Annals of Epidemiology 25 (2015) 147–154

Contents lists available at ScienceDirect

Annals of Epidemiology

journal homepage: www.annalsofepidemiology.org

Original article

# A history of the population attributable fraction and related measures

Charles Poole MPH, ScD *

*Department of Epidemiology, University of North Carolina, Chapel Hill*

ARTICLE INFO

*Article history:*
Received 9 November 2014
Accepted 16 November 2014
Available online 11 December 2014

*Keywords:*
Population attributable fraction
Exposed attributable fraction
Rate difference
Population rate difference
Caseload difference

ABSTRACT

*Purpose:* Since Doll published the first PAF in 1951, it has been a mainstay. Confusion in terminology abounds with regard to these measures. The ability to estimate all of them in case-control studies as well as in cohort studies is not widely appreciated.
*Methods:* This article reviews and comments on the historical development of the population attributable fraction (PAF), the exposed attributable fraction (EAF), the rate difference (ID), the population rate (or incidence) difference (PID), and the caseload difference (CD).
*Results:* The desire for PAFs to sum to no more than 100% and the interpretation of the complement of a PAF as the proportion of a rate that can be attributed to other causes are shown to stem from the same problem: a failure to recognize the pervasiveness of shared etiologic responsibility among causes. A lack of appreciation that "expected" numbers of cases and deaths are not actually the numbers to be expected when an exposure or intervention appreciably affects person-time denominators for rates, as in the case of smoking and nonnormal body mass, makes many CD estimates inflated. A movement may be gaining momentum to shift away from assuming, often unrealistically, the complete elimination of harmful exposures and toward estimating the effects of realistic interventions. This movement could culminate in a merger of the academic concept of transportability with the applied discipline of risk assessment.
*Conclusions:* A suggestion is offered to pay more attention to absolute measures such as the rate difference, the population rate difference, and the CD, when the latter can be validly estimated and less attention to proportional measures such as the EAF and PAF.



## Introduction

The population attributable fraction (PAF) answers the following question: suppose every member of a population who was not in the most favorable level of an exposure or some other condition or event with regard to an adverse outcome had been shifted into that level. By what proportion would the entire population's rate, hazard, risk, prevalence, or caseload have been reduced?

The brief history to follow of the PAF and related measures begins in the early 1950s and ends, not altogether arbitrarily, at the close of the 1980s. The primary focus is on matters of interpretation, with some attention paid to the influence of rather basic features of study design and data analysis. Some statistical issues, such as those pertaining to sampling error and covariate adjustment, are not addressed.

## 1951 to 1953

### *The first PAF*

In 1951, Doll [1] estimated what appears to be the first published PAF in the epidemiologic literature. He used the cases from his preliminary (1948–1949) case-control study with Hill [2] to form the numerators for lung cancer incidence rates in Greater London. To obtain the denominators, he apportioned census figures by the smoking distribution in the study's control group. Within each age stratum, he multiplied the total person-time at risk by the rate among the nonsmokers "to estimate the number of cases that would have been expected to occur if the entire population were nonsmokers" [1]. Summing across the strata, Doll "calculated that 41 cases would have been expected to occur against the 533 which were actually observed. It is therefore estimated that more than 90% of the cases of carcinoma of the lung can be attributed to smoking" [1].

### *The second and third PAFs*

Two years later, Doll [3] generated another PAF using results from the full case-control study (1948–1952) [4]. This time, the

* Corresponding author. Department of Epidemiology, University of North Carolina, Chapel Hill, NC 27599-7435. Tel.: +1 919 966 9294; fax: +1 919 966 2089.
*E-mail address:* cpoole@unc.edu.

http://dx.doi.org/10.1016/j.annepidem.2014.11.015
1047-2797/

target population consisted of all adults aged 25 to 74 years in England and Wales in 1950. Within strata of age, gender, and locale, Doll used the controls from the case-control study as before to break down the person-time at risk by smoking. Then he applied the same approach to the stratified lung cancer deaths across England and Wales, apportioning them according to the smoking distribution of the study cases. An illustrative set of Doll's calculations for one subgroup is listed in Table 1.

After tallying the results across subgroups, Doll estimated that 1875 lung cancer deaths throughout England and Wales "would have been expected, in the absence of smoking" in comparison with the observed figure of 11,189: PAF = (11,189 − 1,875)/11,189 = 83%. This time, he focused his interpretation on the complement of the PAF, stating that "it is, therefore, concluded that about one in five of the lung cancer deaths in persons aged 25 to 74 in 1950, were attributable to causes other than lung cancer" [3].

### *More early PAFs*

Later that same year, Sadowsky et al. [5] showed that Doll had independently rediscovered the same approach Cornfield had described in 1951 [6] for using exposure distributions in a case-control study to estimate smoking-specific rates in an external target population (i.e., in a population other than the source population for the cases and controls). Sadowsky et al. and subsequent authors [7,8] referred to this approach as "the method of Cornfield." In using several case-control studies to generate smoking-specific lung cancer rates, they chose as their target population, as he had done, the combined population of the 10 metropolitan areas of the United States in the study [9–11] that later came to be known as the First National Cancer Survey [12].

### *Levin's PAF formula*

The year 1953 also saw the publication of Levin's [13] novel approach to estimating PAFs from rate ratios rather than from rates and rate differences. As his derivation was somewhat cryptic, the clearer one Leviton [14] gave two decades later is reiterated here.

Let $I$ represents the overall rate in a population, p the proportion of the person-time at risk in the level of a binary exposure variable with the higher rate, $I_1$ the rate in that level, and $I_0$ the rate in the level with the lower rate. If the outcome measure is a hazard, risk, or prevalence, $I$ may be replaced by $H$, $R$, or $P$, respectively. The overall rate is a weighted average of the exposure-specific rates, with weights proportional to the person-time at risk: $I = pI_1 + (1 - p)I_0$. This formula simplifies to $I = p(ID) + I_0$, where $ID = I_1 - I_0$ is the rate difference comparing the two exposure levels.

The population rate (incidence) difference is $PID = I - I_0 = p(ID)$. The PAF is the PID expressed as a proportion of $I$, $PAF = (I - I_0)/I$, or $PAF = p(ID)/[p(ID) + I_0]$. Dividing the numerator and denominator of this expression by $I_0$, we obtain $PAF = p(IR - 1)/[p(IR - 1) + 1]$, where $IR = I_1/I_0$. Illustrative calculations are shown in Table 1.

