Literature review
The attributable fraction in causal inference and genetics

Student: Elisabeth Dahlqwist

Main supervisor: Arvid Sjölander

Co-supervisor: Yudi Pawitan

Department of Medical Epidemiology and Biostatistics

Karolinska Institutet

August 2017

An introduction to the attributable fraction

The attributable fraction (AF) is a population-specific measure of the proportion of preventable outcomes, e.g. disease cases, had all subjects in the population been unexposed of a specific exposure. One aspect that have made the AF popular in epidemiology and public health is that it quantifies the exposure-outcome relationship by taking the exposure prevalence into account. The AF was first used in the 1950's by Morton Levin in a study of the relationship between smoking and lung cancer (Levin, 1953) and later on MacMahon and Pugh (1970) defined the attributable fraction in (1) for a binary outcome Y and a binary exposure X

$$AF = 1 - \frac{P(Y = 1 \mid X = 0)}{P(Y = 1)}, \tag{1}$$

where $P(Y = 1 \mid X = 0)$ is the outcome prevalence among unexposed and $P(Y = 1)$ is the overall outcome prevalence in the population. Reading Levin (1953) it is clear that the AF was intended as a causal measure where the outcome prevalence among unexposed $P(Y = 1 \mid X = 0)$ serve as a proxy for the counterfactual outcome prevalence had everyone been unexposed (with exposure status $X = 0$), $P(Y_0 = 1)$. In absence of confounding $P(Y_0 = 1)$ will equal $P(Y = 1 \mid X = 0)$ but in observational studies we usually expect confounding. If confounding is present the the definition of the AF in (1) cannot have a causal interpretation. A definition of the AF which is applicable for both randomized and observational studies is given in Sjölander and Vansteelandt (2011)

$$AF = 1 - \frac{P(Y_0 = 1)}{P(Y = 1)}. \tag{2}$$

In order to estimate the counterfactual outcome prevalence $P(Y_0 = 1)$ when confounding is present confounding adjustment is necessary. A range of methods have been developed to adjust the AF for confounding (Gefeller, 1992; Benichou, 2001). Let $\mathbf{Z}$ be a set of covariates sufficient for confounding control. If $\mathbf{Z}$ is observed we can estimate $P(Y_0 = 1)$ by adjusting for the covariates $\mathbf{Z}$ keeping $X = 0$ fixed for every subject

$$P(Y_0 = 1) = E_{\mathbf{Z}}\{P(Y = 1 \mid X = 0, \mathbf{Z})\}. \tag{3}$$

One model-based method for adjustment of the AF is to estimate the adjusted probability $P(Y = 1 \mid X = 0, \mathbf{Z})$ by logistic regression. By marginalizing over the sampling distribution of $\mathbf{Z}$ the population average $E_{\mathbf{Z}}\{P(Y = 1 \mid X = 0, \mathbf{Z})\}$ can be estimated (Sturmans et al., 1977; Sjölander and Vansteelandt, 2011). This does not work in case-control studies since the outcomes are fixed by study design. However, the AF can be re-written as

$$E_{X,\mathbf{Z}|Y=1}\{RR(\mathbf{Z})^{-X}|Y = 1\} \tag{4}$$

where $RR(\mathbf{Z}) = \frac{P(Y=1|X=1,\mathbf{Z})}{P(Y=1|X=0,\mathbf{Z})}$ is the risk ratio in every strata of $\mathbf{Z}$. If the disease is rare the risk ratio can be approximated by the odds ratio in every strata of $\mathbf{Z}$, $OR(\mathbf{Z}) = \frac{P(Y=1|X=1,\mathbf{Z})/P(Y=0|X=1,\mathbf{Z})}{P(Y=1|X=0,\mathbf{Z})/P(Y=0|X=0,\mathbf{Z})}$, and

$$E_{X,\mathbf{Z}|Y=1}\{RR(\mathbf{Z})^{-X}|Y = 1\} \approx E_{X,\mathbf{Z}|Y=1}\{OR(\mathbf{Z})^{-X}|Y = 1\}. \tag{5}$$

It is then straight forward to estimate the adjusted odds ratio by logistic regression (Bruzzi et al., 1985).

