traditionally been developed in relation to the classical epidemiological research designs like case-control studies, cross-sectional studies and cohort studies with fixed time to follow-up. Based on the statistical models for analysing time-to-event data the thesis extends and reformulates the traditional definitions of attributable fraction so as to apply also for scenarios where the risk of disease in the population is developing through time and actions against harmful exposure or treatment or other intervention may be implemented at different time points. Thus immediate, later, as well as cumulative effects of an intervention on the disease load in the population are incorporated in these new attributable fraction concepts.

Conclusions

In summary, the thesis discusses many types of attributable fractions to be used for various purposes. The thesis provides methodology for making adequate choices for the question at hand. It also gives new algorithms for calculating attributable fractions extending those of standard statistical software, and it suggests graphical displays that are useful for communicating research results concerning attributable fractions, most of which are not found in standard statistical software of today. Finally, new methodology for dynamic modelling of attributable fractions taking time to disease, time of intervention, or other time-dynamic aspects, into account is suggested by relating the methodology of attributable fractions to established theory of survival analysis. The latter will be an interesting field for further methodological research as will also relating the concepts of attributable fraction to the recent development in causal statistical modelling.

List of articles

I.   Eide, Geir Egil & Gefeller, Olaf (1995) "Sequential and average attributable fractions as aids in the selection of preventive strategies." Journal of Clinical Epidemiology 48, 5, 645-655.

II.   Eide, Geir Egil & Heuch, Ivar (2001) "Attributable fractions: fundamental concepts and their visualization." Statistical Methods in Medical Research 10, 3, 159-193.

III.   Eide, Geir Egil & Heuch, Ivar (2006) "A scaled sample space cube used to illustrate attributable fractions." Biometrical Journal 48, 1, 93-104.

IV.   Eide, Geir Egil & Heuch, Ivar (2006) "Average attributable fractions: a coherent theory for apportioning excess risk to individual risk factors and subpopulations." Biometrical Journal 48, 5, 820-837.

V.   Samuelsen, Sven Ove & Eide, Geir Egil (2007) "Attributable fractions with survival data". Statistics in Medicine. Published Online: 10 Aug 2007 DOI: 10.1002/sim.3022.

Contents

1. Introduction

The wish to quantify the number of the observed cases of diseased or deceased that can be ascribed to a given exposure or risk factor has always been fundamental to biomedical research, and the investigation of potential cause-effect relationships has been the ultimate goal. Levin (1953) seems to be the first to have published a measure for this based on probabilistic considerations. Levin's interest was in quantifying the proportion of lung cancer cases in the population that could be ascribed to smoking. Another later, but early, example is Oftedal et al (1968) who were interested in quantifying the impact of radiation on the number of leukaemia cases. While Levin formulated his measure in terms of the relative risk and the probability of exposure, MacMahon & Pugh (1970) proposed an alternative formulation in terms of the total risk of disease and the risk of disease in the unexposed. The two formulations, that were proved to be algebraically equivalent by Leviton (1973), constitute the classical definitions of the **attributable fraction** (AF), and are given in probabilistic notation as follows:

Levin's formula

$$\lambda = \frac{(RR-1)P(E)}{(RR-1)P(E)+1} \qquad (1)$$

MacMahon and Pugh's formula

$$\lambda = \frac{P(D)-P(D\mid\overline{E})}{P(D)} \qquad (2)$$

Here D denotes the event of disease, E the event of exposure, and over-line complementary events, while the relative risk is RR = P(D|E)/P(D|Ē). For instance, D might be getting or having leukaemia and E being exposed to radiation in some well-defined meaning.

Maximum likelihood estimators (MLE) are easily obtained by substituting probabilities by the corresponding sample proportions. (Detailed proofs can be found

in Eide (1991a).) Thus Levin's formula is suitable when the relative risk can be estimated like in a case-control study, while MacMahon and Pugh's formula is more convenient for a cross-sectional or a one-sample cohort study.

Also alternative formulations have been proposed to optimally serve different sampling designs and a number of these were proved in the appendices of Eide (1991a). A new formulation useful when only exposure probabilities are available was given by Eide and Heuch (2007).

Miettinen (1974) formulated the **attributable fraction in exposed** (AFE) confining the proportion to only those exposed (i.e. all subjects with the event E) and not the total population. The formulations à moduli MacMahon & Pugh (1970) and Levin (1953), respectively, are given as

$$\lambda_E = \frac{P(D \mid E) - P(D \mid \overline{E})}{P(D \mid E)} \qquad (3)$$

and $$\lambda_E = 1 - RR^{-1}. \qquad (4)$$

Miettinen (1974) also related AF to AFE by the equation

$$\lambda = \lambda_E P(E \mid D). \qquad (5)$$

For the AF, Walter (1975; 1976 & 1978) developed asymptotic distributions for the MLEs in the cross-sectional, cohort and case-control designs providing approximate standard errors and confidence intervals. For AFE, being just at transformation of the RR, standard errors and confidence intervals are easily obtained by transforming them from the RR-scale.

The theory so far was univariate, describing the total elimination of only one exposure. Thus, the attributable fraction as defined above is considered to be crude, unadjusted or 'marginal'. However, most often the situation is multi-expositional, i.e. there are many factors influencing the probability of disease, and Walter (1980) was the first to discuss this problem in probabilistic terms. Some exposures may be considered to be modifiable, others not, and the adjusted attributable fraction as first

defined by Whittemore (1982), was designed to quantify the effect of removing one exposure while the others remained unchanged. Whittemore (1982 & 1983) also developed the asymptotic distribution of the maximum likelihood estimator from case-control data. Morgenstern & Bursic (1982) suggested the slightly more general concept of 'potential impact fraction' reflecting the possibility of imperfect prevention of exposure.

Moreover, in the multi-factorial case often a multiple logistic model is estimated for the risk of disease and Bruzzi et al (1985) showed how this could be applied to estimate adjusted attributable fractions with case-control data. Benichou & Gail (1989) were the first to apply the delta method to find the asymptotic variance for a model-based adjusted AF with case-control data and Basu & Landis (1995) extended this methodology to cohort and cross-sectional data.

Despite these developments, much confusion prevailed when trying to apportion an excess risk to single exposures in a multi-expositional setting. Some calculated the crude AF for each exposure, and some calculated the AF for each exposure adjusted for the rest. Each method gives AFs for the single exposures involved that might sum to more than the AF for them all and even to more than 1 (Doll & Peto, 1981). Some authors 'normalized' the calculated single factor AFs so that they were forced to sum to this total combined AF (mentioned, but not advocated, by Kjuus et al (1996)). An elegant solution to this problem was, however, first given by Cox Jr (1984, 1985 & 1987) by adapting a principle from game theory (Shapley, 1953). Also, Kruskal (1987a & 1987b) discussed a parallel solution for ranking independent variables in a multiple regression model according to their individual contributions to the total explained variance.

Recently, Rowe et al (2004) provided an updated discussion, however incomplete as it did not mention the attractive solution from game theory.

At the end of the 1980s there was no standard software available for displaying or calculating estimates of attributable fractions of any kind. With regard to graphic presentation Kjuus et al (1986) and Olsen & Kristensen (1988) included some

instructive figures illustrating basic concepts of AFs by using pie charts and risk vs. exposure plots, respectively, thus paving the way for later development of graphic computer routines.

In the classical epidemiologic literature the AF was mostly a static measure giving the proportion of cases at a given time point that could have been prevented by a hypothesized intervention on the exposure distribution. However, an intervention may have immediate, short term and long term effects on the occurrence of a disease and episodes of disease may more or less come and go (chronic diseases). Also, subjects may be of risk for an exposure for shorter or longer periods of time and with varying amount. Thus there was a need for considerable further development of the classical concept of attributable fraction by more dynamic and flexible modelling of the effects of putative preventive interventions also taking time to disease into account. The increasing availability of statistical methods and software for analysing survival data and other types of longitudinal data has further enhanced the need and possibility for such development.

Finally, some words about terminology. In the literature the concept of attributable fraction has had many names, the most prominent being 'attributable risk' and 'etiologic fraction'. The first is often qualified as the 'population attributable risk' as opposed to the 'attributable risk in exposed'. The preference of 'attributable fraction' throughout this thesis is meant to reflect that it may not necessarily quantify a causal effect (*etiology,* from Greek αίτιολογία, is the study of causation) nor is it always a probability in the usual sense (*risk* is often used as a non-technical term for probability of an adverse event) as it cannot be guaranteed to turn out nonnegative in all instants. Also the term 'attributable proportion' has been abandoned for the latter reason. The use of 'attributable fraction' is consistent with its definition in 'A Dictionary of Epidemiology' (Last, 1988).

2. Aims

In general, the work in this thesis has been aiming at further developing the methodology for attributable fractions for use in practical epidemiological research. This methodology should include parameters formally defined in probabilistic terms, estimators of these parameters and inference theory, interpretations that are easily understood, as well as enlightening graphical and numerical displays of results. For the five articles included the more specific aims have been as follows:

1. The aims for Article I (Eide & Gefeller, 1995) were to establish a probabilistic framework for suggesting estimators of attributable fractions from multi-factorial cross-sectional or cohort data, to suggest such an estimator of the adjusted attributable fraction, to delineate its asymptotic standard error, and to implement the methods of Cox Jr (1985) for apportioning the AF to the single risk factors involved.

2. In Article II (Eide & Heuch, 2001) the aims were to generalize the concepts of AF from Article I to include continuous exposure variables and to generalize and formalize the graphical methods implicitly suggested in the first article.

3. In Article III (Eide & Heuch, 2006a) the aim was to further enhance the graphical methodology for displaying excess risk in two dimensions, as well as to extend it to three dimensions by exploiting an idea of combining the two-dimensional scaled Venn-diagram with the so-called Mosaic-plot for displaying association structures between multiple discrete variables. Moreover, an aim was to consolidate the definitions of attributable fractions with both discrete and continuous explanatory variables.

4. In Article IV (Eide & Heuch, 2006b) the aim was to investigate if the methodology of average AF (AAF) developed in Article I for AF in the population also could be applied to AF in the exposed (AFE), and if so, to see if the resulting average AFs in exposed (AAFE) could be fit into one coherent theoretical framework together with the average AF in the whole population.

5. In Article V (Samuelsen & Eide, 2007) the aims were to further develop the concepts of adjusted AF for more dynamic scenarios as described with survival modelling and to suggest reasonable estimators for them.