The closest Levin [13] came to naming the PAF was in a table legend, where he called it a "proportion attributable." He interpreted it, in his substantive example, as the "maximum proportion of lung cancer attributable to smoking." From the range of PAF estimates, he calculated with results from several case-control studies, Levin concluded that "tobacco smoking may be responsible for from 56% to 92% of lung cancer. If the latter figure is correct, elimination of smoking would almost eliminate lung cancer (other factors remaining the same). If 56% is nearer the true figure, then elimination of smoking would reduce lung cancer by about one half, if smoking is a truly causative agent" [13]. He interpreted the complement of the PAF similarly to Doll [3], observing that a PAF of 50% to 75% "would mean that environmental causes, other than smoking, should be looked for in 25% to 50% of the cases" [13].

Levin [13] also defined, for what seems to have been the first time, the exposed attributable fraction: $EAF = (IR - 1)/IR$. He gave no name to this measure and no derivation, but it is easily derived as the rate difference expressed as a proportion of the exposed rate: $EAF = ID/I_1$. Division of the numerator and denominator by $I_0$ yields $EAF = (IR - 1)/IR$. Levin described the EAF as "the proportion of lung cancer in smokers attributable to smoking." A calculation of this measure is illustrated in Table 1 as well.

## 1954 to 1959

Gefeller wrote that the PAF "fell into oblivion" [15] after 1953. Indeed, the citation to Levin's 1953 article [13] in a 1958 review of statistical methods in cancer research [16] was for another contribution entirely.

Another sign of early indifference to the PAF came in 1954, when the authors of two articles [7,8] used several case-control studies to break down an external target population's overall lung cancer rate

**Table 1**
Incident lung cancer cases and controls and lung cancer mortality in England and Wales, women aged 45 to 64 years, urban areas other than London, 1950, by smoking status [1]

| Group | Measure | Value |
|---|---|---|
| Case-control study | | |
| Cases | Proportion smokers | $p_c = 13/23 = 0.57$ |
| Controls | Proportion smokers | $p = 36/125 = 0.29$ |
| England and Wales | | |
| Nonsmokers | Lung cancer deaths | $A_0 = 0.43\ (533) = 232$ |
| | Person-years at risk | $T_0 = 0.71\ (3{,}507{,}000) = 2{,}496{,}984$ |
| | Rate* | $I_0 = 9.3$ |
| Smokers | Lung cancer deaths | $A_1 = 0.57\ (533) = 301$ |
| | Person-years at risk | $T_1 = 0.29\ (3{,}507{,}000) = 1{,}010{,}016$ |
| | Rate* | $I_1 = 29.8$ |
| | Rate difference* | $ID = 29.8 - 9.3 = 20.5$ |
| | "Expected" lung cancer deaths | $A_1^* = 1{,}010{,}016\ (0.000093) = 94$ |
| | Exposed attributable fraction | $EAF = 20.5/29.8 = (301 - 94)/301 = 69\%$ |
| All | Lung cancer deaths | $A = 533$ |
| | Person-years at risk | $T = 3{,}507{,}000$ |
| | "Expected" lung cancer deaths | $A^* = 3{,}507{,}000\ (0.000093) = 325$ |
| | Rate* | $I = 15.2$ |
| | Population rate difference* | $PID = 15.2 - 9.3 = 5.9$ |
| | "Expected" lung cancer deaths | $3{,}507{,}000\ (0.000093) = 325$ |
| | Population attributable fraction | $PAF = 5.9/15.2 = (533 - 325)/533 = 39\%$ |

* Lung cancer deaths per 100,000 person-years.

by smoking as Cornfield [6] and Doll [3] had done. Although the smoking-rates would have made it easy to produce PAF estimates (as well as estimates of the PID, ID, and EAF), none were calculated. One of these articles was by Levin himself [8].

## 1960 to 1972

### *The ID, EAF, PID, and PAF in an early textbook*

In their 1960 textbook, *Epidemiologic Methods* [17], MacMahon, Pugh, and Ipsen described all four measures of primary interest here. They called the ID the "attributable risk" and the PID the "attributable community risk" but gave no names to the EAF or the PAF. The authors wrote, "The concept of attributable community risk is useful in that it provides an estimate of the maximum reduction of a particular disease rate that might be expected if the specified exposure were removed" [17].

With lung cancer rates from an early report on Doll and Hill's cohort study of British male physicians [18], MacMahon et al. [17] computed estimates of all four measures: not just the ID and EAF but the absolute community measure, PID = 74 lung cancer deaths per 100,000 person-years, and the proportional community measure, PAF = 91%, as well. They also illustrated, with results from a study of prenatal X-ray exposure and childhood cancer by Stewart et al. [19], the same approach Doll [1] had taken to combining census and case-control study data to estimate exposure-specific rates, which then could be used to estimate the ID and PID along with the EAF and PAF.

### *A key article on shared causal responsibility*

In 1968, MacMahon published an article [20] of considerable relevance to Doll [3] and Levin's [8] view that the complement of the PAF for a given cause equals the proportion of a population's rate attributable to other causes. MacMahon's specific context was gene-environment interaction but his argument was completely general. He began by harkening back to the "nature versus nurture" debate in the 1930s in which Hogben had argued for the importance of nurture (i.e., nongenetic causes). Hogben's argument had stressed the role of shared causal responsibility that, as summarized later by Tabery [21], "arises from the combination of a particular genetic constitution with a particular kind of environment." Fisher, in making the case for nature (i.e., genetic causes), while acknowledging the existence of shared causal responsibility, had done his best to minimize its importance [21].

MacMahon [20] simply and eloquently noted the existence of human diseases such as phenylketonuria and favism (i.e., glucose-6-phosphate reductase deficiency) that have two known and universally necessary causes: one genetic and the other environmental. In any population with a nonzero rate of any of these diseases, the entirety of that rate is attributable to the genotype and to the environmental factor as well. That is, PAF = 100% for each of them.

### *The ID, EAF, PID, and PAF in another early textbook*

In the 1970 publication of *Epidemiology: Principles and Methods* by MacMahon and Pugh [22], the treatment of the ID, PID, EAF, and PAF was essentially the same as that of MacMahon et al. [17] a decade earlier, with two principal exceptions. One was a description of yet another rediscovery of Cornfield's [6] approach to using case-control study data to obtain estimates of the rates from which all four measures could be estimated in an external target population. The other was a change in the name of the PID, from the "attributable community risk" to the "population attributable risk" [22].