Model-based adjustment methods for estimating AF have primarily been developed for binary outcomes in cross-sectional and case-control studies but recently methods for estimating AF have been extended to time-to-event outcomes (Chen et al., 2006, 2010; Sjölander and Vansteelandt, 2014; Samuelsen and Eide, 2008). Let T be the time-to-event of interest, e.g. time to death. The AF function is then defined as

$$AF(t) = 1 - \frac{\mathrm{P}(T_0 \leq t)}{\mathrm{P(T} \leq t)}, \tag{6}$$

where $\mathrm{P}(T \leq t)$ is the factual probability of an event at or before time t, and $\mathrm{P}(T_0 \leq t)$ is the counterfactual probability of an event at or before time t had the exposure been eliminated for everyone at baseline.

There are several generalizations of the AF for multiple exposures (Eide and Heuch, 2001; Bruzzi et al., 1985; Eide and Gefeller, 1995; Rämsch et al., 2009) and multiple exposures levels (Eide and Heuch, 2001; Morgenstern and Bursic, 1982; Drescher and Becher, 1997; Taguri et al., 2012) which allow for more advanced scenarios of the intervention effect. In our projects we focus on the causal aspect of the AF and the role of the AF in genetic epidemiology. In project 1 we provide tools for estimating a model-based adjusted AF for different study designs. In project 2 we have developed a model to adjust for cluster shared unmeasured confounding and in project 4 we will develop a sensitivity parameter for unmeasured confounding which is specific for the AF and variation independent of the observed data. In the third project we focus on explaining the relationship between heritability and AF by using the liability threshold model.

Software for estimating the AF based on different study designs

In practice the availability of statistical methods depends on the availability of the method in statistical software. Even though model-based estimation of the AF has been developed for most statistical software (Cox and Li, 2012) there has been no uniform tool for estimating model-based AF for various study designs in the statistical software `R`. When we started this project in 2015 there were only three packages available at `CRAN`: `epiR` (Nunes et al., 2015), `attribrisk` (Schenck et al., 2014) and `paf` (Chen, 2014). However, each of these package had its own limitations. For example, the function `epi.2by2` in the `epiR` package which estimates the AF for cross-sectional, case-control and cohort study designs does not allow for model-based confounder adjustment. Moreover, the `attribrisk` package estimates the AF for case-control and cross-sectional study designs but relies on the 'rare-disease' assumption and is thus in practice restricted to case-control studies. For cohort study designs the AF can be estimated with the `paf` package using Cox proportional hazard regression for confounder adjustment. The main limitation of the `paf` package is that it does not handle big data (in our simulations it breaks down for data with around 20,000 observations or more). Another limitation of all these three packages is that

none of them provides accurate standard errors when data are clustered, e.g. when there are repeated measures on each subject (Dahlqwist et al., 2016). In our package **AF** we aimed at solving the limitations in the other packages and creating a more uniform tool for estimating the AF in different study designs. For example, we made an effort to make it possible to use large datasets and we used the delta method and the sandwich estimator for computing analytical standard error in order to reduce the computation time. We also made it possible to calculated correct standard errors for clustered data. In the latest version of the package the AF is estimated based on the logistic regression (`AFlogit`), conditional logistic regression (`AFclogit`) and a Cox proportional hazard regression (`AFcoxph`). As a part of the second project we also added a function for estimating the AF based on a frailty model (`AFparfrailty`).

Estimating absolute effects adjusted for cluster shared unobserved confounders

As noted in the introduction, the AF assumes that the relationship between the exposure and the outcome is causal and in order to estimate a causal effect confounding adjustment is necessary. In most cases when observational data is used we expect that we have unobserved confounding. For example, the environment when growing up or genes might influence both the exposure X and the outcome Y. Assume we have two subjects indexed by j and j' in the same cluster i. If we assume that these two subjects share the unmeasured confounders U_i but may have different exposure status X_{ij} and $X_{ij'}$, outcome status Y_{ij} and $Y_{ij'}$ and observed confounder status Z_{ij} and $Z_{ij'}$, respectively, we can describe the situation by a directed acyclic graph as