3. Materials and methods

All datasets applied in the Articles I-IV are published. The first and most prominent dataset is from the Hordaland study of obstructive lung disease (HSOLD) (Bakke et al, 1991) and a part of it can be found in detail in Table 1 of Article I (Eide & Gefeller, 1995). This part is reused in Article II (Eide & Heuch, 2001) as well as in Articles III (Eide & Heuch, 2006a) and IV (Eide & Heuch, 2006b). In Article II (Eide & Heuch, 2001) also a dataset published by Lloyd (1996) was used for illustration. Finally, in Article V (Samuelsen & Eide, 2007) a dataset on hearing-impairment described by Nafstad et al (2002) was applied.

Analyses and programming have been done using various software including Excel, Minitab, BMDP, Maple, S-Plus and Stata.

Besides, standard probability and statistical inference theory has been applied.

4. Results and discussion

Article I (Eide & Gefeller, 1995)

In retrospect, primarily there are possibly three significant areas of results to be found in the article.

First, and most central, was a systematic development of terminology and probability theory concerning the effect on a binary response from hypothesized manipulations of one or several risk factors in a population. The manipulations were either lowering the harmful effect of an exposure or reducing the exposure's extent in the population. Such modifications could be done in a stepwise manner by removing one exposure at a time. In this case **the adjusted AF** quantifies the effect of removing one exposure in step one leaving the rest unchanged. Further removal of one exposure at each step in a pre-specified order leads to the important concept of **sequential attributable fractions** (SAF) for all risk factors in an **ordered stepwise strategy**. Also, the **combined adjusted attributable fraction** arises as the measure of the combined effect of removing several (or all) exposures while adjusting for the remaining. In later work the last sequential AF, quantifying the remaining effect of removing an exposure after having removed all the others first, has also been termed the **extra attributable fraction** (EAF) (Eide & Gefeller, 2000). Not giving an exposure any priority in the set of exposures to be eliminated leads to the concept of **average attributable fraction** (AAF) as the average of the SAFs for this exposure over all possible orderings of the risk factors in the set. Both the SAFs and the AAFs for a set of risk factors exhibit the important property of summing to the combined AF for the set. However, only the AAFs are independent of the ordering and thus give a unique apportioning of the combined, possibly adjusted, attributable fraction to the single risk factors eliminated. Thus, in Article I the arguments of Cox Jr (1985) for using sequential or average attributable fractions for apportioning combined attributable fractions to single exposures were restated in an epidemiological context. The usefulness of these measures for evaluating and choosing between different

preventive strategies was described. Further theoretical justification has been provided by the research group of Gefeller (Land & Gefeller, 1997; Gefeller et al 1998; Land et al 2001a) making the connection to the optimality of the Shapley value in game theory (Shapley, 1953). This group has also extended the concept of AAF to a multiplicative, rather than additive, variant (Land & Gefeller, 1998) as well as a variant for grouped exposure variables (Land et al 2001b). A Bayesian extension has also been suggested in which the different ordered strategies were given weights according to an a priori consensus (Llorca & Delgado-Rodrigues, 2004a & 2004b). Recent evaluations of the different approaches have been provided by Rabe & Gefeller (2006) and Rabe et al (2007).

Second, and what has become a central topic for further development in Articles II and III of this thesis, is the suggested 2-dimensional 'risk vs. exposure plot' which at the same time displays the joint exposure distribution and the risk of diseases, as well as the resulting excess risk (or removable fraction of the disease) as areas of rectangles (Figure 1 in Article I). In addition, the graphical methods included a 'components of AF-plot' (Figure 2 in Article I) and the construction of a simple pie chart for displaying the average attributable fractions (Figure 4 in Article I). The latter was further developed in Article IV. In Article I also a flow-chart-like diagram was developed to summarize all possible removal strategies for a multi-exposed population and the sequential attributable fractions associated to them (Figure 3 in Article I).

Third, in Article I the MLE of the adjusted attributable fraction from a one sample cohort or cross-sectional study was developed, filling in a gap in the literature at the time. Its asymptotic distribution was developed using the delta-method, and, although not stated in the article, this is a generalization of the theory for the unadjusted AF estimator developed by Walter (1976). Finally, in Article I, also estimators of all the different types of attributable fractions based on a multiple logistic regression model were stated. For case-control data, asymptotic theory for this model-based adjusted AF estimator was developed already by Benichou & Gail (1989), while the theory of this estimator for cross-sectional or cohort data was published by Basu & Landis

(1995). Model-based estimators from multiple logistic and Poisson regression models have been programmed in Stata (Brady, 1998) and a brief user guide exists (Eide, 2006).

For small samples, the asymptotic distribution of the MLE of the adjusted AF is skewed suggesting that the variance estimation should be based on a transformed adjusted AF and, may be, also by using the re-sampling techniques of jack-knifing or bootstrapping. A recent paper (Lehnert-Batar et al, 2006) summarized the historical development and compared the coverage of different interval estimators. In conclusion they found that confidence intervals based on the computer intensive methods may be worth considering when estimating the adjusted attributable fraction. Lehnert-Batar (2006) also developed the 'pARtial' Package in R for the computation of estimates from these estimators. Also Cox (2006) gave an updated review of model-based estimators for case-control and cohort studies. He showed that two proposed methods for case-control studies, the empirical (Whittemore, 1982 & 1983) and the model-based (Greenland & Drescher, 1993), are in fact identical which allows for a unified approach accommodating stratified sampling as well. Moreover, he treated the cohort and cross-sectional designs possibly with stratified sampling and gave estimators based on the delta method and showed good equivalence with the bootstrap method.

For the average attributable fraction Grömping & Weimann (2004) developed the asymptotic distribution and a SAS-procedure for its estimation that may facilitate extended use in the future. Also the pARtial Package (Lehnert-Batar, 2006) can compute these and, moreover, bootstrap- and jack-knife-estimates.

A preliminary version of Article I was printed as an internal report (Eide, 1991a) and presented on an international conference (Eide, 1991b).

Article II (Eide & Heuch, 2001)

This article establishes very general formulations of the attributable fraction in a population. The definition of the **generalized attributable fraction** in equation 20 of Article II includes the situations with multiple continuous as well as discrete explanatory variables in an m-dimensional vector, **X**, with the options of modifying their joint distribution, F(**x**), the conditional risk function, p(**x**), or both. Equation 26 in Article II gives the corresponding definition of an **adjusted generalized attributable fraction** where the distribution of one set of adjustment variables, $F_2(\mathbf{x})$ is held fixed while the conditional distribution for the rest, $F_1(\mathbf{x_1}|\mathbf{x_2})$, is modified. From this the theory of sequential and average attributable fraction follows along the lines of Article I.

Also, in Article II the risk-exposure plot of Article I (Eide & Gefeller, 1995) is formalized and termed a 'scaled Venn diagram'. This 'scaled Venn diagram' consists of a unit square representing the sample space where the probability metric at the horizontal scale represents the joint distribution of the explanatory variables and at the vertical scale the response (disease/not disease) distribution. For discrete exposure variables the probabilities of various events are mapped as areas of corresponding rectangles in the unit square thus directly depicting the sizes of diseased and exposed groups in the population. A version with a continuous exposure variable is also described. The problem of generalizing the diagram to the case with more than one continuous explanatory variable has, however, not been resolved. To give an exact and correct visual impression of the joint probabilities the scaled Venn diagram should be displayed as a unit square so that probabilities in the response dimension and explanatory dimension get equal weights. Thus, for instance, figure 39 of Article II is somewhat misleading as the disease probability scale is shorter than the exposure probability scale.

The term 'scaled Venn diagram' was chosen because its construction was inspired by, and reminded the authors of, the classical Venn diagram (John Venn 1834-1923). This term is, however, slightly misleading since the classical Venn diagram is different in some important aspects like having no metric (is scale-less), no outer border for the 'universe', a symmetrical appearance, and in that events (sets) are

illustrated by possibly intersecting circles or ellipses rather than rectangles (Edwards, 2004). A contemporary competitor to Venn's diagram was the 'Lewis Carroll diagram' (Louis Carroll was the pseudonym of Charles Lutwidge Dodgson 1832-98) in which the 'universe' was confined within a square and the events were represented by rectangles. A metrical Venn diagram was suggested by Edwards & Edwards (1992) as a squared diagram enabling visual comparison of expected and observed frequencies in a 2×2×2 contingency table. However, neither these, nor later, developments of the classic Venn diagram (Edwards, 2004) are convenient for illustrating excess risk or attributable fractions. Rather, for discrete exposure variables the 'scaled Venn diagram' is a two-dimensional (2D) Mondrian plot (Theus, 1997) where all combinations of the values for the explanatory variables generate the first dimension and the response variable the other. To also embrace the variant with a continuous explanatory variable, the term **scaled sample space square** (SSSS) may be more appropriate.

In Article II examples of the suggested diagrams are given for cross-sectional as well as case-control data, relaxing the requirement of a unit square.

Article II is the first paper in a thematic issue of Statistical Methods in Medical Research on attributable fractions in epidemiology (Gefeller, 2001). A preliminary version of the article can be found as a research report (Eide, 2000). A fair resume of the theory and some of the graphical methods was recently given by Benichou (2007), although the description of average attributable fractions (p.293, termed 'partial attributable risk') seems somewhat flawed.

Article III (Eide & Heuch, 2006a)

This is a purely graphical investigation where the principles of the SSSS of Article II (Eide & Heuch, 2001) were combined with the mosaic plot (Hartigan & Kleiner, 1981 & 1984) to create a three-dimensional (3D) display of the multivariate association structure within the exposure variables as well as between disease

variable and the exposure variables. Moreover, the constructed **scaled sample space cube** (SSSC) directly depicts the sizes of diseased and not diseased subpopulations in the various exposure groups, making it especially useful for illustrating excess risk given any level and combination of exposures.

The mosaic plot is a version of the spine plot (Unwin et al, 2006) made to graphically disclose possible associations in multidimensional contingency tables (Theus & Lauer, 1999). In the mosaic plot (Hartigan & Kleiner, 1981) the tiles are separated by gaps to improve visual discrimination. This is especially useful with empty or infrequent categories, but is not so convenient when the introduction of a probability metric to the plot is desired. Hofmann (2006) discussed several variations of the mosaic plot including the default, 'same binsize', 'fluctuation' and χ^2 diagrams.

Without gaps the mosaic display is termed a Mondrian diagram (reminding of paintings by the Dutch painter Pieter Cornelis Mondriaan, 1872-1944) (Theus, 1997). Hartigan & Kleiner (1981) also suggested a version with horizontal or vertical dotted lines to show the deviance from the situation with independent variables.

