ELSEVIER

Journal of Clinical Epidemiology ■ (2013) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Expected impact of a public health intervention in the presence of synergistic risk factors

Hadrien Charvat[a,b,c,*], François Gueyffier[a,b,c,d], Aurélien Belot[a,b,c,e], Pascal Roy[a,b,c]

[a]*Université Claude Bernard Lyon 1, 43 Boulevard du 11 Novembre 1918, F-69100, Villeurbanne, France*
[b]*CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Equipe Biostatistique-Santé, 16 rue Raphaël Dubois, F-69100, Villeurbanne, France*
[c]*Hospices Civils de Lyon, Service de Biostatistique, 162 avenue Lacassagne, F-69003, Lyon, France*
[d]*INSERM, Centre d'Investigation Clinique 201, 28 Avenue du Doyen Lépine, F-69500, Bron, France*
[e]*Institut de Veille Sanitaire, Département des Maladies Chroniques et des Traumatismes, 12, rue du Val d'Osne, F-94410, Saint-Maurice, France*

Accepted 2 November 2012; Published online xxxx

Abstract

Objective: Elaborate and test a method to extrapolate the population attributable fraction (benefit of an intervention to reduce the exposure of a given population to a given risk factor) to another population allowing for effects of synergistic factors.

Study Design and Setting: Using data from the Systolic Hypertension in the Elderly Program, the present study investigated the impact of a reduction of blood pressure on the occurrence of stroke accounting for the age of the targeted population.

Results: A reduction of blood pressure in populations differing by their age distributions showed that the preventable proportion of strokes increased with age. A 20-mm Hg reduction of blood pressure in a population with mean age 60 years was associated with a 14% reduction of strokes and 18% in a population with mean age 70 years. The difference between these two proportions can be interpreted as the proportion of cases due to the synergistic actions of age and high blood pressure on the occurrence of stroke.

Conclusion: The presented example illustrates how the method may be used by public health practitioners to transpose the potential benefits of interventions estimated in a study population to other populations with different exposures to synergistic risk factors.

Keywords: Hypertension; Population attributable fraction; Risk assessment; Risk factors; Stroke; Synergy

1. Introduction

The study of the relationship between a risk factor and the occurrence of a particular disease is based on several epidemiologic measures. The absolute risk, for example, is a descriptive quantity corresponding to the proportion of cases of the disease in a population exposed to the considered risk factor. The relative risk is based on a comparison of the risk of occurrence of the disease between populations experiencing different patterns of exposure to the risk factor. As such, the relative risk is a particularly valuable epidemiologic tool in evaluating the strength of the association between the exposure to the risk factor and the occurrence of the disease.

However, it is also necessary to have a global measure of the impact of exposure to risk factors in a given population. Actually, even if the relative risk associated with the exposure to a particular risk factor is very high, the effect of this factor may be modest if only a small fraction of the population is exposed. On the contrary, exposure to a risk factor with a relatively moderate relative risk may have a great impact on a population if a great proportion of individuals are exposed to that risk factor. This idea is summarized in the following remark by Rose [1]: ''a large number of people at a small risk may give rise to more cases of disease than a small number who are at a high risk.''

The population attributable fraction (PAF), also known as ''attributable risk'' or ''preventable fraction,'' provides such a measure of the impact of a risk factor at the population level because it combines two sources of information about the risk factor: its effect on the occurrence of the disease and its distribution in the population under study. More specifically, the PAF associated with a risk factor is usually interpreted as the proportion of cases of the disease that could be prevented by suppressing the exposure to this factor in the population

Conflict of interest statement: The authors declare no conflict of interests.

Funding: This work was supported by a grant from the Association pour la Recherche sur le Cancer to H.C.

* Corresponding author. Laboratoire Biostatistique-Santé, Centre Hospitalier Lyon Sud, Pavillon 4D, 165, chemin du Grand Revoyet, F-69310 Pierre-Bénite, France. Tel.: +33-4-78-86-57-75; fax: +33-4-78-86-57-74.
E-mail address: hadrien.charvat@chu-lyon.fr (H. Charvat).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2012.11.004

What is new?

Key finding

- The estimation of the population attributable fraction (PAF) specific of an intervention on a given risk factor should and can be corrected by taking into account the exposure of the target population to one or several other risk factors that may act synergistically.

What this adds to what was known?

- Synergy between risk factors can be simply expressed in terms of the PAFs relative to the single and combined interventions on the risk factors.

What is the implication and what should change now?

- Taking synergistic effects into account allows the extrapolation of the impact of a public health intervention on a specific risk factor to populations differing from the initial study population by one or several characteristics.

under study, supposing that this factor is causally related to the disease. Since its introduction in 1953 by Levin [2], the PAF has been the object of an abundant literature concerning its estimation in various contexts and various applications in epidemiology [3–9].

In particular, the notion of PAF has been extended to a situation in which the exposure to the risk factor is not removed but only modified [6,7]. This adds a third dimension to the PAF that becomes dependent on the choice of exposure modification. More precisely, the estimation of the PAF for a particular exposure modification requires the knowledge of the (1) relationship between exposure and disease occurrence as estimated by the relative risk or odds ratio, (2) exposure to the risk factor observed in the population under study, and (3) new exposure distribution resulting from the chosen modification.