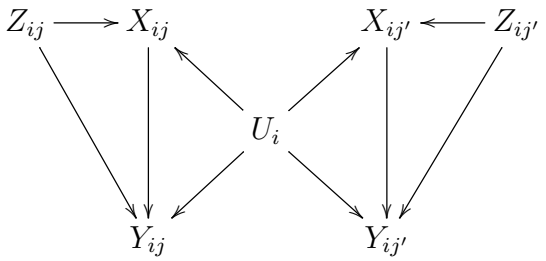

Figure 1: Directed Acyclic Graph (DAG) of the causal relationship between exposure status X, outcome Y and measured confounders Z and cluster shared unmeasured confounders U for subject j and j' in cluster i.

In survival data, the main method to adjust for cluster shared unmeasured confounders U_i as described in Figure 1 is the stratified Cox model. In the stratified Cox proportional hazard model we model the hazard ratio conditioned on the exposure X, covariate Z and cluster i in the following way,

$$\lambda(t \mid X_{ij}, Z_{ij}, \text{cluster } i) = \lambda_{0i}(t) e^{\beta_W X_{ij} + \gamma Z_{ij}} \tag{7}$$

where t is the event-time. However, the stratified Cox model does not model the baseline hazard $\lambda_{0i}(t)$ and thus, we cannot estimate absolute effects necessary to estimate standardized survival curves or the

AF function (Rothman et al., 2008; Sjölander, 2016). An alternative method is the frailty model (which parallels the random effect model for point outcomes)

$$\lambda(t \mid X_{ij}, Z_{ij}, \text{cluster } i) = \lambda_0(t) U_i e^{\beta_W X_{ij} + \gamma Z_{ij}} \tag{8}$$

where it is assumed that we have a cluster-specific 'frailty' effect U_i which we assume follow some distribution (Aalen, 1994; Hougaard, 1995; Aalen et al., 2015). The limitation with the ordinary frailty model is that it does not adjust for cluster shared unobserved confounding (Sjölander et al., 2013). Thus, there are different limitations with the stratified Cox model and the ordinary frailty model which hinge absolute risk estimation. Brumback et al. (2010) propose the 'between-within' model as a possible solution to this problem for binary data. The between-model was first described by Mundlak (1978) and later by Neuhaus and Kalbfleisch (1998). The basic idea with the between-within model is to include a within cluster effect in order to adjust for cluster shared unobserved confounding. Sjölander et al. (2013) have described how the between-within model can be used for time-to-event outcomes. Sjölander et al. (2013) propose assuming a Weibull baseline hazard and a gamma distributed frailty. The between-within frailty model can be defined as

$$\begin{aligned} \lambda(t \mid X_{ij}, Z_{ij}, u_i) &= \lambda_0(t) U_i e^{\beta_W X_{ij} + \beta_B \bar{X}_i + \gamma Z_{ij}} \\ &= \lambda_0(t) \underbrace{U_i e^{\beta_B \bar{X}_i}}_{U_i^*} e^{\beta_W X_{ij} + \gamma Z_{ij}} \end{aligned} \tag{9}$$

where $U_i^* = U_i e^{\beta_B \bar{X}_i}$ and are assumed to be i.i.d with following a gamma distribution $U_i^* \sim \Gamma(\frac{1}{\theta}, \theta \mathrm{e}^{\beta_B \bar{X}_j})$. The between-cluster effect β_B captures if there is unobserved confounding (U) in u_i and the within-cluster effect β_W estimates the within cluster effect of X on Y. We have described how the BW model can be used for standardization and estimation of the AF. We have also updated the version of the package `AF` and `stdReg` with implementation of standardization and estimation of the AF based on a Weibull gamma-frailty model.

