DIGITALCOMMONS @WAYNESTATE

Journal of Modern Applied Statistical Methods

Volume 3 | Issue 1 Article 7

5-1-2004

Quantifying The Proportion Of Cases Attributable To An Exposure

Camil Fuchs

Department of Statistics and Operations Research Tel-Aviv University, Ramat-Aviv, Tel-Aviv Israel, fuchs@math.tau.ac.il

Vance W. Berger

Mathematics and Statistics Department University of Maryland, bergerv@mail.nih.gov

Journal of Modern Applied Statistical Methods
May, 2004, Vol. 3, No. 1, 54-64

1538 – 9472/04/$95.00

Quantifying The Proportion Of Cases Attributable To An Exposure

Camil Fuchs
Department of Statistics and Operations Research
Tel-Aviv University, Ramat-Aviv, Tel-Aviv Israel

Vance W. Berger
Mathematics and Statistics Department
University of Maryland

The attributable fraction and the average attributable fractions, which are commonly used to assess the relative effect of several exposures to the prevalence of a disease, do not represent the proportion of cases caused by each exposure. Furthermore, the sum of attributable fractions over all exposures generally exceeds not only the attributable fraction for all exposures taken together, but also 100%. Other measures are discussed here, including the directly attributable fraction and the confounding fraction, that may be more suitable in defining the fraction directly attributable to an exposure.

Key words: Attributable fraction, average attributable fraction, directly attributable fraction, confounded fraction, fractional complementary attributable risk, multifactor exposure

Introduction

If two identical units are treated differently, and respond differently, then the attribution of the differing responses to the differing treatments follows from the process of elimination, and is unambiguous. The same applies to a situation in which two identical groups are treated differently, even if these groups themselves are heterogeneous. Attribution becomes more of a challenge, however, when the groups differ systematically from each other on many dimensions, or exposures. Various measures of attributable fractions have been proposed in these situations, with many exposures being considered simultaneously; one particularly common one bears the name attributable fraction (AF), and is defined as

$$AF = \{Pr(disease) - Pr(disease|\ no\ exposure)\} / Pr(disease). \quad (1)$$

Camil Fuchs, Sackler School of Exact Sciences, Tel-Aviv University, Ramat-Aviv, Tel-Aviv, 69978, Israel. E-mail: fuchs@math.tau.ac.il. Vance W. Berger is a Mathematical Statistician at the National Cancer Institute. E-mail: bergerv@mail.nih.gov.

See Deubner et al. (1980); Kelsey et al. (1986); and Last (1983). The AF is generally interpreted as an estimate of either the proportion of the cases attributed to (or caused by) the exposure factor or the proportion of the cases that could be prevented if the exposure factor were eliminated. Its importance has grown lately as a measure for interventions, regulations, and lawsuits concerning the effect of the exposure to various factors. Thus, when the Surgeon General warned that 90% of the lung cancer cases are caused by smoking (Gori, 1989), that figure is based on the AF.

In lawsuits, the AF is used in two main contexts. In individual compensation cases, the court may wish to determine the likelihood that the disease of a particular individual was caused by the exposure at issue. The AF has been interpreted as an estimate of this likelihood (Kleinbaum et al., 1982; Greenland & Robins, 1988). Other cases involve class actions, in which states or HMOs sue manufacturers of a presumably hazardous agent for the medical expenses caused by the exposure factor. The medical expenses claimed to have been caused by the exposure factor are usually computed as the sum of the products of the attributable fractions relevant to the specific diseases and the total medical expenses related to those diseases.

The AF was initially termed the attributable risk (Levin, 1953). Other terms include the etiological factor (Kleinbaum et al.,

1982; Schesselman, 1982), the etiological fraction, and the fraction of etiology (Mietienen, 1974). The term attributable risk (e.g. Benichou, 1991), or its variants such as population attributable risk (e.g. Breslow & Day, 1980; MacMahon & Pugh, 1970) or population attributable risk per cent (Cole & MacMahon, 1971; Hennekens et al, 1987) seem to be used less often.

The AF "does not represent disease risk" (Greenland & Drescher, 1993). That is, the AF does not necessarily reflect the proportion of cases caused by the exposure factor; this has been discussed in the statistical and epidemiological literature (Feinstein 1988, 1995; Ashford, 1992; Gori, 1989). One bias originates from shortcomings inherent to epidemiological studies, which invalidate the collected data as representative of the studied populations.