### *A rediscovery of Levin's formula*

In 1971, Cole and MacMahon [23] described how to estimate the EAF and PAF in a case-control study's source population without an estimate of the population's size or overall disease rate. They were apparently unaware that Levin had already shown the same thing and produced exactly the same formula for doing so in 1953 [9]. Cole quickly rectified the oversight in subsequent publications [24,25]. Cole and MacMahon [23] called the EAF the "attributable risk percent" and the PAF the "population attributable risk percent."

## 1973 to 1974

### *Lilienfeld's praise for the PAF*

The PAF received a big boost in 1973 when Lilienfeld, an early adopter, praised its utility in the publication of the American Public Health Association's first Wade Hampton Frost Lecture [26], which he had given the year before. He stated that the PAF is "important to the public health administrator," but did not elaborate on that use of the measure. His main focus was on what he saw as its implications for setting research priorities.

As Lilienfeld described it, the higher the PAF for a given cause, the higher the priority to be given to studying persons unexposed it in searching for additional causes. Conversely, the lower the PAF for a given cause, the higher the priority that should be accorded to studying exposed persons. An estimated PAF = 30% for having at least one of four leukemia risk factors suggested to him "that there are other factors involved but they are principally found among those individuals already exposed to these four risk factors" [26]. In contrast, he concluded that a PAF of 80% to 85% for smoking and lung cancer "indicates that other environmental factors probably play a relatively minor independent role." Lilienfeld would recommend to the epidemiologic researcher, therefore, "as an initial approach…to limit his studies to nonsmoking lung cancer patients" [26].

It is instructive to see how research teams of which Lilienfeld was a member interpreted PAF estimates. An estimated PAF of 17% for maternal radiation exposure and Down syndrome were reported without comment [27]. Estimates of PAF = 5% for more than 20 versus 20 or less diagnostic X-rays and PAF = 9% for more than 10 versus 10 or less diagnostic X-rays in relation to chronic myelocytic leukemia among adult men "indicate that, while small, the proportion of cases attributable to diagnostic irradiation is important. Still, most patients receiving irradiation do not develop chronic myelocytic leukemia. Thus, it seems reasonable to assume that one or more factors other than irradiation, possibly associated with sex and perhaps involving host susceptibility, exposures to viruses, and genetic characteristics are also related to leukemia" [28]. An estimate of PAF of 43% for benign prostatic hypertrophy and prostate cancer among men age 50 years and older in countries with long life expectancy was reported without comment other than, "Our findings warrant further epidemiological and laboratory investigations" [29].

In the inaugural Frost Lecture [26], Lilienfeld initiated some terminologic confusion. As noted previously, MacMahon and Pugh [22] had already followed MacMahon et al. [17] in calling the ID the "attributable risk," had changed their preferred name for the PID from the "attributable community risk" [17] to the "population attributable risk" [22] and had given no names to the EAF or the PAF. Lilienfeld [26] began by suggesting that Levin [9] had called the PAF the "attributable risk," which he had not. (As noted previously, Levin had called it a "proportion attributable.") Then, citing MacMahon and Pugh [22], Lilienfeld took their name for the PID, the "population attributable risk" and transferred it to the PAF.

*A clarifying derivation*

In 1974, Leviton [14], while clarifying Levin's [13] derivation of the PAF and showing its equivalence with that of MacMahon and Pugh [22], correctly attributed to those authors the terms they had actually used: "attributable risk" for the ID and "population attributable risk" for the PID. Unfortunately, Leviton repeated the incorrect assertion [26] that Levin [13] had called the PAF an "attributable risk."

*A new PAF formula*

The year 1974 also saw a publication in which Miettinen [30] fleshed out the theoretical and statistical basis for a new formula that Panayotou et al. [31] had introduced 2 years before for estimating the PAF in a study's source population: PAF = $p_c$EAF, where $p_c$ is the proportion of the cases in the exposure level with the higher rate. Miettinen [30] called the PAF the "etiologic fraction" and the EAF the "etiologic fraction…for those with the marker" of increased risk. He reinforced the assumption that the attributable fraction of a risk (or, as in his two examples, a prevalence [31] and a rate [32]) is equivalent to the fraction of the cases etiologically attributable to the exposure.

## 1975 to 1983

*An early PAF generalization in theory*

In articles in 1975 [33] and 1976 [34], Walter continued with the assumption that attributable fraction of incidence rates are equivalent with attributable fractions of incident caseloads. In the longer article [34], he examined all the measures under consideration here except the PID. He called the ID "Berkson's simple difference" and, following Lilienfeld, called the PAF "the attributable risk" and the EAF "the attributable risk among the exposed."

Walter [34] stated that exposure-specific rates and, therefore, the ID are not "estimable directly" in case-control studies. He referred briefly to the approach, presumably indirect, by which Cole and MacMahon had made such estimates [23].

In Section 6.2 of the second article [34], Walter broke important new ground. He considered the estimation of PAFs for shifts in polytomous exposures less dramatic than moving every single member of a population to the exposure level with the lowest rate.

*An early PAF generalization in practice*

A few years later, Ouellet et al. [35], apparently unaware of this part of Walter's article [34], estimated the measure he had described. First, they used Levin's formula to estimate PAF = 56% in black males and PAF = 52% in black females for stroke incidence in an urban community in Baltimore, Maryland in connection with not receiving hypertension screening. Then the authors relaxed the unrealistic assumption of universal public screening by taking into account private screening and less than complete willingness to attend a public clinic, receive treatment, and persist with it. The resulting PAFs for a public screening program, as opposed to screening itself, were 32% in males and 14% in females.

*More generalized PAF theory*

In 1982, Morgenstern and Bursic [36], acknowledging their intellectual debt to Ouellet et al. [35], formalized the measure the latter authors had estimated and christened it the "potential impact fraction." In this framework, the EAF and PAF become special cases of the potential impact fraction when everyone in a target population actually can be shifted to the exposure level with the lowest rate.

*A famous set of PAFs*

A year before, in 1981, Doll and Peto had come out with one of the most famous sets of PAF estimates ever published, in "*The Causes of Cancer: Quantitative Estimates of Avoidable Risks of Cancer in the United States Today*" [37]. In a section on shared causal responsibility, they noted that it can easily make a set of PAF estimates sum to more than 100%. Nonetheless, they stated that "the amount of 'double counting' in the estimates which we shall present of the proportions of present-day U.S. cancer deaths that can be prevented by various separate methods is small. We shall therefore ignore the anomaly in the rest of this paper and hope that, once it has been pointed out, no one will fall into the trap of adding together proportions that are not, in fact, mutually exclusive" [37]. Their PAFs summed to approximately 96%.