Sensitivity analysis of unmeasured confounding for the AF

The first and second projects concerned ways to adjust for measured and cluster shared unmeasured confounding when estimating the AF. In practice it is not always possible to adjust for unmeasured confounding and this leads to biased estimates. A way to understand the magnitude of the bias due to different scenarios of unmeasured confounding is to conduct a so called 'sensitivity' analysis. Sensitivity analysis of unmeasured confounding were first mentioned in the 1950s by Cornfield et al. (1959) and later on a variety of approaches to sensitivity analysis has been described by Schlesselman (1978); Rosenbaum (1995); Lin et al. (1998) and Greenland (2003). Sensitivity analysis have been developed for Inverse Probability Weighting (Robins et al., 2000), Marginal Structural Models (Brumback et al., 2004) and Propensity Score Models (VanderWeele, 2013). Chiba (2012a,b) have shown how the method proposed

by Brumback et al. (2004) can be applied to conduct sensitivity analysis of the AF. The problem with unmeasured confounding when estimating the AF arises in the estimation of the counterfactual probability of the outcome if all were unexposed, $P(Y_0 = 1)$. Chiba (2012a,b) proposes the following sensitivity parameter δ for assessing the bias in $P(Y_0 = 1)$

$$\begin{aligned} \delta &\equiv P(Y_0 = 1 \mid X = 1, \mathbf{Z}) - P(Y_0 = 1 \mid X = 0, \mathbf{Z}) \\ &= P(Y_0 = 1 \mid X = 1, \mathbf{Z}) - P(Y = 1 \mid X = 0, \mathbf{Z}) \end{aligned} \tag{10}$$

where X denotes the binary exposure which takes the value 1 for exposed and 0 for unexposed, Y_0 denotes the counterfactual outcome had all subjects been unexposed and $\mathbf{Z}$ denotes a set of measured covariates that are adjusted for. The sensitivity parameter δ is assumed to be constant between the strata of $\mathbf{Z}$. It should be noted that while $P(Y_0 = 1 \mid X = 1, \mathbf{Z})$ is a counterfactual quantity the quantity $P(Y_0 = 1 \mid X = 0, \mathbf{Z})$ is factual and equal to $P(Y = 1 \mid X = 0, \mathbf{Z})$ since it is the probability of the outcome among factually unexposed subjects within a given strata of $\mathbf{Z}$. Moreover, Chiba (2012a) show how the true AF (AF) can be expressed as a function of the sensitivity parameter δ and the initial estimate of AF (AF^O) adjusted only for measured confounders $\mathbf{Z}$

$$\begin{aligned} AF &= 1 - \frac{\sum_{\mathbf{Z}} P(Y = 1 \mid X = 0, \mathbf{Z})}{P(Y = 1)} - \frac{\sum_{\mathbf{Z}} \delta P(\mathbf{Z} \mid X = 1) P(X = 1)}{P(Y = 1)} \\ &= AF^O - \frac{P(X = 1)}{P(Y = 1)} \delta. \end{aligned} \tag{11}$$

By giving a range of values of δ the result of the sensitivity analysis can be displayed graphically. However, the range in which the investigator let δ vary will have the *a priori* bounds [-1,1] before the data is observed. After observing the data the possible bounds of δ will depend on the values on the factually observed outcome in the unexposed group, $P(Y = 1 \mid X = 0, \mathbf{Z})$, for every strata of $\mathbf{Z}$. Thus, $P(Y = 1 \mid X = 0, \mathbf{Z})$ is not variation independent of δ and given the observed distribution of Y and X the possible values of δ are restricted. Moreover, the observed value of $P(Y = 1 \mid X = 0, \mathbf{Z})$ is different for different strata of $\mathbf{Z}$ which implies that the values that bound the possible values of δ may change for every strata of $\mathbf{Z}$. This is not a problem if $\mathbf{Z}$ contain few levels but if $\mathbf{Z}$ is continuous finding the bounds of δ is complicated. A solution to this problem is to find another sensitivity parameter which is variation independent of the observed data. One possibility is to define the sensitivity parameter as the coefficient of the exposure effect in a logistic regression. We define the counterfactual logistic regression model as

$$logit\{P(Y_0 = 1 \mid X, \mathbf{Z})\} = \alpha + \boldsymbol{\gamma}\mathbf{Z} + \beta X \tag{12}$$