There are also conceptual problems in the measurement of the effect of the exposure factor in general, and in the measurement of the causal effect in particular. For one thing, the AF lacks the desirable property of additivity; that is, in multifactorial diseases, the sum of AFs of all sources of variation (exposures) will generally exceed not only the AF of all exposures taken together, but also 100%. In fact, "... the total …attributable to the various causes is not 100% but infinity" (Rothman, 1986), which seems to suggest that "…we could prevent more than 100% of any given disease" (Gori, 1989). Many studies focus on a single exposure factor, so this drawback of the AF is not always evident; nevertheless, it remains relevant.

Eide and Gefeller (1995) and Land and Gefeller (2000) propose other measures for assessing the responsibility of the various factors, specifically average attributable fractions (AAFs) and the multiplicative fractional complementary attributable risks (FCARs), respectively. These measures "divide the indivisible" (Pratt,1987), as they allocate the overall reduced probability of disease into fractions whose sum equals the total effect of the considered exposures. This is accomplished by averaging over all sequences of exposures similar to the situation in multiple regressions with correlated regressors when considering the relative importance of terms (Kruskal, 1987, Kruskal & Majors, 1989, Pratt, 1987, Gnizi, 1993).

Although these methods may be appropriate for "solving the problem of shared responsibilities for the prevalence of a disease in the population" (Eide & Gefeller, 1995), additivity is not sufficient to ensure a reasonable measure, and the AAF and the FCAR do not represent adequately the proportion of disease attributable to each exposure separately. The task remains to decompose the attributable fraction for the simultaneous exposure to all exposure factors.

When multiple factors contribute to a disease, the ideal situation of perfect knowledge about the relevant variables and of proper collection of data on those variables at the appropriate levels may be rare. But even in these ideal situations, the AF is not an appropriate measure for the assessment of the proportion of cases that can be attributed to an exposure factor. It is even more certainly not a measure of the proportion of cases caused by the exposure factor.

Proposed here is decomposing the AF for the simultaneous exposure to all factors by using terms that are sequentially conditioned on nested sets of factors. The last term is conditioned on all the previous factors and is called the directly attributable fraction (DAF). The DAF is analogous to the Type III sums of squares (Milliken & Johnson, 1984) in linear model theory, in that the variation attributable to an exposure is limited to the variation that cannot be explained by the totality of all other exposures taken together.

The confounded fraction (CF) is CF=AF-DAF; the AF of any exposure may be decomposed into a DAF and a CF. It is argued here that the DAF is a more appropriate measure of the proportion of cases that can be directly attributed to the exposure factor than the AF measure defined in (1) above. The overall effect of the exposure factor on the probability of disease is adequately represented by the pair (DAF, CF).

Methodology

First consider the case in which the risk of disease is potentially affected by a single exposure factor A at L levels, and by M adjusting factors (usually demographic variables such as gender, age, residence, etc.). By convention, the first level of the factor A corresponds to no exposure. Each configuration of a level of exposure and a specific combination of levels of the adjusting variables can be presented as a cell E_{sk} in a two way table, s=1,2,…,S; l=1,2,…,L. The rows r_1, r_2, …,r_S are the strata constructed from the combinations of levels of the adjusting factors, $S=G_1 \cdot G_2 \cdot \cdot \cdot G_M$ and the columns are the levels of A.

The attributable fraction for A, adjusted for the confounded variables, can be written as:

$$AF_A = \{Pr(D) - \Sigma_s Pr(D|E_{s0}) \cdot Pr(r_s)\} / Pr(D), \quad (2)$$

(Whittemore, 1982).

Furthermore, the contribution of each cell and of each column in the table to AF_A can be computed. Following Eide and Gefeller (1995), define

$$\lambda_{sl}=\{Pr(D|E_{sl})-Pr(D|E_{s0})\}Pr(E_{sl})/Pr(D) \text{ and } \lambda_l = \Sigma_s\lambda_{sl}. \quad (3)$$

Thus, λ_{sl} is the contribution of the configuration E_{sl} to $AF_A = \Sigma_{sl}\lambda_{sl}$ and λ_l is the contribution of the l-th level of exposure to the risk attributable to A. In particular, if A has only two levels, then the only contribution is due to the second (exposed) level. The extension to the general case of F exposures and M adjusting factors is immediate. The adjusted AF for each factor and for the joint effect of several factors can be computed using the appropriate two-way table representation. The columns of the two-way table are now the combinations of levels for the factors whose joint effect is to be computed. The other exposure factors is added to the set of adjusting variables and set the rows of the table as the combination of levels of the newly defined set of adjusting variables.