*The dawn of regulatory risk assessment*

Meanwhile, regulatory agencies and their scientific advisors began to look at population-level causal attribution from a very different perspective. They did not start with epidemiologic studies and wonder how to make the results relevant to public health policy making. Instead, their starting point was their responsibility for developing and implementing interventions to improve population health. They then looked for health research, especially epidemiologic research when available, to help estimate the impacts of those interventions. Thus was born the discipline of risk assessment.

Some early regulatory risk assessments contemplated exposure elimination, as in a 1980 proposal to remove friable (e.g., sprayed on) asbestos insulation from all US school buildings [38]. Others aimed at exposure reduction, as in Nicholson's 1983 risk assessment for a proposal to lower the United States workplace asbestos exposure standard from 2.0 fibers/cm$^3$ to 0.1 fibers/cm$^3$ [39]. These risk assessments tended to focus on measures such as the ID and PID and not on the EAF or PAF.

A key event in the early days of this field was the 1983 publication of the National Research Council's *Risk Assessment in the Federal Government: Managing the Process* [40]. This landmark publication became known in risk-assessment circles as the Red Book.

*Recognition that "expected" cases are not expected cases*

It is difficult to say when it was first recognized that an "expected" caseload calculated by multiplying exposed person-time by an unexposed rate, such as the "expected" value of 325 lung cancer deaths in Table 1, is not actually the number of expected cases that would have occurred if an exposed population had been shifted to unexposed. There is no indication of any awareness of this point in very early treatments, which date back to Westergaard in 1916 [41] and Woodbury in 1922 [42]. It had become well, if not widely, known by 1983, when Berry [43] wrote of "the fact that what, by long usage, is referred to as the 'expected number of deaths' is not equal to the number of subjects expected to be dead."

The reason is that, if an exposure affects a rate, it affects not only the incident cases or deaths in the numerator but the person-time at risk in the denominator as well. Exposure effects on competing risks can amplify this discrepancy. In Table 1, for instance, the smokers experienced 1,010,016 person-years at risk. In reality, we know today that they would have experienced substantially more person-time at risk if they had never smoked. But the same figure,

1,010,016 person-years, is multiplied by the rate among the nonsmokers. The "expected" number of lung cancer deaths so calculated can be divided into the observed number to calculate the rate ratio estimate: IR = 29.8/9.3 = 301/94 = 3.2. Similarly, the observed and "expected" lung cancer deaths can be used to calculate the EAF estimate: EAF = (29.8 − 9.3)/29.8 = (301 − 94)/301 = (3.2 − 1)/3.2 = 69%.

Nonetheless, considerably more than 1,010,016 person-years at risk would have been experienced in the absence of smoking and, therefore, 94 substantially overestimates the actual number of lung deaths to have been expected if no one had ever smoked. Hence, although (301 − 94)/301 = 69% is a convenient way to calculate the attributable fraction of the lung cancer rate among the smokers, it is not the attributable fraction of the lung cancer deaths among the smokers.

Berry [43] went on to state, "It would be possible to calculate the expected number of deaths using, instead of observed years at risk, expected years at risk based on the reference death rates. This would be the life-table approach, and an expectation calculated in this way would not be a random variable and would equal the number of subjects expected to be dead."

## 1984 to 1989

### *More PAF theory and generalized PAF practice*

In 1986, Rothman [44] criticized Doll and Peto [37] for ignoring shared causal responsibility and thereby forcing the sum of their PAFs for causes of cancer not to exceed 100%. In an update, Peto raised the ceiling to 200% [45].

The 1980s also saw other researchers begin to follow in the footsteps of Ouellet et al. [35] and estimate attributable proportions for shifts in exposure distributions more realistic than moving everyone to the exposure level with the lowest rate. In one article, Wahrendorf [46] estimated proportional rate reductions for interventions such as moving 10% of those in each category of saturated fat intake down to the next lower category. Such examples were notable for their rarity.

### *Why attributable cases are not necessarily etiologic cases*

In 1988 and 1989, Greenland and Robins published a series of articles identifying two additional prerequisites for an attributable fraction of a risk or caseload to be an etiologic fraction [47–49]. First, the exposure must not hasten the incidence time of any incident outcomes that were already going to occur in the risk period. The exposure causes these cases, but they do not cause the risk to be increased. Second, there must not be any cases prevented during the risk period to offset caused cases. In some settings, either or both of these assumptions is unwarranted.

## Comment

Although by the 1990s, the basic elements of the points emphasized here were in place in the literature on the PAF and related measures, that literature has continued to blossom. Particularly, notable contributions have been made by Greenland [50–56], Rockhill et al. [57,58].

### *Terminology*

The terminologic confusion in this area is most unfortunate. Gefeller [15] counted 16 different names for the PAF. The current edition of an epidemiologic dictionary has 12 entries that start with the word "attributable" [59]. In about two-thirds of the articles in a systematic review of 334 publications in 1966 to 1996 reporting variations on this theme, "no precise reference was made and no explicit definition was given" [60].

Particularly, lamentable is the tradition of calling a proportion or fraction of a risk (or of a rate, prevalence, or hazard). A fraction of a risk is not a risk and should not be called one [55,58]. As shown previously, although this terminologic blunder is often mistakenly traced to Levin [13], it started with Lilienfeld [26].

### *Summing PAFs and the complement of a PAF*

It is wrong to assume that the complement of a PAF equals the proportion of a population in which other causes are operating. It is also a mistake to assume to cap the sum of a set of PAFs at 100%. These errors are two sides of the same coin. They stem from the exceedingly deleterious assumption that a lung cancer that was caused by smoking, for instance, could not have been caused by anything else as well.

That a set of PAFs can add up to well more than 100% was recently described as one of the measure's "key limitations" [61]. To the contrary, a desire for the sum of a set of PAFs to be capped at 100% is a key limitation of understanding. Shared responsibility among causes requires that PAFs can add up to more than 100%. Shared causal responsibility is not a rarity, as special case, or an "anomaly" [37]. It is a fact of etiologic life.