We can then define the sensitivity parameter as,

$$\beta = logit\{P(Y_0 = 1 \mid X = 1, \mathbf{Z})\} - logit\{P(Y = 1 \mid X = 0, \mathbf{Z})\}. \tag{13}$$

The proposed sensitivity parameter β is variation independent of $P(Y, X, \mathbf{Z})$ since the parameters in a logistic regression are not bounded. Moreover, we can then estimate the counterfactual probability

$P(Y_0 = 1)$ as,

$$P(Y_0 = 1) = \int_{\mathbf{Z}} \{P(Y_0 = 1 \mid X = 0, \mathbf{Z})P(X = 0 \mid \mathbf{Z}) + P(Y_0 = 1 \mid X = 1, \mathbf{Z})P(X = 1 \mid \mathbf{Z})\} P(\mathbf{Z}) \quad (14)$$

which we can estimate by two different logistic regression models. The factual probability $P(Y = 1 \mid X = 0, \mathbf{Z})$ can be estimated by $expit(\alpha + \boldsymbol{\gamma}\mathbf{Z})$ and the counterfactual probability $P(Y_0 = 1 \mid X = 1, \mathbf{Z})$ can be estimated by $expit(\alpha + \boldsymbol{\gamma}\mathbf{Z} + \beta)$ where β can be set by the researcher. Similarly, the exposure model $P(X \mid \mathbf{Z})$ can be estimated with $expit(\alpha^* + \boldsymbol{\gamma}^*\mathbf{Z})$ from the logistic regression model $logit\{P(X = 1 \mid \mathbf{Z})\} = \alpha^* + \boldsymbol{\gamma}^*\mathbf{Z}$. Thus, by using the definition of $P(Y_0 = 1)$ in Equation 14 we can show that $P(Y_0) = P(Y = 1 \mid X = 0)$ if there is no unmeasured confounding, that is if $\beta = 0$

$$\begin{aligned} P(Y_0 = 1) = &\int_{\mathbf{Z}} [expit(\alpha + \boldsymbol{\gamma}\mathbf{Z})\{1 - expit(\alpha^* + \boldsymbol{\gamma}^*\mathbf{Z})\} + expit(\alpha + \boldsymbol{\gamma}\mathbf{Z} + \beta)expit(\alpha^* + \boldsymbol{\gamma}^*\mathbf{Z})] \, P(\mathbf{Z}) \\ &\text{and if } \beta = 0 \\ =&\int_{\mathbf{Z}} expit(\alpha + \boldsymbol{\gamma}\mathbf{Z})P(\mathbf{Z}) \\ =&P(Y = 1 \mid X = 0). \end{aligned} \quad (15)$$

By this approach the range of values of the sensitivity parameter β will not be restricted by the observed data and by expressing $P(Y_0 = 1)$ as a function of β we can get a graphical description of the magnitude of confounding bias in the AF.

The AF as a function of heritability

Heritability measures familial resemblance of a trait in a population and is a popular measure in genetic epidemiology. However, heritability tends to be difficult to interpret as it is a ratio of two variances and it is often misinterpreted, sometimes as the AF (Visscher et al., 2008). Even though both measures quantify the effect of a genetic factor on a trait in a specific population the measures estimate different quantities. While the heritability estimate the proportion of variance in phenotype that can be explained by genetic variation in a population the AF estimates the proportion of a trait that would be avoided had a genetic component been eliminated in the population. To our knowledge the exact relationship between these measures has not been described in the literature. Previous work using the AF in a genetic context can roughly be divided into estimating the AF for an intervention on the overall genetic effect and for an intervention on carriers of specific single-nucleotide polymorphisms (SNPs). For example, Ramakrishnan and Thacker (2012) use twin data to estimate the AF for the overall genetic effect. A limitation with their approach is that the exposure is defined as elimination of the outcome in the affected co-twin while the actual exposure of interest is the genetic set-up which confound the association in outcome status between the twin-pairs. Another approach described in Witte et al. (2014) use genetic

data on SNPs to estimate the AF. In this application the exposure is defined as carriers of certain SNPs which implies that the AF is not estimated for the overall genetic effect.