In particular, the attributable fractions for the F exposure factors, and especially for the factor of interest A can be computed. Thus, for the computation of the AF of the first exposure factor, the table has $L_2 \cdot L_3 \cdot \cdot \cdot L_F \cdot S$ rows and L_1 columns. The table for the second factor has number $L_1 \cdot L_3 \cdot \cdot \cdot L_F \cdot S$ rows and L_2 columns, while the table for the assessment of the joint AF of the first two factors has $L_3 \cdot L_4 \cdot \cdot \cdot L_F \cdot S$ rows and $L_1 \cdot L_2$ columns. An important special case assesses the AF for the joint effect of all the exposure factors for which data were collected. The two-way table has S rows and $L_1 \cdot L_2 \cdot \cdot \cdot L_F$ columns. The first column represents the category of exposure to none of the risk factors.

Estimation of the various AF's

In a cohort study, let n_{sl} be the number of individuals sampled in the E_{sl} configuration with $n_s = \Sigma_l n_{sl}$ and $n = \Sigma_s n_s$. The maximum likelihood estimate (MLE) for the AF under the logistic regression model-adjustment (Miettinen 1974; Walter 1975; Bruzzi et al 1985; Greenland 1987; Benichou & Gail 1990; Greenland & Drescher 1993) is obtained by substituting the proper estimates in equation (2) above. Let $Pr(Y=1|E_{sl})$ be the probability of disease at E_{sl}, where Y is an indicator variable taking the value one if the person is diseased, and zero otherwise. If the vector of carriers $\boldsymbol{x}$ is extended to include $x_1 \equiv 1$, then these probabilities are assumed to follow the logistic model:

$$\pi_{sl} = Pr(d=1|\boldsymbol{x}) = \exp(\boldsymbol{x\beta})/\{1+\exp(\boldsymbol{x\beta})\}. \quad (4)$$

For the $(s,l)^{th}$ configuration of covariate levels, let E_{s0} be the configurations of levels that a subject with configuration levels E_{sl} would have if not exposed to the studied factors (e.g. factor A). Furthermore, let p_{sl} be the MLE for π_{sl} and DIS be the proportion of diseased in the sample. The MLE for the AF for the studied factors is given by

$$AF = \{DIS - \Sigma_s p_{s0}(n_s / n)\} / DIS \quad (5)$$

The weighted-sum adjustment (Walter, 1976; Whittemore, 1982, 1983; Benichou, 1991) is a special case of the logistic regression model-adjustment with the fitted model being the saturated model. In this case the relative

frequencies d_{s0}/n_s substitute the estimated probabilities p_{sl} , where d_{sl} is the observed number of diseased in the configuration E_{sl}. Walter (1980) denotes the weighted-sum adjusted AF by the "proportional effect of *A*", and reserves the term "attributable fraction" for the unadjusted measure. In a case-control study with randomly sampled n_D cases (diseased) and n_C controls, the AF can be computed from equation (2) by dividing its numerator and denominator by the probability of disease in the no exposure configuration, i.e.

$$AF_A = 1 - \Sigma_s\{RR_{s0} \cdot Pr(r_s)\} / \Sigma_{sl}\{RR_{sl} \cdot Pr(E_{sl})\} . \quad (6)$$

For the estimation of AF_A, under the usual rare disease assumption, the estimates of the proportions of the various exposure configurations are replaced with the appropriate proportions in the sample of controls. The relative ratios RR_{sl} are approximated by the corresponding odds ratios from the sample.

Allocation of the overall effect

Consider two nested sets of exposure variables Q_1 and Q_2, with the second set being $Q_2 = Q_1 \cup A$, i.e. the second set includes all the variables in Q_1 and the extra factor *A*. The difference $AF_{Q1} - AF_{Q2}$, measures the conditional effect of *A*, given that all the factors in Q_1 have been removed, i.e. set at the non-exposure level.

In general, for a given ordered set of *F* exposure variables A_1, A_2,..., A_F, with sequentially nested sets $Q_1 = A_1$, $Q_2 = A_1 \cup A_2$,…,$Q_F = \cup_j A_j$, the factors can be remove one at a time to compute the *F* sequentially attributable fractions (*saf*s) $AF_{Q(j+1)} - AF_{Qj}$ (Eide and Gefeller,1995, Gefeller & Eide, 1998). The set of exposure factors can be extended to include $Q_0 = \phi$ by defining $AF_{Q0} = 0$. This extension properly defines the *AF* for the factor A_1 as the difference between $AF_{Q0} - AF_{Q1}$.