Consider how one pair of experts interpreted the complement of a PAF. In 2001, Magnus and Beaglehole [62] argued that if established risk factors such as hypertension, smoking, and blood lipids have a combined PAF of 50% for coronary heart disease, the implication would be "that another unexplained 50% exists, that is, that important undiscovered risk factors make up this apparent deficit, and that high priority should be given to further social, molecular, or other basic research." But if, as the authors contended, the combined PAF for established risk factors is 75%, "whether another 25% exists, and how much of it we could expect to explain, are matters for debate." A year later, the same two authors [63] called not just for implementation of population-level interventions to improve population distributions of the established risk factors but for a reduced priority on the search for new causes as well.

We may contrast this with view with Lilienfeld's argument that the greater the PAF for a given cause, the greater the priority to be placed on searching for new risk factors among persons unexposed to that cause. This argument runs aground when taken to its logical extreme and we find a cause for which the PAF = 100%. At that point, as with the PAF of 100% for human immunodeficiency virus and AIDS, our attention naturally correctly turns to the search for new causes among exposed persons. A more tenable perspective would be that the greater the PAF for a given cause or set of causes, the more advisable it is to look for new causes among exposed persons, especially if the cause or cause(s) with the high PAF are resistant to favorable change at the population level.

### *Caseload effects*

Smoking is not the only exposure that can affect the denominator of an incidence rate appreciably while affecting the numerator of the rate. In estimate after estimate of the effects of obesity, overweight, and underweight on overall and cause-specific mortality in large populations such as the United States [64–67], the starkly unrealistic assumption has been made that the person-time at risk would have remained unchanged if no one had been allowed to stray outside the narrow band of body mass index that is defined as normal. To the contrary, the person-time at risk and the expected (as opposed to "expected") number of deaths would have been considerably greater and, as a

consequence, the actually attributable numbers of deaths would have been substantially smaller than the values reported in these articles.

For this reason, no attempt has been made to estimate the attributable number of deaths liver cancer deaths in the example in Table 2 [68]. Elimination of hepatitis B infection (as measured by the presence of the hepatitis B surface antigen) would have too great an effect on the person-time at risk for the "expected" liver cancer deaths to be a reasonable approximation of the liver cancer deaths that actually would be expected.

Sometimes an exposure can appreciably affect even the denominator of a risk or a prevalence. The most obvious example would be an effect on a postnatal outcome (e.g., infant mortality) of artificial insemination or *in vitro* fertilization, if that intervention is compared with no infertility treatment. Another example is provided by the study by Panayotou et al. [31] of one or more prior induced abortions in relation to the prevalence of ectopic pregnancy. If the prior pregnancies had not been terminated medically, parity would have increased. By the well-known inverse effect of parity on subsequent pregnancies, there would been fewer pregnancies in the population cross-section.

Thus, the caseload difference (CD) estimates in Tables 3 and 4 are reported with some trepidation. If the elimination of waist size greater than 102 cm in Swedish men (Table 3) [69] or of *Schistosoma japonicum* infection in Filipino children (Table 4) [70] would have materially altered the numbers of persons available for the cross-sectional samples in those studies, the CDs in Tables 3 and 4 are overestimates.

### *Intervention connections*

As we have seen, Levin [13], MacMahon et al. [17], and MacMahon and Pugh [22] all wisely used the word "maximum" to qualify the purported relevance of the EAF and PAF to public health interventions. In the textbook by Gordis, which devotes an entire chapter to the value of the PAF in "estimating the potential for prevention," starts off with, "Assume that we know how to eliminate smoking" [71]. Although that assumption is never questioned or relaxed, qualifiers such as "at best" and "the best we could hope to achieve" are given along the way.

Other enthusiastic PAF users are not as circumspect. Consider these claims: "Mounting evidence indicates that more than 50% of cancer could be prevented if our current knowledge of risk factors were successfully implemented to reduce risk factor prevalence" [72]. "As many as half of all cancers could be prevented if we just implemented what we already know about cancer's causes" [73]. "We currently have sufficient knowledge of cancer causes and prevention to reduce cancer burden in the United States by over 50%" [74].

Here is the evidence: "How did we determine the extent to which mortality can be reduced? We began by identifying the lowest rates for various types of cancer among large international populations that keep reliable figures on death from cancer. The incidence of many of the most common cancers in the U.S. and Europe is much lower in Japan and China. To compile a list of estimated 'baseline' cancer incidences, then, we chose the lowest rate for each type of cancer from among the data for the U.S., Japan and China. Then we calculated the difference between the highest rate and the baseline. From these comparisons, we conclude that it should be possible to reduce cancer mortality by approximately 60 percent in the U.S." [75]. Apparently, these experts know how to import low cancer rates from China and Japan into the United States. Implementation of this knowledge is all that is needed.

Another example: "From our results, we estimate that 90,000 deaths due to cancer could be prevented in the United States each year if men and women could maintain normal weight. It is unlikely

**Table 2**
Primary liver cancer mortality by gender and baseline hepatitis B surface antigen (HBsAg) status, baseline age 25 to 69 years, Haimen, China, 1992 to 2006 [68]

| Group and measure | Men | Women |
|---|---|---|
| HBsAg negative | | |
| Liver cancer deaths | 423 | 51 |
| Person-years at risk | 660,324 | 317,464 |
| Rate* | 64 | 16 |
| HBsAg positive | | |
| Liver cancer deaths | 1113 | 216 |
| Person-years at risk | 107,511 | 62,415 |
| Rate* | 1035 | 346 |
| All persons | | |
| Liver cancer deaths | 1536 | 267 |
| Person-years at risk | 767,835 | 379,888 |
| Rate* | 200 | 70 |
| Rate difference* | 971 | 330 |
| PID* | 136 | 19 |
| EAF (%) | 86 | 95 |
| PAF (%) | 37 | 77 |

* Liver cancer deaths per 100,000 person-years.