The resulting scaled sample space cube illustrates the joint exposure distribution as a Mondrian diagram in the two first dimensions and the conditional response distribution in the third dimension. Probabilities of the various events appear as volumes of 3D rectangular boxes within the cube. In Article III it was demonstrated how the scaled sample space cube can be used to illustrate excess risk and the potential impact on risk of disease from hypothesized interventions on the exposures in the population. An example was given by Eide, Heuch & Eagan (2002) and is shown in Figure 1. The figure uses data of the Hordaland Study of Obstructive Lung Disease (Eagan et al, 2002) on 11 years incidence of attacks of dyspnoea and illustrates the effect of removing exposures according to all possible ordered preventive strategies (Eide & Gefeller, 1995). The volume of the yellow boxes represents the proportions of disease that might have been prevented. In the first column of cubes the largest yellow volume is found in the second row. It illustrates that the most effective strategy would be to remove smoking first (direct estimated

adjusted AF: 0.297, 95% CI: (0.173, 0.420)). From this there are three options shown by arrows to the second column of cubes, i.e. to the cubes in the first, fourth and fifth row, and it is seen that among these the largest yellow volume is the last which also has an arrow from the female cube in the first column illustrating that female gender is the best to prevent in combination with smoking (adjusted combined AF: 0.441, 95% CI: (0.242, 0.640)). Gender is, however, not a realistic target for a preventive campaign and this strategy is mostly of academic interest. However, one might speculate if female gender in this case could be a proxy for other, possibly modifiable, exposures that are yet to be identified. For the probably more realistic combination, dust/gas and smoking, the combined adjusted AF is 0.344 with 95% CI: (0.201, 0.488) which is illustrated in Figure 1 by the first SSSC in the 2nd column.
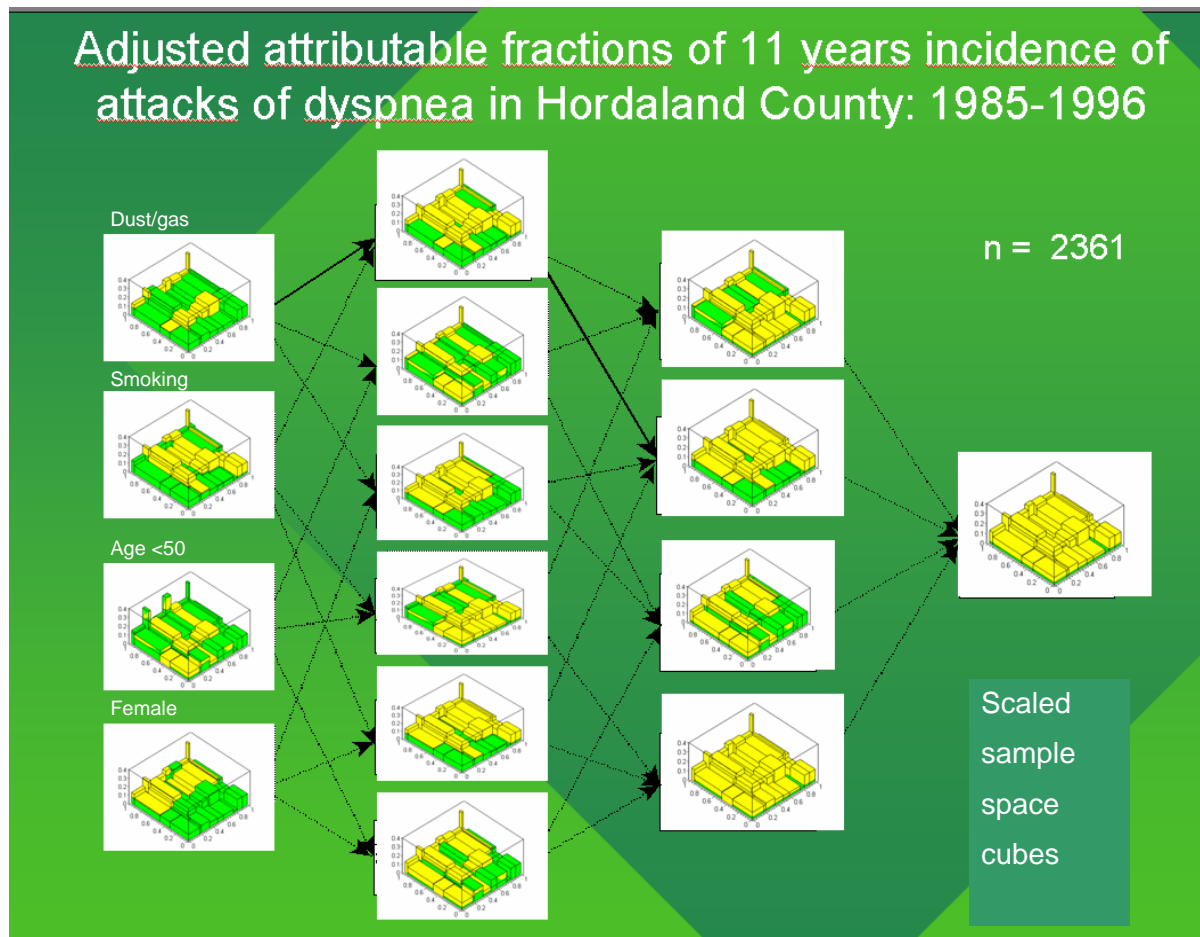

Figure 1 Scaled sample space cubes showing the effects of stepwise removal of exposures; Volumes of yellow boxes are potentially removed excess risks.

 (Notice that changing the incidence in the persons being under 50 years of age in 1985 to that of those at least 50 gave positive attributable fractions as shown in Figure 1, but that choosing below 30 years of age in 1985 as the reference level, as initially done, increased the overall incidence and gave negative attributable fractions for age.)

Also, scaled sample space cubes may be used for comparing the observed and the modelled (estimated) response distribution as shown in Figure 2.
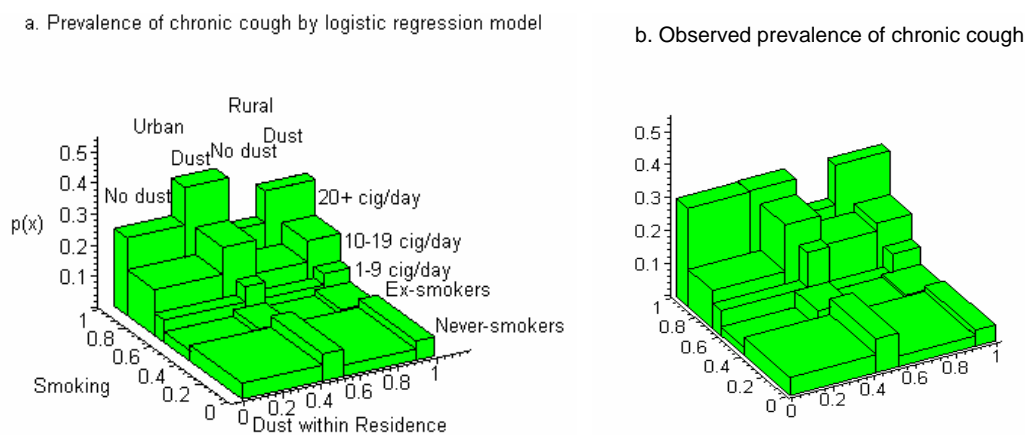
Figure 2 Scaled sample space cubes illustrating modelled (a) and observed (b) probabilities of disease.

The regularity of the scaled sample space cube (here without roof and side walls) for the modelled prevalence (left), which is based on a logistic regression model with no interactions, is striking compared to that for the observed distribution of the same variables (right), and one may identify exposure classes where the discrepancy between model predicted risk and data is large (for instance heavy-smoking, not dust exposed, urban residents). Of course, also the empirical illustration here (right) is a simplification of the observed situation since the explanatory variables included are chosen by a statistical selection procedure before making the plot. In a regression model the response is usually modelled conditionally on the explanatory variables, implicitly doing no smoothing of the joint distribution of the explanatory variables. With categorical explanatory variables further smoothing of the cube could be done by applying the technique of log-linear modelling of multidimensional contingency

tables removing nonsignificant interaction terms between the explanatory variables. Indeed, in a random sample from a population, also the explanatory variables are random variables and their possible dependencies can be tested and estimated, but this is rarely done in practice.

The scaled sample space cube is especially beneficial when used interactively on a computer screen. Rotation and zooming often disclose hidden features of the data or the model. Interactivity is even more important with the volume-based scaled sample space cube than with the area-based scaled sample squares of Article II. In print, however, the advantages of such interactive manipulation are not so apparent, but Figure 3 shows an example of rotating the scaled sample space cube of Figure 1
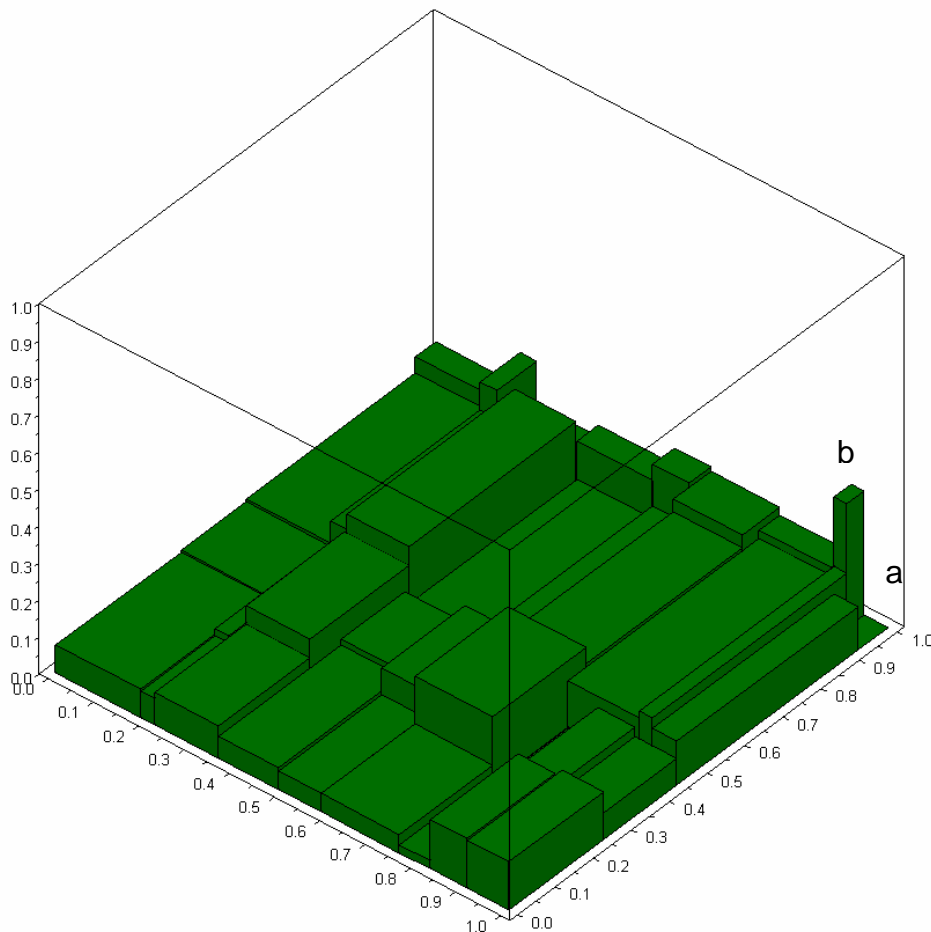

Figure 3 Rotation of scaled sample space cube to better display the extreme risks in the small exposure groups in the back of Figure 1. For females over 50 years there were no cases of breathlessness in the smoking group (a), but a large proportion among the ex-smokers (b).

that discloses the small group of women who were above 50 years and smokers in 1985 but had no incident cases of breathlessness during 1985-1996; a survival of the fittest effect?