A way to capture the complexity of multifactorial traits is to use a 'liability model'. In the estimation of heritability for binary traits it is usually assumed that there is an underlying continuous variable, the liability, which is composed of genetic and environmental factors that aggregate the liability to the outcome. The observed outcome status can then be regarded as an indicator of whether the individual has a liability that precedes a certain threshold on the liability scale (Falconer and Mackay, 1996). A simplification of the liability model is the commonly used ACE liability model.

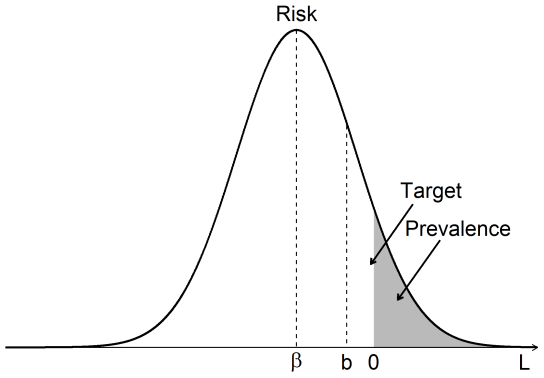

Figure 2: *The liability threshold model with threshold 0, location intervention of size Δ and target population for the intervention defined as b.*

In the ACE liability threshold model it is assumed that several small genetic (A), common environmental (C) and random environmental effects (E) add up to the liability (L). In our approach we combine the common environmental and random environmental effects into overall environmental effects (E). We assume that the liability to the outcome is composed of the baseline risk β and additive genetic (A) and environmental (E) effects such that $L = \beta + A + E$. We further assume that the univariate distribution of A and E are normally distributed with zero mean and variances σ_A^2 and σ_E^2 respectively. From our assumption of the relationship between the liability, the genetic and the environmental component the distribution of the underlying liability is also normally distributed with mean β and variance $\sigma_A^2 + \sigma_E^2$. In the liability model the joint distribution of the underlying liability (L) and the genetic component (A) is assumed to be bivariate normal and the correlation between L and A can be shown to equal the square root of the heritability, h. This parametrization of the liability model can then be used to link the heritability to the AF. In the model the observed prevalence of the outcome Y represents the proportion of all subjects that precedes a certain threshold on the underlying liability scale (L). Let $P(Y = 1)$ denote the outcome prevalence Y in a certain population and let 0 be the risk threshold for getting the outcome. We can then express the outcome prevalence Y as $P(Y = 1) = P(L > 0)$. In order to capture the overall genetic effect we propose to use the ACE liability model to estimate the AF. However, in order to estimate the AF we have to define some intervention as well as a target group for the intervention. In Figure 2 we describe a hypothetical scenario of the relationship between liability,

prevalence and target group of the intervention.

Let Y_Δ denote the counterfactual outcome under the scenario in which an intervention is given to all subjects in the target group. We can then define the AF where the intervention is limited to a specific target group as

$$\begin{aligned} AF &= \frac{P(Y=1, \text{target group}) - P(Y_\Delta = 1, \text{target group})}{P(Y=1)} \\ &= \frac{P(L>0, G>b\sigma_A) - P(L_\Delta>0, G>b\sigma_A)}{P(Y=1)} \\ &= \frac{\Phi\left[\Phi^{-1}\{P(Y=1)\}, -b; h\right] - \Phi\left[\Phi^{-1}\{P(Y=1)\} - kh, -b; h\right]}{P(Y=1)} \end{aligned} \tag{16}$$

Thus, the AF for a limited target group can be expressed as a function of heritability, intervention effect size, size of target group and prevalence of the outcome. Moreover, by assuming a fixed value of intervention effect size and target group size the relationship between the AF can graphically be expressed as a function of heritability and outcome prevalence,