**Table 3**
Type 2 diabetes prevalence, men aged 35 to 56 years, outskirts of Stockholm, Sweden, 1996 to 1998, by family history of type 2 diabetes and waist circumference [69]

| Group and measure | Positive family history | Negative family history |
|---|---|---|
| Waist circumference <102 cm | | |
| Diabetes cases | 31 | 8 |
| Persons | 1235 | 1280 |
| Prevalence (%) | 2.5 | 0.6 |
| Waist circumference ≥102 cm | | |
| Diabetes cases | 20 | 6 |
| Persons | 178 | 130 |
| Prevalence (%) | 11.2 | 4.6 |
| All men | | |
| Diabetes cases | 51 | 14 |
| Persons | 1413 | 1410 |
| Prevalence (%) | 3.6 | 1.0 |
| Prevalence difference (%) | 8.7 | 4.0 |
| Population prevalence difference (%) | 1.1 | 0.4 |
| CD | 16 | 5 |
| EAF (%) | 78 | 86 |
| PAF (%) | 30 | 37 |

**Table 4**
Anemia prevalence, Leyte, The Philippines, aged 7 to 18 years, circa 2000, by hookworm and *Schistosoma japonicum* infection [70]

| Group and measure | Hookworm positive | Hookworm negative |
|---|---|---|
| *S. japonicum* positive | | |
| Anemia cases | 11 | 31 |
| Children | 15 | 116 |
| Prevalence (%) | 73 | 27 |
| *S. japonicum* negative | | |
| Anemia cases | 11 | 30 |
| Children | 28 | 219 |
| Prevalence (%) | 39 | 14 |
| All children | | |
| Anemia cases | 22 | 61 |
| Children | 43 | 337 |
| Prevalence (%) | 51 | 18 |
| Prevalence difference (%) | 34 | 13 |
| Population prevalence difference (%) | 12 | 4 |
| CD | 5 | 15 |
| EAF (%) | 46 | 49 |
| PAF (%) | 23 | 25 |

that this goal can be achieved without concerted effort and substantial investment on the part of policymakers, educators, clinicians, employers, and schools to promote physical activity and healthful dietary practices as a cultural norm" [76]. A more realistic view is that only a totalitarian police state could achieve this goal.

Some PAF estimates would be laughable if they were not offered with such seriousness. One case-control study of endometrial cancer [77] reported a PAF of 2% for a positive family history. If all family history of endometrial cancer were eliminated, the only women who would develop that disease would be those whose mothers and sisters, if any, have all died without having developed it themselves. That PAF would be considerably greater than 2%.

In response to an article by a deputy editor of a major journal extolling the PAF for its ability to "predict the impact of medical and public health interventions on the health status of a population" [78], a reader wrote in to note the value of estimating the effects of partial reductions in the prevalence of harmful exposures, when that is all that can be achieved [79]. As of September 22, 2014, according to the Web of Science, the article had been cited 90 times, and the letter had been cited twice.

Many thousands of readers learned from a popular book that "a healthy diet teamed up with regular exercise and no smoking can eliminate 80 percent of heart disease and 70 percent of some cancers" [80]. Many fewer read book review's accurate characterization of this claim as "hyperbole" [81].

More optimistic signs are present, however. One is the continued progress of risk assessment methods. The Red Book [40] has been supplanted by a Silver Book [82] to summarize the advancing state of that art. Another hopeful trend is the growing interest in estimating effects of realistic interventions [83–86]. The current buzzword for this activity in academic circles is "transportability" [87]. One can envision a time at which it will merge with risk assessment, for they are in principle one and the same.

### *Attributable fractions or attributable absolutes?*

The preference for proportional measures such as the EAF and PAF over absolute measures such as the ID and PID forms a tributary feeding into the epidemiologic river I have called "risk relativism" [88]. To the extent that any of these measures actually can guide the setting of public health policy priorities, it might be worthwhile to reconsider that preference.

It is often pointed out, as Walter [34] did, that the rates needed to calculate measures such as the ID and PID cannot be estimated "directly" in case-control studies [61]. The matter is usually left there, as he left it. One might ask why more consideration is not given to estimating these measures "directly" in cohort studies and "indirectly" in case-control studies, as Cornfield [6], Doll [1,3], Sadowski et al. [5], Breslow et al. [7], Levin [8], MacMahon et al. [17], MacMahon and Pugh [22], and Cole and MacMahon [23] did. Today, a wide variety of underutilized study designs and statistical methods are available for doing so [89–97], but in practice the art seems all but lost [98–100].

To contrast the absolute measures with the proportional measures, consider the examples in Tables 2 and 4 In Table 2, the EAF and PAF are higher among women, but the ID and PID are higher among men. In Table 3, the EAF and PAF are higher when a family history is negative, but the prevalence difference, population prevalence difference, and CD are higher among men with a positive family history. In Table 4, where the EAF and PAF are about the same among the children with and without hookworm infection, the prevalence difference and population prevalence difference are much higher in the presence of hookworm infection but the CD is much higher in its absence. Who would argue that the EAF and PAF are the better guides to public health importance in these examples? Not I.