An example of the usefulness of zooming is given with the scaled sample space square in Figures 33-38 of Article II (Eide & Heuch, 2001).

Like for the mosaic plot, or the double-decker plot (Hofmann, 2001), plotting the scaled sample space cube for different orderings of the explanatory factors, may disclose information that otherwise might go undetected. Indeed, in the context of attributable fractions this can give directions to more efficient preventive strategies. In the cubes in Figure 1 the orderings of the variables in the mosaic plot in the base is the same for all preventive strategies to make visual comparisons between the different removal strategies easier. When illustrating only one selected preventive strategy one could adopt the convention of having the same order in the base mosaic as in the actual strategy because eliminating one exposure from the mosaic will give the mosaic display corresponding to the remaining exposure factors. Another ambiguity is the ordering of categories within each categorical variable. For an ordinal variable it seems natural to stick to the original ordering, but for a nominal variable the ordering does not come automatically. One reasonable convention would then be to order the categories according to increasing risk of disease, be it hypothesized, observed or estimated.

As of today neither mosaic plots, nor spine plots, Mondrian diagrams, or SSSCs can be found in standard statistical software, although specialized software exists (http://www.rosuda.org/GOLD/Software.html). Indeed, this author could not even find a standard statistical package with the option of creating a univariate histogram with bin widths varying between the categories, nor could the option of a 3D bivariate histogram with irregular baseline grid be found. To incorporate such flexibilities in the plotting routines, for instance drawing a, possibly two-dimensional, histogram with varying pre-selected or data-driven bins, should be a natural challenge for statistical software developers.

Statistical graphics are distinguished from other graphics by their universality (Theus, 2006) and should be valid for any data measured on a nominal, ordinal or continuous scale and not tailored for only one specific application. The 3D scaled sample space cube is such a method that, like a histogram, can be used also for continuous distributions provided that reasonable categorizations, or possibly smoothing, are made. Developing further interactive facilities like querying, selection and linking, and varying the plot characteristics by rescaling, resizing, zooming, reordering or re-colouring (Unwin, 2006) might further enhance its applicability.

A first version of Article III can be found as a Centre for Clinical Research (CCR) Research Report (Eide & Heuch, 2002).

Article IV (Eide & Heuch, 2006b)

From the beginning there has been a duality between AF and AFE and also between attributing risk to variables and to categories. While the theory of adjusted, sequential and average AFs was developed in Articles I and II and illustrated in Articles II and III, a similar development was not done for the AFE although for the unadjusted versions equation (5) gives a simple relationship. In Article IV a probabilistic development of adjusted, sequential and average AFEs is carried out in parallel to that for the AF. In principle then, the AFE for each exposure subclass is partitioned into average AFEs for all the factors responsible for the excess risk of this subclass and by piecing together the AFEs from all the subclasses for one particular factor the average AF in the total population for this factor is restored. Thus, the relationships between average AF concepts for the exposed and for the population are disclosed, and a complete theory for adjusted, sequential and average attributable fractions is established. This does not only resolve the duality between partitioning the AF and AFE to multiple risk factors, but also the duality between attributing risk to variables or to categories. Indeed, the complete decomposition of the AF, as for example given in Table 3 of Article IV, can be visualized in a pie chart putting the emphasis on the

variables or on the categories as reproduced from Eide & Heuch (2004) in Figure 4, panel A and B, respectively.
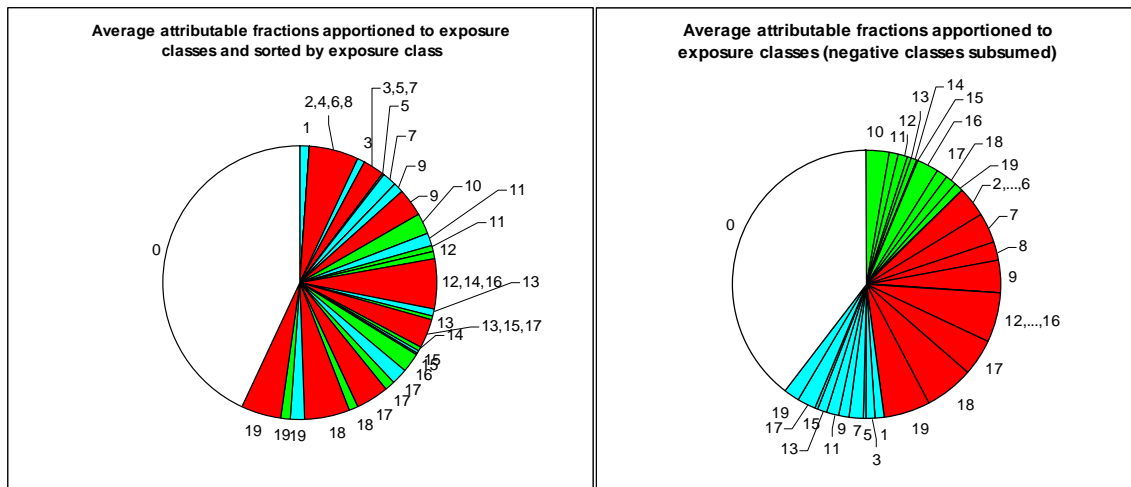
Figure 4 Pie chart showing complete decomposition of total AF = 60.18 % to risk factors and exposure classes. Left panel sorted according to exposure class numbering, right panel according to the risk factors, i.e. residence (green: 12.75%), smoking (red: 34.96%) and occupational exposure to dust or gas (blue: 12.47%).

Finally, it was demonstrated that the methodology of average attributable fractions in exposed has favourable characteristics compared with other methods proposed in the literature and some ad hoc proposals. These results have been substantiated by other recent research (Rabe & Gefeller, 2006; Rabe et al, 2007).

A first version of this article can be found as a CCR Research Report (Eide & Heuch, 2004).

Article V (Samuelsen & Eide, 2007)

A major limitation of the classical concept of attributable fraction is that it has no time dimension. Traditional epidemiology usually concerns relationships between exposure and disease that are causal and a common requisite for an association to be causal is that the individual should be exposed before a disease occurs (Hill, 1965). However, time from exposure to disease is not an ingredient in the definitions of the

attributable fractions that were discussed in the first four articles of this thesis. Moreover, the time from a suggested intervention to its impact on the population studied is traditionally not taken into account when estimating attributable fractions. Thus the attributable fraction is a static measure considering the disease situation at one point in time and what it could have been if the risk distribution at the same time point or some other fixed timed point had been different. The need for a more dynamic attributable fraction measure is obvious. Indeed, from the field of event history analysis (or survival analysis) a set of models is offered that describes evolution of disease in individuals over time. Indeed, such models set out to describe the time from one event (starting time) to another (disease) and estimation procedures are especially developed that account for the feature of censored observation times that is so frequently present in survival data.

Article V set out to transfer the concepts of attributable fraction to the situation with survival data. To accomplish this, three different concepts were proposed: the **attributable hazard fraction** (AHF), the **attributable fraction before** (AFB) **time t**, and the **attributable fraction within study** (AFW). Traditionally one would think of these AFs as the potential effects of removing an exposure at the starting time, that is that some kind of intervention takes place at the intervention time $s = 0$. In Article V this situation was modelled within the time to event analysis framework and further extended to include the possibility of an intervention at a later time point, i.e. at a time s where possibly $s > 0$. Thus, rather general models are formulated.

The different definitions are interpreted and their features analysed and illustrated, as well as how they relate to each other. With intervention at time s all measures are zero at time-points $t < s$. For $t \geq s$ AHF(t,s) is monotonically decreasing and approaches zero, while the effect on the cumulative risk function (1-S(t)) described by AFB(t,s) is increasing continuously from zero to a maximum cumulative effect before decreasing to zero in the long run. The attributable fraction within study is a measure that depends on the censoring scheme for the particular study at hand and may or may not approach zero when maximum follow-up time is approaching infinity.

Also, the bias from the common practice of substituting the adjusted hazard ratio from a Cox proportional hazards model for survival data (Cox, 1972) for the RR in the classical definition of Levin (1953) was demonstrated.

Finally, Article V proposed estimators for the three AF measures and showed how various point-wise confidence intervals can be obtained using the bootstrap methodology. The proposed concepts and their estimators are illustrated using a real data set on age when being granted cash benefit for hearing impairment. Receiving cash benefit here serves as a proxy for developing hearing impairment and the impact on the risk of being hearing impaired preventing harm from low birth weight in several scenarios is estimated.

The main ideas and results of this paper were presented internationally for the first time as an oral contribution at the International Conference on Reliability and Survival Analysis in Columbia, USA (Samuelsen & Eide, 2003).

Some general remarks

As noticed already in Article I (Eide & Gefeller, 1995) the average attributable fraction is an application of a principle described by Kruskal (1987a & 1987b) for determining relative importance of explanatory variables (regressors) in a multiple linear regression model with correlated regressors based on decomposing the determination coefficient, R^2 (Kruskal & Majors, 1989). The background and a summary of further development of this methodology within linear regression modelling have recently been provided by Grömping (2007a). Grömping (2007a), ascribing the principle to Lindeman, Merenda & Gold (1980), denoted it the LMG method and compared it with the method of 'proportional marginal variance decomposition' (PMVD) proposed by Feldman (2005). The PMVD is shown to be a weighted averaging method, as opposed to the LMG method where all rankings of regressors have equal weight, and is shown to fulfil four postulated desirable requirements while the LMG method fulfils three of them. The four requirements,

restated from Grömping (2007a), are: a) 'proper decomposition': the model variance is to be decomposed into shares, that is, the sum of all shares has to be the model variance, b) 'non-negativity': all shares have to be non-negative, c) 'exclusion': the share allocated to a regressor X_j with $\beta_j = 0$ should be 0, and d) 'inclusion': a regressor X_j with $\beta_j \neq 0$ should receive a nonzero share. Requirements a) and c) is quite similar to two of the three axioms of Cox Jr (1985) which were informally stated as A1 and A3, respectively, in Article I. According to Grömping (2007a) the LMG method does not fulfil the 'exclusion'-requirement. However, examples of causal models are given where a regressor X_j has $\beta_j = 0$ in the regression model but still is a causal factor and so it is not reasonable to assign it a zero share. Consequently, Grömping (2007a) recommended the LMG method unless there are specific reasons to exclude an X_j with $\beta_j = 0$ from the causal model. Further discussion can be found in Menard (2007) and Grömping (2007b).