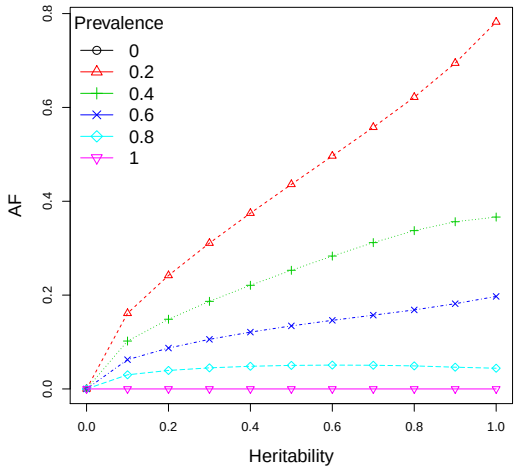

Figure 3: *The AF as a function of heritability and outcome prevalence given an intervention effect size of 2 and 1 as the target group cut-off.*

Likewise we can express the AF as a function of target group size and intervention size effect for a given prevalence and heritability. This is a potentially useful tool to graphically explain how the AF can increase with heritability and to understand the impact of intervention effect size and target group size for the AF for a specific disease and scenario.

Finally, it should be noted that the reliability of heritability estimates from the ACE-model have been questioned since this estimation strategy is based untestable assumptions of bivariate and univariate normality (Benchek and Morris, 2013). Therefore we note that the approach for estimating the AF as a function of heritability is limited to scenarios in which we have multifactorial traits where all risk factors are small and act additively on the overall risk of the trait.

References

Aalen, O. O. (1994). Effects of frailty in survival analysis. *Statistical Methods in Medical Research* **3,** 227–243.

Aalen, O. O., Valberg, M., Grotmol, T., and Tretli, S. (2015). Understanding variation in disease risk: the elusive concept of frailty. *International Journal of Epidemiology* **44,** 1408–1421.

Benchek, P. H. and Morris, N. J. (2013). How meaningful are heritability estimates of liability? *Human Genetics* **132,** 1351–1360.

Benichou, J. (2001). A review of adjusted estimators of attributable risk. *Statistical Methods in Medical Research* **10,** 195–216.

Brumback, B. A., Dailey, A. B., Brumback, L. C., Livingston, M. D., and He, Z. (2010). Adjusting for confounding by cluster using generalized linear mixed models. *Statistics & Probability Letters* **80,** 1650–1654.

Brumback, B. A., Hernn, M. A., Haneuse, S. J. P. A., and Robins, J. M. (2004). Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Statistics in Medicine* **23,** 749–767.

Bruzzi, P., Green, S. B., Byar, D. P., Brinton, L. A., and Schairer, C. (1985). Estimating the Population Attributable Risk for Multiple Risk Factors Using Case-Control Data. *American Journal of Epidemiology* **122,** 904–914.

Chen, L. (2014). paf: Attributable Fraction Function for Censored Survival Data.

Chen, L., Lin, D. Y., and Zeng, D. (2010). Attributable fraction functions for censored event times. *Biometrika* **97,** 713–726.

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

Chiba, Y. (2012a). Sensitivity analysis for unmeasured confounding of attributable fraction. *Epidemiology (Cambridge, Mass.)* **23,** 175–176.

Chiba, Y. (2012b). A Simple Method for Sensitivity Analysis of Unmeasured Confounding. *Journal of Biometrics & Biostatistics* **3,**.

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, E. L. (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer Institute* **22,** 173–203.

Cox, C. and Li, X. (2012). Model-Based Estimation of the Attributable Risk: A Loglinear Approach. *Computational statistics & data analysis* **56,** 4180–4189.

Dahlqwist, E., Zetterqvist, J., Pawitan, Y., and Sjölander, A. (2016). Model-based estimation of the attributable fraction for cross-sectional, case-control and cohort studies using the R package AF. *European Journal of Epidemiology* **31,** 575–582.

Drescher, K. and Becher, H. (1997). Estimating the Generalized Impact Fraction from Case-Control Data. *Biometrics* **53,** 1170–1176.

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

Eide, G. E. and Heuch, I. (2001). Attributable fractions: fundamental concepts and their visualization. *Statistical Methods in Medical Research* **10,** 159–193.

Falconer, D. S. and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Longman.

Gefeller, O. (1992). Comparison of adjusted attributable risk estimators. *Statistics in Medicine* **11,** 2083–2091.