## References


[1] Doll R. On the aetiology of cancer of the lung. Acta Union Internat Contra Cancrum 1951;7(1 Spec. No.):39–50.
[2] Doll R, Hill AB. Smoking and carcinoma of the lung: preliminary report. BMJ 1950;30(2):739–48.
[3] Doll R. Mortality from lung cancer among non-smokers. Br J Cancer 1953;7:303–12.
[4] Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. BMJ 1952;2:1271–86.
[5] Sadowsky DA, Gilliam AG, Cornfield J. Statistical association between smoking and carcinoma of the lung. J Natl Cancer Inst 1953;13:1237–58.
[6] Cornfield J. A method of estimating comparative rates from clinical data: application to cancer of the lung, breast, and cervix. J Nat Cancer Inst 1951;11:1269–75.
[7] Breslow L, Hoaglin L, Rasmussen G, Abrams HK. Occupations and cigarette smoking as factors in lung cancer. Am J Public Health 1954;44:171–81.
[8] Levin ML. Etiology of lung cancer: present status. New York State Med J 1954;54:769–77.
[9] Dorn HF. Illness from cancer in the United States. Public Health Rep 1944;59:33–48.
[10] Dorn HF. Illness from cancer in the United States (con.). Public Health Rep 1944;59:65–77.
[11] Dorn HF. Illness from cancer in the United States. Public Health Rep 1944;59:97–115.
[12] Lilienfeld DE. Harold Fred Dorn and the First National Cancer Survey (1937-1939): the founding of modern cancer epidemiology. Am J Public Health 2008;98:2150–8.
[13] Levin ML. The occurrence of lung cancer in man. Acta Union Internat Contra Cancer 1953;9:531–41.
[14] Leviton A. Definitions of attributable risk (letter). Am J Epidemiol 1973;98:231.
[15] Gefeller O. An annotated bibliography on the attributable risk. Biom J 1992;34:1007–12.
[16] Lombard HL. Statistical research in cancer control: a review. Cancer Res 1958;18:4–19.
[17] MacMahon B, Pugh TF, Ipsen J. Epidemiologic methods. Little, Brown and Company; 1960. p. 229–30. 260–261.
[18] Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking: a second report on the mortality of British doctors. BMJ 1956;2:1071–81.
[19] Stewart A, Webb J, Hewitt D. A survey of childhood malignancies. BMJ 1958;1:1495–508.
[20] MacMahon B. Gene-environment interaction in human disease. J Psychiatr Res 1968;6:393–402.
[21] Tabery J. R.A. Fisher, Lancelot Hogben, and the origin(s) of genotype-environmental interaction. J Hist Biol 2008;41:717–61.
[22] MacMahon B, Pugh TF. Epidemiology: principles and methods. Little, Brown and Company; 1970. p. 233–4. 268–275.
[23] Cole P, MacMahon B. Attributable risk percent in case-control studies. Br J Prev Soc Med 1971;25:242–4.
[24] Schmauz R, Cole P. Epidemiology of cancer of the renal pelvis and ureter. J Natl Cancer Inst 1974;52:1431–4.
[25] Cole P. Primary prevention of cancer. Bull NY Acad Med 1975;51:75–9.
[26] Lilienfeld AM. Epidemiology of infectious and non-infectious disease: some comparisons. Am J Epidemiol 1973;97:135–47.
[27] Cohen BJ, Lilienfeld AM. The epidemiological study of mongolism in Baltimore. Ann NY Acad Sci 1970;171:320–7.
[28] Gibson R, Graham S, Lilienfeld A, Schuman L, Dowd JE, Levin ML. Irradiation in the epidemiology of leukemia among adults. J Natl Cancer Inst 1972;48:301–11.
[29] Armenian HK, Lilienfeld AM, Diamond EL, Bross IDJ. Relation between prostatic hyperplasia and cancer of the prostate: a prospective and retrospective study. Lancet 1974;2:115–7.
[30] Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. Am J Epidemiol 1974;99:325–33.
[31] Panayotou PP, Kaskarelis DB, Miettinen OS, Kalandidi AK. Induced abortion and ectopic pregnancy. Am J Obstet Gyecol 1972;114:507–10.
[32] Jick H, Miettinen OS, Neff RK, Shapiro S, Heinonen OP, Slone D. Coffee and myocardial infarction. N Engl J Med 1973;289:63–7.
[33] Walter SD. The distribution of Levin's measure of attributable risk. Biometrika 1975;62:371–4.
[34] Walter SD. The estimation and interpretation of attributable risk in health research. Biometrics 1976;32:829–49.
[35] Ouellet RP, Apoltolides AY, Entwisle G, Hebel JR. Estimated community impact of hypertension control in a high risk population. Am J Epidemiol 1979;109:531–8.
[36] Morgenstern H, Bursic ES. A method for using epidemiologic data to estimate the potential impact of an intervention on the health status of a target population. J Commun Health 1982;7:292–309.
[37] Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. J Natl Cancer Inst 1981;66:1193–308.