The LMG method does neither seem to be developed for regression models with discrete regressors on a nominal (or ordinal) scale represented by a set of indicator variables in the model, nor for models with interaction terms $(X_i \cdot X_j)$. Although both indicators and multiplicative terms can be accommodated within the general framework of the LMG method, the interpretation of the assigned shares for these becomes obscure. For the AAF, however, the theory for such cases is thoroughly developed in Article IV. A similar development for the LMG method could be warranted.

An important difference between the use of AF in epidemiology and the general use of the determination coefficient, R^2, in linear regression is that while R^2 is a measure of the proportion of the total variation around the expected value of the response variable explained by variation in the explanatory variables, the AF is a measure of the proportion of the total risk explained by raised levels of exposure (proportionate excess risk). Thus, to transfer the ideas of the AF to the situation with a linear regression model one might define an R^2 using a chosen reference value for the response representing the expected (normal) level in an unexposed population. In a linear regression model with the continuous explanatory variables normalised to have

$E(X_j) = 0$ as the 'normal' value and categorical variables represented by indicators of exposure, this reference value would be $\mu_0 = E(Y|\mathbf{X=0})$, the total variation would be $SST_0 = \Sigma_i(Y_i-\mu_0)^2$ and the explained variation $SSR_0 = \Sigma_i(\mu_i-\mu_0)^2$ where $\mu_i = E(Y|\mathbf{x_i})$, so that the attributable squared variation becomes $AV = SSR_0/SST_0$. More generally, in a linear regression model one could define the attributable squared variation (AV) due to a specified change in the joint distribution of the regressors in the same way as for the generalized attributable fraction. The AV will coincide with the usual R^2 when μ_0 is the mean response. Further development of this theory could be interesting, but are not found in the literature as of today.

In the effort of understanding the various kinds of attributable fractions discussed in this thesis graphical illustrations have been of great help. A number of displays have been suggested, especially in Articles II and III, and one might hope that some of these can find their way to standard statistical software. Both 2D and 3D visualizations are possible and can supplement each other in a graphical menu for attributable fractions. Modern ways of interactively manipulating graphs on the computer screen like querying, selection and linking, and varying the plot characteristics by rescaling, resizing, zooming, reordering or re-colouring could be implemented. Moreover, a possibility of flexible choice of preventable exposures and their ranking in a preventive strategy would certainly aid in understanding more complex modelling and estimation of attributable fractions and in discussing pros and cons for various preventive strategies of interest.

Taking into account time from exposure or intervention to effect on disease risk, as suggested in Article V, is a major step forward in the development of attributable fractions and also other authors have now taken up this idea (Chen et al, 2006). Further development in this direction will be natural by linking the attributable fraction to even more realistic intervention scenarios and to causal models (Hernán & Robins, 2006). Survival models with time-dependent effects and covariates will be central and one possibility is intervention at a fixed time-point s that has, however, an impact after a stochastic latency period L. Also, introducing an intervention at a fixed date, d, may induce a random intervention time V in the model. For instance, if the

accrual of individuals in the study is uniformly distributed over a given calendar period (a,b) the intervention time V will also be uniformly distributed, i.e. on (a+d,b+d), and combined with a latency period an effect of intervention might not be observed until a random time V+L after time zero. Thus the attributable fractions should be defined accordingly. One can also foresee to study a series of interventions at successive dates or time-points and to evaluate the effects of different orderings or selections of interventions in order to obtain maximum preventive effects after minimum elapse of time.

In survival analysis there is also the problem of selection bias in the meaning that the more fit patients are the best survivors so that the comparison of hazard functions is gradually more biased as time increases. To remedy this problem some authors have proposed that comparing expected survival times, E(T), is more appropriate. In this perspective one might define 'attributable fraction of lost life time' in a population (AF(T)) due to an exposure as

$$AF(T) = \frac{E(T) - E(T \mid \overline{E})}{E(T)} \qquad . \qquad (6)$$

Further development of theory for estimating this or similar parameters is in demand.

5. Main conclusions

This investigation describes reasonable and practical methods for apportioning excess risk to different risk factors as well as to groups of risk factors and to subpopulations identified by given risk profiles. For the attributable fraction in the classical sense this is developed in Articles I, II and IV and the synthesis is given in Article IV.

Secondly, a graphical methodology for describing excess risk attributable to different exposure groups in a multi-expositional setting has been developed. This methodology provides means for dynamic illustrations of the effects on the risk of disease in a population of stepwise removing exposures in an ordered preventive strategy. It also paves the way for making simple pie charts quantifying the portions of the total risk to different exposures or risk profiles. The methodology is described in Articles II and III, and a synthesis is given in Article IV.

Finally, Article V points at how the classic concept of attributable fraction from epidemiology can be further developed from being a static measure to being used with dynamic scenarios encountering risks developing in time or effects of preventive interventions on risk changing over time.

In summary, the practice of simple and crude calculations of attributable fractions in epidemiology should be abandoned. Modern statistical software should be upgraded to be able to illustrate the different kinds of attributable fractions as well as with procedures for estimating the more advanced types. Finally, further development of the statistical theory for attributable fractions with survival data and possibly linking it to causal modelling will be of great interest.

6. Errata

Article I (Eide & Gefeller, 1995)

> Eide, Geir Egil & Gefeller, Olaf (1995) "Sequential and average attributable fractions as aids in the selection of preventive strategies." Journal of Clinical Epidemiology 48, 5, 645-655.

p. 649: Subscripts "c" and "s" in the logistic regression formulas should be "C" (for Chronic cough) and "B" (for Breathlessness), respectively.

Table 1: For stratum 1, exposure class 9, the predicted #yes should be 36.71, not 30.00.

Table 1: The average attributable fraction in the lower rightmost corner should be 0.512 (or 51.2%), not 5.12.

Fig. 4: The sector labels "Residence" and "Occupational exposure" should be interchanged so that "Residence" gives 13 % and "Occupational exposure" 12 %.

p.655: In the definition of θ 'θ_{11K}' should be 'θ_{22S}' and '4K' should be '4S'.

p.655: The line with the expectation and the covariance matrix should read:

$E(\mathbf{X}) = n\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}_{\mathbf{X}} = n[\text{diag}(\boldsymbol{\theta}) - \boldsymbol{\theta}\boldsymbol{\theta}']$.

Article II (Eide & Heuch, 2001)

> Eide, Geir Egil & Heuch, Ivar (2001) "Attributable fractions: fundamental concepts and their visualization." Statistical Methods in Medical Research 10, 3, 159-193.

Fig. 4: The heading 'V' should be 'Ω' and '12P*(E) should be '1-P*(E)' at the q-axis.

Fig. 26: The colour codes under the figure should be interchanged so that eliminating E first is the darker.

Article III (Eide & Heuch, 2006a)

Eide, Geir Egil & Heuch, Ivar (2006a) "A scaled sample space cube used to illustrate attributable fractions." Biometrical Journal 48, 1, 93-104.

p. 98: In the equation, the sign of 0.37126 should be negative (-), not positive (+).

Fig. 8d: This should not be equal to Fig. 8c, but look like
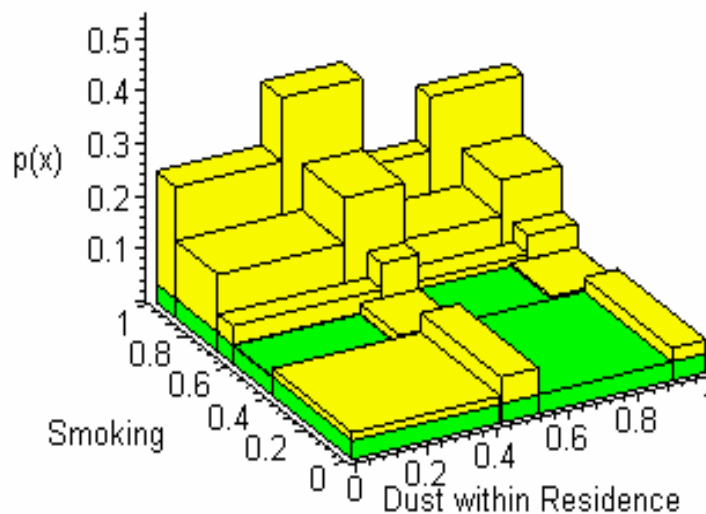
as correctly shown in Animation 1 on the website

http://www.helse-bergen.no/avd/kkf/statistiskforskning/cube.htm.

p.104: The page numbers for the reference of Uter & Pfahlberg (2001) should be 231-237, not 231-231.

Article IV (Eide & Heuch, 2006b)

Eide, Geir Egil & Heuch, Ivar (2006b) "Average attributable fractions: a coherent theory for apportioning excess risk to individual risk factors and subpopulations." Biometrical Journal 48, 5, 820-837.

p. 826: Line 13: '[1]' should be '(Eide and Gefeller, 1995)'.

Article V (Samuelsen & Eide, 2007)

> Samuelsen, Sven Ove & Eide, Geir Egil (2007) "Attributable fractions with survival data". Statistics in Medicine. Published Online: 10 Aug 2007 DOI: 10.1002/sim.3022.

Section 4.3, paragraph 2, line 2: Delete 'are'.

Section 5.1, paragraph 1, line 2: Insert 'for' in '… to use for covariates …'.

Chapter 6, paragraph 2, line 4: Delete first occurrence of 'other'.

Chapter 7, paragraph 1, line 6-7 should read: 'That the time perspective has not been taken into account has been typical for the use of AFs so far.'

Chapter 7, paragraph 2, line 3: Insert a comma after 'dates'.

Chapter 7, paragraph 5, line 15: Replace 'the models' with 'their'.

Reference 38: Insert 'Research' in 'Centre for Clinical Research Research Report'.

7. References (alphabetical)

Bakke P, Eide GE, Hanoa R & Gulsvik A (1991) Occupational dust or gas exposure and prevalences of respiratory symptoms and asthma in a general population. **European Respiratory Journal** 4, 3, 273-278.