The *j*-th difference represents the conditional effect of the variable *A* positioned in the *j*-th location in the ordering, given that the previous *j*-1 exposures have been removed. Note that, with the exception of the last exposure, the *saf* for a variable depends on the original ordering.

By considering all *F*! possible orderings, the *saf*s for each variable can be computed, with all the combinations of other exposures being removed prior to its own removal. (Note that since a variable's *saf* depends only on the prior exposures, subsets of its *F*! *saf*s will have equal values.)

Cox (1985), Eide and Gefeller (1995), Gefeller and Eide, (1998) propose to compute the average of all possible *saf*s relate to each factor, and suggest that those *F* average attributable fractions (AAFs) are a reasonable measure of the responsibilityof the various factors when it is desired to share the disease load in the population among the analyzed exposures. The AAFs satisfy the important requirement that the *AF* for the joint effect of all the exposures equals the sum of the allocations (Cox, 1985).

A related approach for allocating the responsibility among several exposure factors has been lately proposed by Land and Gefeller (2000). Using a multiplicative Shapley value, they factorize the $1 - AF_{QF}$ into a set of *F* terms called factorial complementary attributable risks (FCARs) which, under this representation, measure the relative contributions of the exposure to the overall load of disease. Unlike the usual AFs, a small FCAR value represents a large effect of the respective factor.

For each factor, the *Pr*(*disease*|*no exposure*) is now substituted in equation (1) by *Pr*(*disease)**FCAR. The resulting ratio, called factorial attributable risk, equals FAR=1-FCAR. Those measures of shared responsibility do not possess the property that joint effect of all the exposures equals the sum of the allocations.

Directly attributable and confounded fractions

It has been mentioned before that since the AFs for the various exposure do not sum to the total attributable fractions for those exposures, the AFs cannot be considered as proper measures of the cases attributable to a factor. Furthermore, the same reason precludes the AFs from being a proper measure for apportioning, when some factors have to share together the responsibility (Gefeller & Eide, 1993, Eide & Gefeller, 1995).

The AAFs (and/or the FCARs and FARs) may be reasonable measures for solving

the problem of sharing the responsibility for the prevalence of a disease in a population, bur they are not good estimates of the effect of a single specific exposure. The allocation the total attributable effect does, what in the regression models context was called the "division of the indivisible" (Pratt, 1987), with the emphasis on the "indivisible".

To continue the parallel from regression models, note that in those models, the direct effect of a factor is commonly assessed by the extra sums of squares yielded when the factor in question is the last to be included in the model. Similarly, it is suggested that since the estimation of the effect for an exposure factor requires the removal of that factor, its directly attributable effect must be interpreted as the disease reduction when the factor is the last to be removed, and not the first.

Thus, if the attributable responsibility of *A* is considered to representthe segment of the probability of disease which is not explained by the other exposure factors, a more appropriate measure is obtained by ordering the set of exposures with the factor of interest as A_F, and defining the directly attributable fraction (DAF) as the last sequentially attributable fraction. The use of the last *saf* has been also recently proposed by Wilson et al (1998). They termed that special sequentially attributable fraction, resulting when the factor of interest is the last to be removed, “extra attributable fraction” (see also Eide & Gefeller, 2000). This is indeed appropriate in the estimation of the effect of a factor, derived by methods similar to the extra sum of squares in the linear regression models. Used here is the term directly attributable fraction, in the subject matter context, which assesses the attributability of the various fractions of the total probability of disease, and partitions the fraction in which that factor is involved into a directly attributable and a confounded fraction.

As noted before, the *saf* for the last exposure, does not depend on the original ordering. The calculation of the DAFs does not require ,the calculation of the intermediary *saf*s. The DAF for the factor of interest *A* is defined as the difference of two well defined *AF*s, i.e.

$$DAF_A = AF_{QF} - AF_{QF\sim A}, \quad (7)$$

where $QF\sim A$ is the set of all the exposure factors, except the factor *A*. This directly attributable fraction is the conditional attributable fraction for *A*, after removing the effects of all the other exposure factors.