[38] Office of Pesticides and Toxic SubstancesUnited States Environmental Protection Agency. Health effects and magnitude of exposure. In: Support document: asbestos-containing materials in schools. United States Government Printing Office; 1980. EPA-560/12-80-003.
[39] Nicholson W. Quantitative risk assessment for asbestos-related cancers. Occup Saf and Health Adm Off Carcinogen Stand. OSHA asbestos rulemaking docket H033C Exhibits 84–349 and 84-392, August 1983, revised October 1983.
[40] National Research Council. Risk assessment in the federal government: managing the process. National Academy Press; 1983.
[41] Westergaard H. Scope and method of statistics. Publ Am Statist Assoc 1916;15:229–76.
[42] Woodbury RM. Westergaard's method of expected deaths as applied to the study of infant mortality. J Am Statist Assoc 1922;18:366–76.
[43] Berry G. The analysis of mortality by the subject-years method. Biometrics 1983;39:173–84.
[44] Rothman KJ. Modern epidemiology. First edition. Little, Brown and Company; 1986.
[45] Peto R. The preventability of cancer. In: Vessey MP, Gray M, editors. Cancer risks and prevention. Oxford University Press; 1985. p. 1–14.
[46] Wahrendorf J. An estimate of the proportion of colorectal and stomach cancers which might be prevented by certain changes in dietary habits. Int J Cancer 1987;40:625–8.
[47] Greenland S, Robins J. Conceptual problems in the definition and interpretation of attributable fractions. Am J Epidemiol 1988;128:1185–97.
[48] Robins JM, Greenland S. Estimability and estimation of excess and etiologic fractions. Stat Med 1989;8:845–9.
[49] Robins JM, Greenland S. The probability of causation under a stochastic model for individual risks. Biometrics 1989;45:1125–38.
[50] Greenland S. The relation of the probability of causation to the relative risk and the doubling dose: a methodologic error that has become a social problem. Am J Public Health 1999;89:1166–9.
[51] Greenland S. Attributable fractions: bias from broad definition of exposure. Epidemiology 2001;12:518–20.
[52] Greenland S. Choosing sensible effect measures for practice and policy. J Clin Epidemiol 2003;56:392–3.
[53] Greenland S. Epidemiologic measures and policy formulation: lessons from potential outcomes (with discussion). Emerg Themes Epidemiol 2005;2:1–4.
[54] Greenland S, Robins JM. Epidemiology, justice, and the probability of causation. Jurimetrics 2000;40:321–40.
[55] Greenland S, Rothman KJ, Lash TL. Measures of effect and measures of association. Chapter 4. In: Rothman KJ, Greenland S, Lash TL, editors. Modern epidemiology. Third edition, 54. Lippincott-Raven; 2008. p. 62–7.
[56] Greenland S. Applications of stratified analysis methods. Chapter 16. In: Rothman KJ, Greenland S, Lash TL, editors. Modern epidemiology. Third edition. Lippincott-Raven; 2008. 295–298.
[57] Rockhill B, Weinberg CR, Newman B. Population attributable fraction estimation for established breast cancer risk factors: considering the issues of high prevalence and unmodifiability. Am J Epidemiol 1998;147:826–33.
[58] Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. Am J Public Health 1998;88:15–9.
[59] Porta M, editor. A dictionary of epidemiology. Fifth edition. International Epidemiological Association; 2008. p. 9–12.
[60] Uter W, Pfahlberg A. The application of methods to quantify attributable risk in medical practice. Stat Methods Med Res 2001;10:231–7.
[61] Keyes KM, Galea S. Epidemiology matters: a new introduction to methodological foundations. Oxford University Press; 2014. p. 84–90.
[62] Magnus P, Beaglehole R. The real contribution of the major risk factors to the coronary epidemics: time to end the "only 50%" myth. Arch Intern Med 2001;161:2657–60.
[63] Beaglehole R, Magnus P. The search for new risk factors for coronary heart disease: occupational therapy for epidemiologists? Int J Epidemiol 2002;31:1117–22.
[64] Allison DB, Fontaine KR, Manson JE, Stevens J, VanItallie TB. Annual deaths attributable to obesity in the United States. JAMA 1999;282:1530–8.
[65] Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med 2003;348:1625–38.
[66] Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. JAMA 2004;291:1238–45 (errata: JAMA 2005;293:293-294).
[67] Flegal KM, Graubard BI, Williamson DF, Gail MH. Excess deaths associated with underweight, overweight, and obesity. JAMA 2005;292:1861–7.
[68] Wang N, Zhang Y, Yu X, Lin W, Chen Y, Jiang Q. Sex-modified effect of hepatitis B virus infection on mortality from primary liver cancer. Am J Epidemiol 2009;169:990–5.
[69] Hilding A, Eriksson AK, Agardh EE, Grill V, Ahlbom A, Efendic S, et al. The impact of family history of diabetes and lifestyle factors on abnormal glucose regulation in middle-aged Swedish men and women. Diabetologia 2006;49:2589–98.
[70] Ezeamama AE, McGarvey ST, Acosta LP, Zierler S, Manalo DL, Wu HW, et al. The synergistic effect of concomitant schistosomiasis, hookworm, and trichuris infections on children's anemia burden. Plos Negl Trop Dis 2008;2(6):e245.
[71] Gordis L. Epidemiology. Third edition. Elsevier Saunders; 2004. p. 191–8.
[72] Colditz GA, Atwood KA, Emmons K, Monson RR, Willett WC, Trichopoulos D, et al. Harvard report on cancer prevention. Volume 4. Harvard Cancer Risk Index. Cancer Causes Control 2000;11:477–88.
[73] Colditz G. The science of cancer (letter). The New Yorker July 16, 2001;7.
[74] Colditz GA, Samplin-Saldago M, Tomeo Ryan C, Dart H, Fisher L, Tokuda A, Rockhill B. Harvard Report on Cancer Prevention, Volume 5: fulfilling the potential for cancer prevention: policy approaches. Cancer Causes Control 2002;13:199–212.
[75] Willett WC, Colditz GA, Mueller NE. Strategies for minimizing cancer risk. Sci Am 1996;275:88–95.
[76] Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med 2003;348:1625–38.
[77] Hemminki K, Bermejo JL, Granström C. Endometrial cancer: population attributable risks from reproductive, familial and socioeconomic factors. Eur J Cancer 2005;41:2155–9.
[78] Northridge ME. Public health methods: attributable risk as a link between causality and public health action. Am J Public Health 1995;85:102–4.
[79] Hill GB. The value of the population attributable risk percentage (letter). Am J Public Health 1996;86:1483.
[80] Willett WC. Eat, drink, and be healthy: the Harvard Medical School guide to healthy eating. Free Press; 2005. p. 15.
[81] Swartzberg J, Margen S. Book review: "Eat, Drink, and Be Healthy: The Harvard Medical School Guide to Healthy Eating," by Walter C. Willett. Am J Epidemiol 2001;154:1160–1.
[82] National Research Council. Science and decisions: advancing risk assessment. National Academy Press; 2009.
[83] Hernán MA. Hypothetical interventions to define causal effects: afterthought or prerequisite? Am J Epidemiol 2005;162:618–20.
[84] Spiegelman D, Hertzmark E, Wand HC. Point and interval estimates of partial population attributable risks in cohort studies: examples and software. Cancer Causes Control 2007;18:571–9.
[85] Loehr LR, Rosamond WD, Poole C, McNeill AM, Chang PP, Dewal A, et al. The potentially modifiable burden of incident heart failure due to obesity: the Atherosclerosis Risk in Communities Study. Am J Epidemiol 2010;172:781–9.
[86] Westreich D. From exposures to population interventions: pregnancy and response to HIV therapy. Am J Epidemiol 2014;179:797–806.
[87] Hernán MA, VanderWeele TJ. Compound treatments and transportability of causal inference. Epidemiology 2011;22:368–77.
[88] Poole C. On the origin of risk relativism. Epidemiology 2010;21:3–9.
[89] Benichou J, Gail M. Methods of inference for estimates of absolute risk derived from population-based case-control studies. Biometrics 1990;51:182–94.
[90] Nurminen M. Assessment of excess risks in case-base studies. J Clin Epidemiol 1992;45:1081–92.
[91] Carroll RJ, Wang S, Wang CY. Prospective analysis of logistic case-control studies. J Am Statist Assoc 1995;90:157–69.
[92] Wacholder S. The case-control study as data missing by design: estimating risk differences. Epidemiology 1996;7:144–50.
[93] Langholz B, Borgan O. Estimation of absolute risk from nested case-control data. Biometrics 1997;53:767–74.
[94] Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. Am J Epidemiol 2004;160:301–5.
[95] Langholz B. Case-control study, nested. Encyclopedia of Biostatistics. Second edition, Volume 1; 2005. p. 646–55.
[96] Langholz B. Use of cohort information in the design and analysis of case-control studies. Scand J Statist 2007;34:120–36.
[97] Kovalchik SA, De Matteis S, Landi MT, Caporaso NE, Varadhan R, Consonni D, et al. A regression model for risk difference estimation in population-based case-control studies clarifies gender differences in lung cancer risk of smokers and never smokers. BMC Med Res Methodol 2013;13:143.
[98] Habel LA, Shak S, Jacobs MK, Capra A, Alexander C, Pho M, et al. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. Breast Cancer Res 2006;8:R25.
[99] Lacey Jr JV, Sherman ME, Rush BB, Ronnett BM, Ioffe OB, Duggan MA, et al. Absolute risk of endometrial carcinoma during 20-year follow-up among women with endometrial hyperplasia. J Clin Oncol 2010;28:788–92.
[100] Rhea S, Weber DJ, Poole C, Cairns C. Risk factors for hospitalization after dog bite injury: a case-cohort study of emergency room visits. Acad Emerg Med 2014;21:196–203.