Basu S & Landis JR (1995) Model-based estimation of population attributable risk under cross-sectional sampling. **American Journal of Epidemiology** 14, 12, 1338-1343.

Benichou J & Gail MH (1989) A delta method for implicitly defined random variables. **The American Statistician** 43, 1, 41-44.

Benichou J (2007) Biostatistics and epidemiology: measuring the risk attributable to an environmental or genetic factor. **Comptes Rendus Biologies** 330, 4, 281-298.

Brady AR (1998) sbe21: Adjusted population attributable fractions from logistic regression. **Stata Technical Bulletin** 42 (STB-42), 8-12.

Bruzzi P, Green SB, Byar DP, Brinton LA & Schairer C (1985) Estimating the population attributable risk for multiple risk factors using case-control data. **American Journal of Epidemiology** 122, 5, 904-914.

Chen YQ, Hu C & Wang Y (2006) Attributable risk function in the proportional hazards model for censored time-to-event. **Biostatistics** 7, 4, 515-529.

Cox C (2006) Model-based estimation of the attributable risk in case-control and cohort studies. **Statistical Methods in Medical Research** 15, 6, 611-625.

Cox DR (1972) Regression models and life-tables. **Journal of the Royal Statistical Society, Series B** 34, 2, 187-202.

Cox Jr. LA (1984) Probability of causation and the attributable proportion of risk. **Risk Analysis** 4, 3, 221-230.

Cox Jr. LA (1985) A new measure of attributable risk for public health applications. **Management Science** 31, 7, 800-813.

Cox Jr. LA (1987) Statistical issues in the estimation of assigned shares for carcinogenesis liability**. Risk Analysis** 7, 1, 71-78.

Doll R & Peto R (1981) The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. **Journal of the National Cancer Institute** 66, 6, 1191-1308.

Eagan TML, Gulsvik, A, Eide GE & Bakke, PS (2002) Occupational airborne exposure and the incidence of respiratory symptoms and asthma. **American Journal of Respiratory and Critical Care Medicine** 166, 7, 933-938.

Edwards AWF (2004) *Cogwheels of the Mind. The Story of Venn Diagrams.* The John Hopkins University Press, Baltimore and London 2004.

Edwards AWF & Edwards JH (1992) Metrical Venn diagrams. **Annals of Human Genetics** 56, 71-75.

Eide GE (1991a) The selection of preventive strategy based on stepwise calculation of population attributable fractions from a multivariate regression model for a binary response. Internal report. **Institute of Mathematics and Statistics, Norwegian School of Economics and Business Administration & Section for Medical Informatics and Statistics, University of Bergen**, 1991-09-19, 50 pages.

Eide GE (1991b) Strategies for stepwise estimation of population attributable fractions. Poster at the **Joint Meeting of the International Society for Clinical Biostatistics and Society for Clinical Trials**, Brussels, July 8-12, 1991.

Eide GE (2000) Fundamental concepts of attributable fractions and their visualization. **Centre for Clinical Research Research Report** 2000-01. Haukeland Hospital, Bergen.

Eide GE (2006) How to estimate attributable fractions in Stata: A simple introduction. **Centre for Clinical Research Research Report** 2006-04. Haukeland University Hospital, Bergen.

Eide GE & Gefeller O (1995) Sequential and average attributable fractions as aids in the selection of preventive strategies. **Journal of Clinical Epidemiology** 48, 5, 645-655.

Eide GE & Gefeller O (2000) Re: Attributable risk in practice. [Letter] **American Journal of Epidemiology** 151, 2, 208-209.

Eide GE & Heuch I (2001) Attributable fractions: fundamental concepts and their visualization. **Statistical Methods in Medical Research** 10, 3, 159-193.

Eide GE & Heuch I (2002) Displaying attributable fractions in three dimensions using the scaled Venn cube. **Centre for Clinical Research Research Report** 2002-01. Haukeland University Hospital, Bergen.

Eide GE & Heuch I (2004) Average attributable fractions: a coherent theory for apportioning attributable fractions in populations and exposed to subpopulations and individual risk factors. **Centre for Clinical Research Research Report** 2004-01. Haukeland University Hospital, Bergen.

Eide GE & Heuch I (2006a) A scaled sample space cube used to illustrate attributable fractions. **Biometrical Journal** 48, 1, 93-104.

Eide GE & Heuch I (2006b) Average attributable fractions: a coherent theory for apportioning excess risk to individual risk factors and subpopulations. **Biometrical Journa**l 48, 5, 820-837.

Eide GE & Heuch I (2007) Computation of attributable fractions on the basis of exposure probabilities. Commentary. **Sleep** 30, 4, 386-386.

Eide GE, Heuch I & Eagan TML (2002) The scaled Venn cube for displaying attributable fractions. In: **Proceedings of the 17th International Workshop on Statistical Modelling** 215-226, Chania, Greece, July 2002, Eds: M Stasinopoulos & G Touloumi.

Feldman B (2005) Relative importance value. Preliminary manuscript March 19, 2005. Available online at **http://prismanalytics.com/docs/RelativeImportance050319.pdf.**

Gefeller O (2001) The concept of attributable risk in epidemiology: yesterday, today and tomorrow. Editorial. **Statistical Methods in Medical Research** 10, 3, 157-158.

Gefeller O, Land M & Eide GE (1998) Second thoughts: averaging attributable fractions in the multifactorial situation: assumptions and interpretation. **Journal of Clinical Epidemiology** 51, 5, 437-441.

Greenland S. & Drescher, K (1993) Maximum likelihood estimation of the attributable fraction from logistic models. **Biometrics** 49, 3, 865-872.

Grömping U (2007a) Estimators of relative importance in linear regression based on variance decomposition. **The American Statistician** 61, 2, 139-147.

Grömping U (2007b) Response to: Menard S. (2007) "Two simple estimators of relative importance in linear regression based on variance decomposition." Letter to the editor. **The American Statistician** 61, 3, 282-283.

Grömping U & Weimann U (2004) The asymptotic distribution of the partial attributable risk in cross-sectional studies. **Statistics** 38, 5, 427-438.

Hartigan JA & Kleiner B (1981) Mosaics for contingency tables. **13th Symposium on the Interface**, Springer Verlag, New York, 268-273.

Hartigan JA & Kleiner B (1984) A mosaic of television ratings. **The American Statistician** 38, 1, 32-35.

Hernán MA & Robins JM (2006) Estimating causal effects from epidemiological data. **Journal of Epidemiology and Community Health** 60, 578-586.

Hofmann H (2001) Generalized odds ratios for visual modeling. **Journal of Computational and Graphical Statistics** 10, 4, 628-640.

Hofmann H (2006) Multivariate categorical data – mosaic plots. Ch. 5 in: **Graphics of Large Datasets. Visualizing a Million.** Springer, Singapore 2006, p.107-141.

Kjuus H, Hauge OA, Kongerud J & Aasen TB (1996) Vurdering av årsaksforhold ved yrkesrelaterte lungesykdommer. (In Norwegian). **Tidsskrift for den Norske Lægeforening** 116, 6, 736-738.

Kjuus H, Langård S & Skjærven R (1986) A case-referent study of lung cancer, occupational exposures and smoking - III Etiologic fraction of occupational exposures. **Scandinavian Journal of Work, Environment and Health** 12, 3, 210-215.

Kruskal W (1987a) Relative importance by averaging over orderings. **The American Statistician** 41, 1, 6-10.

Kruskal W (1987b) Correction. **The American Statistician** 41, 4, 341-341.

Kruskal W & Majors R (1989) Concepts of relative importance in recent scientific literature. **The American Statistician** 43, 1, 2-6.

Land M & Gefeller O (1997) A game-theoretic approach to partitioning attributable risks in epidemiology. **Biometrical Journal** 39, 7, 777-792.

Land M & Gefeller O (1998) A multiplicative approach to partitioning the risk of disease. In: I Balderjahn, R Mathar, M Schader (eds.): **"Classification, Data Analysis and Data Highways"**, 73-80. Berlin: Springer-Verlag.

Land M, Vogel C & Gefeller O (2001a) Partitioning methods for multifactorial risk attribution. **Statistical Methods in Medical Research** 10, 3, 217-230.

Land M, Vogel C & Gefeller O (2001b) A multifactorial variant of the attributable risk for groups of exposure variables. **Biometrical Journal** 43, 4, 461-481.

Last JM (ed.) (1988) *A Dictionary of Epidemiology.* 2nd edition. Oxford University Press, New York 1988.

Lehnert-Batar A (2006) **The pARtial Package**. Version 0.1 October 16, 2006, pp 11.

Lehnert-Batar A, Pfahlberg A & Gefeller O (2006) Comparison of confidence intervals for adjusted attributable risk estimates under multinomial sampling. **Biometrical Journal** 48, 5, 805-819.

Levin ML (1953) The occurrence of lung cancer in man. **Acta Unio Internationalis Contra Cancrum** 9, 3, 531-541.

Leviton A (1973) Definitions of attributable risk. Letter to the editor. **American Journal of Epidemiology** 98, 3, 231-231.

Lindeman RH, Merenda PF & Gold RZ (1980) *Introduction to Bivariate and Multivariate Analyses.* Glenview, IL: Scott, Foresman.

Llorca J & Delgado-Rodríguez M (2004a) A new way to estimate the contribution of a risk factor in populations avoided nonadditivity. **Journal of Clinical Epidemiology** 57, 5, 479-483.

Llorca J & Delgado-Rodríguez M (2004b) Erratum. **Journal of Clinical Epidemiology** 57, 8, 873-873.

Lloyd CJ (1996) Estimating attributable response as a function of a continuous risk factor. **Biometrika** 83, 3, 563-573.

MacMahon B & Pugh TF (1970) *Epidemiology: Principles and Methods*. Little, Brown and Company, Boston.

Miettinen OS (1974) Proportion of disease caused or prevented by a given exposure, trait or intervention. **American Journal of Epidemiology** 99, 5, 325-332.

Menard S (2007) Two simple estimators of relative importance in linear regression based on variance decomposition. Comment to Grömping, U. (2007) Estimators of relative importance in linear regression based on variance decomposition. Letter to de editor. **The American Statistician** 61, 3, 281-282.

Morgenstern H & Bursic E (1982) A method for using epidemiologic data to estimate the potential impact of an intervention on the health status of a target population. **Journal of Community Health** 7, 4, 292-309.

Nafstad P, Samuelsen SO, Irgens LM & Bjerkedal T (2002) Birth weight and hearing impairment in Norwegians born from 1967 to 1993. **Pediatrics** 110, 3, Art.No.e30.