The difference between the attributable fraction AF_A and the directly or conditional attributable fraction (DAF) as the confounded fraction (CF) of *A*, i.e. is defined as:

$$CF_A = AF_A - DAF_A \quad (8)$$

The confounded fraction is the segment of the probability of disease which is marginally attributed to *A*, but which is confounded and could just as well be attributed to the effect of the other exposure factors. The confounded factor can also be written as:

$$CF_A = AF_{QF\sim A} - (AF_{QF} - AF_A). \quad (9)$$

The confounded fraction for *A* can thus be interpreted as a difference of two AF terms related to the *notA* exposure factors. The first is the attributable fraction to all the factors which are *notA*, and the second is the effect of those same factors. after the removal of *A* (i.e. conditioned on *A*).

A related measure of conditional exposure effect

The conditional AF’s defined above are intuitively appealing since they represent the decomposition of the overall effect of the *F* exposure factors. An additional measure of the conditional exposure effect (CEE) is suggested here, not as an alternative to those presented above, but rather as yielding complementary information. The overall incidence rate after the removal of $A_1,A_2,\ldots A_t$ is:

$$Pr(D|A_1,A_2,\ldots A_t)=\{1 - Pr(D)\cdot AF_{Qt}\}. \quad (10)$$

The conditional exposure effect (CEE) can thus be defined as:

$$CEE_{A(t+1)|A1,A2,\ldots At} = (AF_{Q(t+1)} - AF_{Qt})\ Pr(D)\cdot / Pr(D|A_1,A_2,\ldots A_t) \quad (11)$$

If the correlations between *A* and the other exposure factors are roughly constant, the

various CEE's which correspond to a specific exposure factor, conditioned on various other effects, can be expected to differ only slightly from each other. Note that as in the case of the directly attributable fraction, the CEE's can be defined for an exposure factor conditioned on any subset of the F-1 variables, not only on their union.

$$Pr(D|E_{sl}) - Pr(D|E_{s0}) \} \cdot Pr(E_{sl})$$

Examples

The computations and the interpretation of the statistics presented in the previous sections are illustrated with a hypothetical example originated from Walter (1980, Table3). The data contain three dichotomous exposure factors (A, B, C). Complete information is provided on the proportions in the population for each configuration of levels of the exposure factors (the estimates of the $Pr(E_l)$'s) and with the respective incidence rates (the estimates of the $Pr(D|E_l)$'s). The original $Pr(\mathbf{E})$'s –vector was slightly altered to illustrate the fact that $\Sigma_j \mathrm{AF}_j$ j=1,..,F can exceed 100%. All the attributable fractions were computed with weighted-sum adjustments.

Panel (a) of Table 1 presents the data and the sequential vectors of estimated proportions in the populations exposed to each factor, following the various removals of factors. There are three factors which can be removed in stage 1, and the resulting statistics are denoted with the notation of (*|A), (*|B), (*|C) according to the respective removed factor. Similarly, one of the pairs of factors AB or AC or BC, is removed at the end of the second stage.. At stage 3 the remaining factor is removed and the conditional probability of disease is obtained, with all the factors being at the not exposed level.

First note that $Pr(D) = \Sigma_l Pr(D|E_l) \cdot Pr(E_l)$ = 0.4%, and that when all the three factors are controlled for, $\Sigma_l Pr(D|E_0) \cdot Pr(E_l)$ = 0.1%, yielding an overall attributable fraction for A+B+C of AF_{A+B+C} = 75%, i.e. the three factors together "can explain" 75% of the overall incident rates. The unconditional individual attributable fractions are 38.1%, 43.1% and 41.3%, respectively, whose sum is 122.5%. Panel (b) of Table 1 presents all the possible sequences of removal of factors.

Assume that the factor of interest is C. The unconditional AF seems to indicate that exposure to C is responsible for 41.3% of all the disease cases. However, when the effects of variables B and C are controlled for, only 9.4% of the cases can be directly attributable to C, and that the remainder of 31.9% is confounded effect with the other two factors.

Table 1 also presents the conditional exposure effects (CEEs) for all the stages. Unlike the conditional AFs, the CEEs are not necessarily monotonic and they vary less as a function of the removed variables.

Case Control Studies

The calculations are illustrated with the data on the oral cancer distributions among persons at the four configurations of (exposed, not exposed) to the alcohol and tobacco factors. The original data set of Rothman and Keller (1972) and Keller and Terris (1965) contained 598 case-control pairs. The data were further analyzed by Walter (1983).