Greenland, S. (2003). The Impact of Prior Distributions for Uncontrolled Confounding and Response Bias: A Case Study of the Relation of Wire Codes and Magnetic Fields to Childhood Leukemia. *Journal of the American Statistical Association* **98,** 47–54.

Hougaard, P. (1995). Frailty models for survival data. *Lifetime Data Analysis* **1,** 255–273.

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

Lin, D. Y., Psaty, B. M., and Kronmal, R. A. (1998). Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* **54,** 948–963.

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

Morgenstern, H. and Bursic, E. S. (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,** 292–309.

Mundlak, Y. (1978). On Pooling Time Series and Cross Section Data. *Econometrica* **1,** 69–85.

Neuhaus, J. M. and Kalbfleisch, J. D. (1998). Between- and Within-Cluster Covariate Effects in the Analysis of Clustered Data. *Biometrics* **54,** 638–645.

Nunes, M. S. w. c. f. T., Heuer, C., Marshall, J., Sanchez, J., Thornton, R., Reiczigel, J., Robison-Cox, J., Sebastiani, P., Solymos, P., Yoshida, K., Jones, G., and Firestone, S. P. a. S. (2015). epiR: Tools for the Analysis of Epidemiological Data.

Ramakrishnan, V. and Thacker, L. R. (2012). Population Attributable Fraction as a Measure of Heritability in Dichotomous Twin Data. *Communications in statistics: Simulation and computation* **41,**.

Rämsch, C., Pfahlberg, A. B., and Gefeller, O. (2009). Point and interval estimation of partial attributable risks from case-control data using the R-package 'pARccs'. *Computer Methods and Programs in Biomedicine* **94,** 88–95.

Robins, J. M., Rotnitzky, A., and Scharfstein, D. O. (2000). Sensitivity Analysis for Selection bias and unmeasured Confounding in missing Data and Causal inference models. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, The IMA Volumes in Mathematics and its Applications, pages 1–94. Springer, New York, NY.

Rosenbaum, P. R. (1995). *Observational Studies*. Springer, New York.

Rothman, K. J., Greenland, S., and Lash, T. L. (2008). *Modern Epidemiology*. Lippincott Williams & Wilkins.

Samuelsen, S. O. and Eide, G. E. (2008). Attributable fractions with survival data. *Statistics in Medicine* **27,** 1447–1467.

Schenck, L., Atkinson, E., Crowson, C., and Therneau, T. (2014). attribrisk: Population Attributable Risk.

Schlesselman, J. J. (1978). Assessing effects of confounding variables. *American Journal of Epidemiology* **108,** 3–8.

Sjölander, A. (2016). Regression standardization with the R package stdReg. *European Journal of Epidemiology* **31,** 563–574.

Sjölander, A., Lichtenstein, P., Larsson, H., and Pawitan, Y. (2013). Between-within models for survival analysis. *Statistics in Medicine* **32,** 3067–3076.

Sjölander, A. and Vansteelandt, S. (2011). Doubly robust estimation of attributable fractions. *Biostatistics (Oxford, England)* **12,** 112–121.

Sjölander, A. and Vansteelandt, S. (2014). Doubly robust estimation of attributable fractions in survival analysis. *Statistical Methods in Medical Research* .

Sturmans, F., Mulder, P. G., and Valkenburg, H. A. (1977). Estimation of the possible effect of interventive measures in the area of ischemic heart diseases by the attributable risk percentage. *American Journal of Epidemiology* **105,** 281–289.

Taguri, M., Matsuyama, Y., Ohashi, Y., Harada, A., and Ueshima, H. (2012). Doubly robust estimation of the generalized impact fraction. *Biostatistics* **13,** 455–467.

VanderWeele, T. J. (2013). Unmeasured confounding and hazard scales: sensitivity analysis for total, direct, and indirect effects. *European journal of epidemiology* **28,** 113–117.

Visscher, P. M., Hill, W. G., and Wray, N. R. (2008). Heritability in the genomics era concepts and misconceptions. *Nature Reviews Genetics* **9,** 255–266.

Witte, J. S., Visscher, P. M., and Wray, N. R. (2014). The contribution of genetic variants to disease depends on the ruler. *Nature Reviews Genetics* **15,** 765–776.