Oftedal P, Magnus K & Hvinden T (1968) On the probability of radiation being the cause of leukemia. **British Journal of Radiology** 41, 489, 711-712.

Olsen O & Kristensen TS (1988) Den ætiologiske fraktion - definition og teori. I **Hjerte/karsygdomme og arbejdsmiljø, Bind 3**. Hvor stor betydning har arbejdsmiljøet for hjerte/karsygdomme i Danmark? 39-188.

Rabe C & Gefeller O (2006) The attributable risk in a multifactorial situation. Evaluation of different methods of partitioning. **Methods of Information in Medicine** 45, 4, 404-408.

Rabe C, Lehnert-Batar A & Gefeller O (2007) Generalized approaches to partitioning the attributable risk of interacting risk factors can remedy existing pitfalls. **Journal of Clinical Epidemiology** 60, 5, 461-468.

Rowe AK, Powell KE & Flanders D (2004) Why population attributable fractions can sum to more than one. **American Journal of Preventive Medicine** 26, 3, 243-249.

Samuelsen SO & Eide GE (2003) Attributable fractions for survival data. Oral presentation at **International Conference on Reliability and Survival Analysis (ICRSA)**, Columbia, South Carolina May 23, 2003.

Samuelsen SO & Eide GE (2007) Attributable fractions with survival data. **Statistics in Medicine**. Published Online: 10 Aug 2007 DOI: 10.1002/sim.3022.

Shapley LS (1953) A value for n-person games. **Annals for Mathematics Studies** 28, 307-317.

Theus M (1997) Visualizing categorical data. In: Bandilla, W. & Faulbaum, F. (eds.), **Softstat '97: Advances in Statistical Software** 6, Lucius and Lucius, Stuttgart, 47-55.

Theus M (2006) Statistical graphics. Ch. 2 in: **Graphics of Large Datasets. Visualizing a Million.** Springer, Singapore 2006, 31-54.

Theus M & Lauer SRW (1999) Visualizing loglinear models. **Journal of Computational and Graphical Statistics** 8, 3, 396-412.

Unwin A (2006) Interacting with graphics. Ch. 4 in: **Graphics of Large Datasets. Visualizing a Million.** Springer, Singapore 2006, 73-101.

Unwin A, Theus M & Hofmann H (2006) *Graphics of Large Datasets. Visualizing a Million*. Springer, Singapore 2006.

Walter SD (1975) The distribution of Levin's measure of attributable risk. **Biometrika** 62, 2, 371-374.

Walter SD (1976) The estimation and interpretation of attributable risk in health research. **Biometrics** 32, 4, 829-849.

Walter SD (1978) Calculation of attributable risks from epidemiological data. **International Journal of Epidemiology** 7, 2, 175-182.

Walter SD (1980) Prevention for multifactorial disease. **American Journal of Epidemiology** 112, 3, 409-416.

Whittemore AS (1982) Statistical methods for estimating attributable risk from retrospective data. **Statistics in Medicine** 1, 3, 229-243.

Whittemore AS (1983) Estimating attributable risk from case-control studies. **American Journal of Epidemiology** 117, 1, 76-85.

8. Appendices

I.   The mission (by Per S. Bakke).

II.   References in chronological order.

III.   English – Norwegian nomenclature.

Appendix I: The mission (by Per S. Bakke)

Mottatt 880913

Geir

Kan du tenke på:

1) hvordan beregne konfidens intervaller for population attributable risk (PAR)?

2) hvordan beregne PAR ved interaksjon mellom ulike risikofaktorer?

3) hvordan beregne PAR etter at har fjernet den del av en symptomforekomst som tilskrives en annen risiko faktor?
med andre ord: hvis PAR for yrkesstøv eksposisjon på forekomst av luftveissympt er 20% og PAR for sigarett røyking på forekomst av luftveissympt er 60%. Hva er PAR for yrkesstøv på forekomst av luftveissympt gitt at effekten av sigarettrøyking på forekomsten av luftveissympt er fjernet?

Per

Appendix II: References in chronological order

Levin M.L. (1953) The occurrence of lung cancer in man. **Acta Unio Internationalis Contra Cancrum** 9, 3, 531-541.

Shapley, L.S. (1953) A value for n-person games. **Annals for Mathematics Studies** 28, 307-317.

Oftedal, P.; Magnus, K. & Hvinden, T. (1968) On the probability of radiation being the cause of leukemia. **British Journal of Radiology** 41, 489, 711-712.

MacMahon, B. & Pugh, T. F. (1970) *Epidemiology: Principles and Methods*. Little, Brown and Company, Boston.

Cox, D. R. (1972) Regression Models and Life-Tables. **Journal of the Royal Statistical Society, Series B** 34, 2, 187-202.

Leviton, Alan (1973) Definitions of attributable risk. Letter to the editor. **American Journal of Epidemiology** 98, 3, 231-231.

Miettinen, O. S. (1974) Proportion of disease caused or prevented by a given exposure, trait or intervention. **American Journal of Epidemiology** 99, 5, 325-332.

Walter, S. D. (1975) The distribution of Levin's measure of attributable risk. **Biometrika** 62, 2, 371-374.

Walter, S. D. (1976) The estimation and interpretation of attributable risk in health research. **Biometrics** 32, 4, 829-849.

Walter, S. D. (1978) Calculation of Attributable Risks from Epidemiological Data. **International Journal of Epidemiology** 7, 2, 175-182.

Lindeman, R. H.; Merenda, P. F. & Gold, R. Z. (1980) *Introduction to Bivariate and Multivariate Analyses.* Glenview, IL: Scott, Foresman.

Walter, S. D. (1980) Prevention for multifactorial disease. **American Journal of Epidemiology** 112, 3, 409-416.

Doll, Richard & Peto, Richard (1981) The Causes of Cancer: Quantitative Estimates of Avoidable Risks of Cancer in the United States Today. **Journal of the National Cancer Institute** 66, 6, 1191-1308.

Hartigan, J. A. & Kleiner, Beat (1981) Mosaics for contingency tables. **13th Symposium on the Interface**, Springer Verlag, New York, pp. 268-273.

Morgenstern, H. & Bursic, E. (1982) A method for using epidemiologic data to estimate the potential impact of an intervention on the health status of a target population. **Journal of Community Health** 7, 4, 292-309.

Whittemore, A. S. (1982) Statistical methods for estimating attributable risk from retrospective data. **Statistics in Medicine** 1, 229-243.

Whittemore, A. S. (1983) Estimating attributable risk from case-control studies. **American Journal of Epidemiology** 117, 1, 76-85.

Cox Jr., L. A. (1984) Probability of causation and the attributable proportion of risk. **Risk Analysis** 4, 3, 221-230.

Hartigan, J. A. & Kleiner, Beat (1984) A Mosaic of Television Ratings. **The American Statistician** 38, 1, 32-35.

Bruzzi, P.; Green, S. B.; Byar, D. P.; Brinton, L. A. & Schairer, C. (1985) Estimating the population attributable risk for multiple risk factors using case-control data. **American Journal of Epidemiology** 122, 5, 904-914.

Cox Jr., L.A. (1985) A new measure of attributable risk for public health applications. **Management Science** 31, 7, 800-813.

Kjuus, H.; Langård, S. & Skjærven, R. (1986) A case-referent study of lung cancer, occupational exposures and smoking - III Etiologic fraction of occupational exposures. **Scandinavian Journal of Work, Environment and Health** 12, 3, 210-215.

Cox Jr., L. A. (1987) Statistical issues in the estimation of assigned shares for carcinogenesis liability**. Risk Analysis** 7, 1, 71-78.

Kruskal, W. (1987a) Relative importance by averaging over orderings. **The American Statistician** 41, 1, 6-10.

Kruskal, W. (1987b) Correction. **The American Statistician** 41, 4, 341-341.

Last, John M. (ed.) (1988) A Dictionary of Epidemiology. 2nd edition. Oxford University Press, New York 1988.

Olsen, O. & Kristensen, T. S. (1988) Den ætiologiske fraktion - definition og teori. I **Hjerte/karsygdomme og arbejdsmiljø, Bind 3**. Hvor stor betydning har arbejdsmiljøet for hjerte/karsygdomme i Danmark? 39-188.

Benichou, J. & Gail, M.H. (1989) A Delta Method for Implicitly Defined Random Variables. **The American Statistician** 43, 1, 41-44.

Kruskal, William & Majors, Ruth (1989) Concepts of Relative Importance in Recent Scientific Literature. **The American Statistician** 43, 1, 2-6.

Bakke, P.; Eide, G.E.; Hanoa, R. & Gulsvik, A. (1991) Occupational dust or gas exposure and prevalences of respiratory symptoms and asthma in a general population. **European Respiratory Journal** 4, 3, 273-278.

Eide, Geir Egil (1991a) The Selection of Preventive Strategy Based on Stepwise Calculation of Population Attributable Fractions from a Multivariate Regression Model for a Binary Response. Internal report. **Institute of Mathematics and Statistics, Norwegian School of Economics and Business Administration & Section for Medical Informatics and Statistics, University of Bergen**, 1991-09-19, 50 pages.

Eide, Geir Egil (1991b) Strategies for Stepwise Estimation of Population Attributable Fractions. Poster at the **Joint Meeting of the International Society for Clinical Biostatistics and Society for Clinical Trials**, Brussels, July 8-12, 1991.

Edwards, A. W. F. & Edwards, J. H. (1992) Metrical Venn diagrams. **Annals of Human Genetics** 56, 71-75.

Greenland, S. & Drescher, K. (1993) Maximum likelihood estimation of the attributable fraction from logistic models. **Biometrics** 49, 3, 865-872.

Basu, Srabashi & Landis, J. Richard (1995) Model-based Estimation of Population Attributable Risk under Cross-sectional Sampling. **American Journal of Epidemiology** 14, 12, 1338-1343.

Eide, Geir Egil & Gefeller, Olaf (1995) Sequential and average attributable fractions as aids in the selection of preventive strategies. **Journal of Clinical Epidemiology** 48, 5, 645-655.

Kjuus, Helge; Hauge, Odd A.; Kongerud, Johny & Aasen, Tor B. (1996) Vurdering av årsaksforhold ved yrkesrelaterte lungesykdommer. (In Norwegian). **Tidsskrift for den Norske Lægeforening** 116, 6, 736-738.

Lloyd, C. J. (1996) Estimating attributable response as a function of a continuous risk factor. **Biometrika** 83, 3, 563-573.

Land, Matthias & Gefeller, Olaf (1997) A Game-Theoretic Approach to Partitioning Attributable Risks in Epidemiology. **Biometrical Journal** 39, 7, 777-792.