The data summarized in Table 2 presents as initial data the four odds ratios (used to approximate the relative risks) and the proportions of controls in the four configurations (as estimates for $Pr(E_l)$'s). Panel (a) of Table 2 also presents the $P^{(t)}_g$-values for t=1,2.. Panel (b) presents the attributable fractions for the various levels of conditioning. It can be seen from the table that the two individual AFs for alcohol and tobacco are 66.2% and 72.1%, while the AF for the two factors taken together is 76.2%. Walter (1983) noticed that "very little additional is gained by removing tobacco and alcohol exposure as opposite to preventing exposure to just one of them".

Thus, one can expect that the computed individual AFs decompose into small directly attributable fractions and much larger confounded fractions. The entries in Table 2 confirm this expectation. The DAFs for alcohol and tobacco are 4.2% and 10% as opposed to the initial AFs of 66.2% and 72.1%. The remaining roughly 62% for both alcohol and tobacco are confounded fractions.

Table 1. Computation of the all the possible AFs, DAFs, CFs, and CEEs, for the hypothetical data wit dichotomous factors. The *P*-, *I*-, AF-, DAF-, CF-, and CEE-values are percentages.

Panel (a)

Design Factors			***Initial***		***Proportions in the population in the 1st, 2nd and 3rd stages***						
A	***B***	***C***	***Pr(D|E)***	***Pr(E)***	$P^{(1)}_{\|A}$	$P^{(1)}_{\|B}$	$P^{(1)}_{\|C}$	$P^{(2)}_{\|AB}$	$P^{(2)}_{\|AC}$	$P^{(2)}_{\|BC}$	$P^{(3)}_{\|ABC}$
0	0	0	20	0.1	45.0	32.5	32.5	62.5	70	50	100
0	0	1	13	0.2	25.0	17.5	0	37.4	0	0	0
0	1	0	13	0.3	17.5	0	17.5	0	30	0	0
0	1	1	5	0.8	12.5	0	0	0	0	0	0
1	0	0	25	0.2	0	30.0	37.5	0	0	50	0
1	0	1	13	0.5	0	20.0	0	0	0	0	0
1	1	0	5	0.6	0	0	12.5	0	0	0	0
1	1	1	8	1.8	0	0	0	0	0	0	0

Panel (b)

	A	***B***	***C***	***AB***	***AC***	***BC***	***ABC***
AF	38.1	43.1	41.3	65.6	60.0	62.5	75.0
$AF_{\|A}$		27.5	21.9			36.9	
$AF_{\|B}$	22.2		19.4		31.9		
$AF_{\|C}$	18.8	21.3		33.8			
$AF_{\|AB}$			9.4				
$AF_{\|AC}$		15.0					
$AF_{\|BC}$	12.5						
DAF	**12.5**	**15.0**	**9.4**				
CF	***25.6***	***28.1***	***31.9***				
AAF	**23.7**	**27.5**	**23.8**				
FAR	**35.8**	**40.4**	**40.7**				
CEE	38.1	43.1	41.3	65.6	60.0	62.5	75.0
$CEE_{\|A}$		44.4	35.4			59.6	
$CEE_{\|B}$	39.6		34.1		56.0		
$CEE_{\|C}$	31.9	36.2		57.4			
$CEE_{\|AB}$			27.3				
$CEE_{\|AC}$		37.5					
$CEE_{\|BC}$	33.3						

Table 2. Computation of the all the possible AFs, DAFs, CFs, and CEEs, for the case-control oral cancer data with two dichotomous factors.

Panel (a)

Tobacco	***Alcohol***	*Pr(E)-controls*	*RR(from OR)*	$P^{(1)}_{\|Alcohol}$	$P^{(1)}_{\|Tobacco}$	$P^{(2)}_{\|A+T}$
No User	***No User***	9	1.00	19	24	100
	User	10	1.23	0	76	0
User	***No User***	15	1.52	81	0	0
	User	66	5.71	0	0	0

Panel (b)

	Alcohol	***Tobacco***	***Both***
AF	66.2	72.1	76.2
$AF_{\|Alcohol}$		10.0	
$AF_{\|Tobacco}$	4.2		
DAF	**4.2**	**10.0**	
CF	**62.0**	**62.1**	
AAF	**35.2**	**41.0**	
FAR	**46.3**	**55.7**	
CEE	66.2	72.1	76.2
$CEE_{\|Alcohol}$		14.9	
$CEE_{\|Tobacco}$	29.6		

The analysis of the CEEs also reveals an interesting pattern. In the example presented in Table 1, the CEEs were relatively stable as a function of the extra conditioning, and did not differ dramatically from the AFs. On the other hand, in this example, the proportion of the incidence rates explained by the second term (alcohol or tobacco) is very low not only when the denominator is the overall incidence rate, but also when the denominator is the incidence rates remained after the first variable was removed. This is another facet of the highly confounding pattern in this data set.