Theus, M. (1997) Visualizing categorical data. In: Bandilla, W. & Faulbaum, F. (eds.), **Softstat '97: Advances in Statistical Software** 6, Lucius and Lucius, Stuttgart, 47-55.

Brady, Anthony R. (1998) sbe21: Adjusted population attributable fractions from logistic regression. **Stata Technical Bulletin** 42 (STB-42), 8-12.

Gefeller, O.; Land, M. & Eide, G. E. (1998) Second Thoughts: Averaging Attributable Fractions in the Multifactorial Situation: Assumptions and Interpretation. **Journal of Clinical Epidemiology** 51, 5, 437-441.

Land, M. & Gefeller, O. (1998) A Multiplicative Approach to Partitioning the Risk of Disease. In: I. Balderjahn, R. Mathar, M. Schader (eds.): **"Classification, Data Analysis and Data Highways"**, 73-80. Berlin: Springer-Verlag.

Theus, Martin & Lauer, Stephan R. W. (1999) Visualizing Loglinear Models. **Journal of Computational and Graphical Statistics** 8, 3, 396-412.

Eide, Geir Egil (2000) Fundamental Concepts of Attributable Fractions and Their Visualization. **Centre for Clinical Research Research Report** 2000-01. Haukeland Hospital, Bergen.

Eide, Geir Egil & Gefeller, Olaf (2000) Re: Attributable risk in practice. [Letter] **American Journal of Epidemiology** 151, 2, 208-209.

Eide, Geir Egil & Heuch, Ivar (2001) Attributable fractions: fundamental concepts and their visualization. **Statistical Methods in Medical Research** 10, 3, 159-193.

Gefeller, Olaf (2001) The concept of attributable risk in epidemiology: yesterday, today and tomorrow. Editorial. **Statistical Methods in Medical Research** 10, 3, 157-158.

Hofmann, Heike (2001) Generalized Odds Ratios for Visual Modeling. **Journal of Computational and Graphical Statistics** 10, 4, 628-640.

Land, Matthias; Vogel, Christine & Gefeller, Olaf (2001a) Partitioning methods for multifactorial risk attribution. **Statistical Methods in Medical Research** 10, 3, 217-230.

Land, Matthias; Vogel, Christine & Gefeller, Olaf (2001b) A Multifactorial Variant of the Attributable Risk for Groups of Exposure Variables. **Biometrical Journal** 43, 4, 461-481.

Eagan, Tomas M. L.; Gulsvik, Amund; Eide, G. E. & Bakke, Per S. (2002) Occupational Airborne Exposure and the Incidence of Respiratory Symptoms and Asthma. **American Journal of Respiratory and Critical Care Medicine** 166, 7, 933-938.

Eide, Geir Egil & Heuch, Ivar (2002) Displaying Attributable Fractions in Three Dimensions Using the Scaled Venn Cube. **Centre for Clinical Research Research Report** 2002-01. Haukeland University Hospital, Bergen.

Eide, Geir Egil; Heuch, Ivar & Eagan, Tomas Mikal Lind (2002) The Scaled Venn Cube for Displaying Attributable Fractions. In: **Proceedings of the 17th International Workshop on Statistical Modelling** 215-226, Chania, Greece, July 2002, Eds: Mikis Stasinopoulos & Giota Touloumi.

Nafstad, Per; Samuelsen, Sven O.; Irgens, Lorentz M. & Bjerkedal, Tor (2002) Birth Weight and Hearing Impairment in Norwegians Born From 1967 to 1993. **Pediatrics** 110, 3, Art.No.e30.

Samuelsen, Sven Ove & Eide, Geir Egil (2003) Attributable Fractions for Survival Data. Oral presentation at **International Conference on Reliability and Survival Analysis (ICRSA)**, Columbia, South Carolina May 23, 2003.

Edwards, A. W. F. (2004) *Cogwheels of the Mind. The Story of Venn Diagrams.* The John Hopkins University Press, Baltimore and London 2004.

Eide, Geir Egil & Heuch, Ivar (2004) Average attributable fractions: a coherent theory for apportioning attributable fractions in populations and exposed to subpopulations and individual risk factors. **Centre for Clinical Research Research Report** 2004-01. Haukeland University Hospital, Bergen.

Grömping, Ulrike & Weimann, Ulla (2004) The asymptotic distribution of the partial attributable risk in cross-sectional studies. **Statistics** 38, 5, 427-438.

Llorca, Javier & Delgado-Rodríguez, Miguel (2004a) A new way to estimate the contribution of a risk factor in populations avoided nonadditivity. **Journal of Clinical Epidemiology** 57, 5, 479-483.

Llorca, Javier & Delgado-Rodríguez, Miguel (2004b) Erratum. **Journal of Clinical Epidemiology** 57, 8, 873-873.

Rowe, Alexander K.; Powell, Kenneth E. & Flanders, Dana (2004) Why Population Attributable Fractions Can Sum to More Than One. **American Journal of Preventive Medicine** 26, 3, 243-249.

Feldman, Barry (2005) Relative Importance Value. Preliminary manuscript March 19, 2005. Available online at **http://prismanalytics.com/docs/RelativeImportance050319.pdf.**

Chen, Ying Qing; Hu, Chencheng & Wang, Yan (2006) Attributable risk function in the proportional hazards model for censored time-to-event. **Biostatistics** 7, 4, 515-529.

Cox, Christopher (2006) Model-based estimation of the attributable risk in case-control and cohort studies. **Statistical Methods in Medical Research** 15, 6, 611-625.

Eide, Geir Egil (2006) How to estimate attributable fractions in Stata: A simple introduction. **Centre for Clinical Research Research Report** 2006-04. Haukeland University Hospital, Bergen.

Eide, Geir Egil & Heuch, Ivar (2006a) A Scaled Sample Space Cube Used to Illustrate Attributable Fractions. **Biometrical Journal** 48, 1, 93-104.

Eide, Geir Egil & Heuch, Ivar (2006b) Average Attributable Fractions: A Coherent Theory for Apportioning Excess Risk to Individual Risk Factors and Subpopulations. **Biometrical Journa**l 48, 5, 820-837.

Hernán, Miguel A. & Robins, James M. (2006) Estimating causal effects from epidemiological data. **Journal of Epidemiology and Community Health** 60, 578-586.

Hofmann, Heike (2006) Multivariate Categorical data – Mosaic plots. Ch. 5 in: **Graphics of Large Datasets. Visualizing a Million.** Springer, Singapore 2006, p.107-141.

Lehnert-Batar, Andrea (2006) **The pARtial Package**. Version 0.1 October 16, 2006, pp 11.

Lehnert-Batar, Andrea; Pfahlberg, Annette & Gefeller, Olaf (2006) Comparison of Confidence Intervals for Adjusted Attributable Risk Estimates under Multinomial Sampling. **Biometrical Journal** 48, 5, 805-819.

Rabe, C. & Gefeller, O. (2006) The Attributable Risk in a Multifactorial Situation. Evaluation of Different Methods of Partitioning. **Methods of Information in Medicine** 45, 4, 404-408.

Theus, Martin (2006) Statistical Graphics. Ch. 2 in: **Graphics of Large Datasets. Visualizing a Million.** Springer, Singapore 2006, 31-54.

Unwin, Antony; Theus, Martin & Hofmann, Heike (2006) *Graphics of Large Datasets. Visualizing a Million***.** Springer, Singapore 2006.

Unwin, Antony (2006) Interacting with Graphics. Ch. 4 in: **Graphics of Large Datasets. Visualizing a Million.** Springer, Singapore 2006, 73-101.

Benichou, Jacques (2007) Biostatistics and epidemiology: measuring the risk attributable to an environmental or genetic factor. **Comptes Rendus Biologies** 330, 4, 281-298.

Eide, Geir Egil & Heuch, Ivar (2007) Computation of Attributable Fractions on the Basis of Exposure Probabilities. Commentary. **Sleep** 30, 4, 386-386.

Grömping, Ulrike (2007a) Estimators of Relative Importance in Linear Regression Based on Variance Decomposition. **The American Statistician** 61, 2, 139-147.

Grömping, Ulrike (2007b) Response to: Menard S. (2007) "Two Simple Estimators of Relative Importance in Linear Regression Based on Variance Decomposition." Letter to de editor. **The American Statistician** 61, 3, 282-283.

Menard, Scott (2007) Two Simple Estimators of Relative Importance in Linear Regression Based on Variance Decomposition. Comment to Grömping, U. (2007) Estimators of Relative Importance in Linear Regression Based on Variance Decomposition. Letter to de editor. **The American Statistician** 61, 3, 281-282.

Rabe, Christina; Lehnert-Batar, Andrea & Gefeller, Olaf (2007) Generalized approaches to partitioning the attributable risk of interacting risk factors can remedy existing pitfalls. **Journal of Clinical Epidemiology** 60, 5, 461-468.

Samuelsen, Sven Ove & Eide, Geir Egil (2007) Attributable Fractions with Survival Data. **Statistics in Medicine**. Published Online: 10 Aug 2007 DOI: 10.1002/sim.3022.

Appendix III: English – Norwegian nomenclature

Attributable fraction (AF)   Tilskrivbar andel (TA)

Adjusted AF   Justert TA

Sequential AF   Sekvensiell TA

Combined AF   Samlet TA

Average AF   Gjennomsnittlig TA

AF in exposed   TA blant eksponerte

Relative risk   Relativ sannsynlighet

Relative odds   Relativ odds

Risk ratio   Sannsynlighetsforhold

Odds ratio   Oddsforhold

Scaled sample space square   Skalert utfallsromskvadrat

Scaled sample space cube   Skalert utfallsromskube

Survival time   Levetid

Ordered stepwise strategy   Ordnet stegvis strategi

Generalized AF   Generalisert TA

9. Articles I-V

I.   Eide, Geir Egil & Gefeller, Olaf (1995) "Sequential and average attributable fractions as aids in the selection of preventive strategies." Journal of Clinical Epidemiology 48, 5, 645-655.

II: Eide, Geir Egil & Heuch, Ivar (2001) "Attributable fractions: fundamental concepts and their visualization." Statistical Methods in Medical Research 10, 3, 159-193.

III.   Eide, Geir Egil & Heuch, Ivar (2006) "A scaled sample space cube used to illustrate attributable fractions." Biometrical Journal 48, 1, 93-104.

IV.   Eide, Geir Egil & Heuch, Ivar (2006) "Average attributable fractions: a coherent theory for apportioning excess risk to individual risk factors and subpopulations." Biometrical Journal 48, 5, 820-837.

V.   Samuelsen, Sven Ove & Eide, Geir Egil (2007) "Attributable fractions with survival data". Statistics in Medicine. Published Online: 10 Aug 2007 DOI: 10.1002/sim.3022.