This simple example also illustrate the contention that the AAFs value may provide an equitable solution for the problem of allocation of shared responsibility but is inappropriate for assessing the attributable fraction for a specific exposure. The corresponding AAFs are 35.1% and 41.0% which sum to the total effect of 76.2%. However, if only one exposure is considered, for example alcohol, its AAF of 35.1% is the average of 66.2% (the original AF) and the value of 4.2% (the DAF, which is the *saf* in the second step). It is very difficult to defend this value with any degree of confidence as representing the percent of cases attributable to alcohol. The same is true for smoking where the AAF of 41.0% is the average of 72.1% and 10%.

Conclusion

In the discussion following the analysis oral cancer presented above, Walter (1980) stated that "although the sum (of the AF's) exceeds 100%, this does not invalidate the individual (AF's) estimates; indeed, this phenomenon is more likely as more factors are considered and confounding becomes inevitable. Each measure must be interpreted as the disease reduction if the factor in question were the first to be removed".

However, when the purpose of the research is the assessment of causation and of attributable responsibility of a specific factor, the fact that the total contribution may exceed 100% does invalidate the AF's as interpretable measurements.

Assume that while assessing the effect of consumption of alcohol, one controls first for the effect of smoking by assessing the remaining incidence rates after all persons stopped smoking. Following this adjustment, the percent of cases for which the alcohol consumption is still "responsible" is assessed. The computations presented above show that the estimate of the percent of cases for which alcohol is found now responsible is 4%, instead of the initial 66%. The controlling for the tobacco variable didn't assume any change in the drinking behavior of the population.

Nevertheless, following the control for the smoking behavior, one witnesses a very significant decrease in the percent of cases attributable to alcohol consumption. It is thus clear that a significant proportion of the fraction initially attributed to drinking, can in fact be attributed to the effect of smoking, and vice versa.

The AAFs (and/or the FCARs and FARs) may be reasonable measures for solving the shared responsibility problem, but they are not proper estimates of the effect of a single specific exposure.

In contrast, the DAF has the clear interpretation as the fraction that can be attributed to that factor and which cannot be attributed to any of the other factors on which there are data in the sample. The complementary confounding fraction indicates the portion of the extra cases in which the factor in question may have been involved, but about which it is impossible to distinguish between its effect and the effects of the other factors.

Finally, note that for all measures of attributable fractions, the assumption that the data include all the relevant variables is cardinal for the validity of the results. As an illustration, constructed in the oral cancer is an artificial latent variable X, and set for the four combinations of X and alcohol (regardless of smoking) the RRs to be 1, 2, 10 and 20. The percents exposed to X in the four combinations of smoking and alcohol were set to be 17%, 6%, 31% and 68%, respectively. The collapsed table over X returns the previous pattern, but when X is considered, the AFs for Alcohol, Tobacco and X are 46%, 0% and 83%, with the AF for Alcohol*Tobacco*X (and also Alcohol*X) explaining 90.5% of the total load, a certainly different picture than in the previous analysis.

References

Ashford, J. R. (1992). Problems in assessing smoking-related mortality. *Smoking-Related Diseases, 33*, 263-274.

Benichou, J. (1991). Methods of adjustment for estimating the attributable risk in case-control studies: A review. *Statistics in Medicine, 10*, 1753-1773.

Benichou, J. & Gail, M. H. (1990). Variance calculations and confidence intervals for estimates of the attributable risk based on logistic models. *Biometrics, 46*, 991-1003.

Breslow, N. E., & Day, N. E. (1980). Statistical Methods in Cancer Research Vol 1: The Analysis of Case-control Studies, International Agency for Research on Cancer. *Scientific Publications*, No. 32, Lyon.

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*, 904-914.

Cole, P. & Macmahon, B. (1971). Attributable risk percent in case-control studies. *British Journal of Preventive and Social Medicine, 25*, 242-244.

Deubner, D. C., Wilkinson, W.E., Helms, M. J., Tyroler, H. A., & Hames, C. G. (1980). Logistic model estimation of death attributable to risk factors for cardiovascular disease in Evans County, Georgia. *American Journal of Epidemiology, 112*, 135-143.

Eide, E. E., & 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., & Gefeller, O. (2000). Re: Attributable risk in practice, (Letter, with replies of Walter, S. D., Wilson, P. D., Loffredo, C.A, Correa-villasenor A. & Ferencz, C.). *American Journal of Epidemiology, 151*, 208-212.

Feinstein, A. R. (1988) Scientific Standards in Epidemiologic Studies of the Menace of Daily Life. *Science, 242*, 1257-1263.

Feinstein, A. R. (1995). Biases introduced by confounding and imperfect retrospective and prospective exposure assessments, In: J. D. Graham, (Ed.), *The Role of Epidemiology in Regulatory Risk Assessment, Elsevier Science*, B. V., 29-42.

Gefeller, O., & Eide, G. E. (1993). Estimation of adjusted attributable risk in case-control studies. Letter. *Statistics in Medicine, 12*, 91-94

Gefeller, O., Land, M., & Eide, G. E. (1998). Averaging attributable fractions in the multifactorial situation: Assumptions and Interpretation. *Journal of Clinical Epidemiology, 51*, 437-441.

Gnizi, A. (1993). Decomposition of R^2 in multiple regression with correlated regressors. *Statistica Sinica, 3*, 407-420.

Gori, G. B. (1989). Epidemiology and Concept of Causation in Multifactorial Diseases. *Regulatory Toxicology and Pharmacology, 9*, 263-272.

Greenland, S. (1987). Variance estimators for attributable fraction estimates, consistent in both large strata and sparse data, Statistics in Medicine, 6, 701-708.

Greenland, S., and Drescher, K. (1993). Maximum Likelihood Estimation of the Attributable Fraction from Logistic Models. *Biometrics, 49*, 865-872.

Greenland, S., & Robins, J. M. (1988). Conceptual problems in the definition and interpretation of attributable fractions. *American Journal of Epidemiology, 128*, 1185-1197.

Hennekens, C. H., & Buring, J. E. (1987). *Epidemiology in Medicine*. Little, Brown & Co, Boston/Toronto.

Keller, A. Z., & Terris, M. (1965). The association of alcohol and tobacco with cancer of the mouth and pharynx. *American Journal of Public Health, 55*, 1578-1578.

Kelsey, J. L., Thompson, W. D., & Evans, A. S. (1986). *Methods in Observational Epidemiology.* New York: Oxford University Press.

Kleinbaum, D. G., Kupper, L. L., & Morgenstern, H. (1982). *Epidemiologic Research: Principles and Quantitative Methods*. Belmont, California: Lifetime Learning.

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

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

Kuritz, S. J., & Landis, J. R. (1988a). Summary attributable risk estimation from unmatched case-control data. *Statistics in Medicine, 7*, 507-517.

Kuritz, S. J., & Landis, J. R. (1988b). Attributable risk estimation from matched case-control data. *Biometrics, 44*, 355-367.

Land, M., & Gefeller O. (2000). A multivariate variant of the Shapley value for factorizing the risk of disease, In: Patrone, F., Garcia-jurado, I. & Tijs, S. (Eds.), *Game theory: Contributions from applied game theory*. Kluwer Academic Publishers.

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

Macmahon, B., & Pugh, T. F. (1970). *Epidemiology; Principles and Methods.* Little, Brown and Co, Boston.

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

Milliken, G. A., & Johnson, D. E. (1984). *Analysis of messy data volume 1, Designed Experiments*. Van Nostrand Reinhold, New York.

Pratt, J. W. (1987). Dividing the indivisible: using simple symmetry to partition variance explained, In: T. Pukkila and S. Puntanen (Eds.), *Proceedings of the Second International Conference in Statistics* (Edited by), 245-260. University of Tampere, Tampere, Finland.

Rothman, K. J. (1986). *Modern epidemiology*. Little, Brown & Co, Boston.

Rothman, K., & Keller, A. (1972). The effect of joint exposure to alcohol and tobacco on risk of cancer of the mouth and pharynx. *Journal of Chronic Diseases, 25*, 711-716.

Schlesselman, J. J. (1982). *Case-control studies: Design, conduct, analysis.* New York: Oxford University Press.

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

Walter, S. D. (1980). Prevention for Multifactorial Diseases'. *American Journal of Epidemiology, 112*, 409-416.

Walter, S. D. (1983). Effects of interaction, confounding and observational error on attributable risk estimation. *American Journal of Epidemiology, 117*, 598-604.

Wilson, P. D., Loffredo, C. A, Correa-villasenor, A., & Ferencz, C. (1998). Attributable fractions for cardiac malformations. *American Journal of Epidemiology, 148*, 414-423.

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