Besides, when two or more risk factors are causally related to the disease under study, the estimation of the PAF associated with any of these risk factors usually depends on the distribution of the other factors. This situation may be described as the existence of *synergy* between the effects of the risk factors. Failing to take these synergistic effect into account may lead to inappropriate conclusions about the way ARs combine: the global PAF cannot be calculated as the sum of the factor-specific PAFs [10]. This simple observation suggests the necessity to correct the PAF estimates to take this synergy between factors into account.

In this article, we use a corrected PAF, also known as the ''sequential attributable fraction'' [11], to extrapolate the impact of an intervention to populations differing from the population under study by their distribution of exposure to a set of synergistic risk factors. This corrected PAF proves particularly useful when one or more synergistic risk factors are nonmodifiable; this is illustrated by a study of the proportion of strokes preventable through reduction of systolic blood pressure (SBP) according to the age distribution of the target population.

2. Materials and methods

2.1. The PAF: concept and formulas

In its original form [2,3], the PAF was defined as the proportion of disease cases preventable by complete elimination of the exposure to a risk factor X in a given population. The PAF may be expressed as follows [3]:

$$\mathrm{PAF}_X = \frac{\Pr(D) - \Pr(D|X=0)}{\Pr(D)}, \tag{1}$$

where Pr (D) is the probability of the disease in that population and Pr $(D|X=0)$ is the probability of the disease in the absence of exposure to X.

In the case of a binary risk factor, another formulation can show the dependence of the PAF on the strength of the relationship between the factor and disease (via the relative risk, RR_x) and on the prevalence p_x of exposure to X [2]:

$$\mathrm{PAF}_X = \frac{\mathrm{p}_X(\mathrm{RR}_X - 1)}{1 + \mathrm{p}_X(\mathrm{RR}_X - 1)}. \tag{2}$$

Note here that PAF is nil in the absence of relationship between the factor and the disease ($\mathrm{RR}_X = 1$) or when the prevalence p_x is nil and that PAF ranges between 0 and 1 for vulnerability factors (whose $\mathrm{RR}_X > 1$). In the case of protective factors, alternative definitions may be used [4] to avoid dealing with negative quantities.

It is clear from formula (2) that the PAF of a risk factor with a high relative risk can be small if the proportion of the population exposed to that factor is very small: suppressing the exposure to that factor in the population would have a very limited benefit in reducing disease cases. On the contrary, that benefit would be much greater if the risk factor was largely spread in the population even with a low relative risk.

However, it is important to note that the PAF is a theoretical quantity that gives an upper bound to the proportion of cases that could be prevented by suppressing the exposure to the risk factor [2].

2.2. PAF in the case of a change in exposure to the risk factor

The above definition of PAF has been already extended to cases in which exposure was not suppressed but only reduced by a specific intervention [6,7]. The PAF was then

defined as the proportion of disease cases preventable through a change in exposure from an observed distribution to a counterfactual one. Let $\Pr^*(D)$ denotes the probability of disease in a population with the counterfactual distribution, then the PAF can be written as:

$$\mathrm{PAF}_A = \frac{\Pr(D) - \Pr^*(D)}{\Pr(D)}. \quad (3)$$

This definition allows much more flexibility than the former one (expressed by formula 1)—suppression being the special case of reduction of exposure to zero. It allows the public health practitioner to estimate the theoretically maximum proportion of cases that would be prevented under different interventions (e.g., a halving of the exposure to a causal risk factor or, more reasonably, a reduction of exposure by, say, 10%). More formally, the extended definition highlights the fact that there are as many PAF values as possible interventions (i.e., counterfactual distributions), which in turn makes clearer the fact that the PAF is not a measure of what is attributable to the factor but what is attributable to a change of its distribution in the population.

2.3. PAF in the case of exposure to several risk factors

Generally, many risk factors are related to a specific disease and one would want to estimate the PAF to a reduction of exposure to one risk factor while taking into account the distribution of exposure to the other risk factors. This is achievable by calculating an adjusted PAF [7–9]. Basically, this corresponds to a weighted sum of the PAFs of reduction of exposure to that risk factor in each combination of exposure levels of the other risk factors.

Usually, the sum of the individual adjusted PAFs is greater than the global PAF associated with the combined modification of exposure distributions; that is, in the case of two risk factors, X and Y, $\mathrm{PAF}_X + \mathrm{PAF}_Y > \mathrm{PAF}_{XY}$. Moreover, the sum of the individual adjusted PAFs may exceed 100%, whereas the PAF associated with the combined intervention on two or more risk factors cannot [10]. The reason of this apparently counterintuitive phenomenon, which is in reality just a consequence of the mathematical definition of the PAF, is the fact that two or more risk factors may act synergistically to promote the disease. Indeed, if a proportion of cases is due to the combined exposure to X and Y, then suppressing the exposure to either factor could theoretically prevent that proportion of cases from occurring. Hence, both PAF_X and PAF_Y include that proportion and summing these two PAFs result in counting that proportion twice, whereas it is counted only once in the PAF associated with the suppression of both exposures. On the contrary, in the case of antagonism, we have $\mathrm{PAF}_X + \mathrm{PAF}_Y < \mathrm{PAF}_{XY}$ because the cases prevented by the exposure to both factors may occur when suppressing the exposure to either of the factors (but not when removing both factors).

2.4. Synergy: concept and quantification

Two binary risk factors are said to have additive effects if $(\mathrm{RR}_{11} - 1) = (\mathrm{RR}_{10} - 1) + (\mathrm{RR}_{01} - 1)$, where subscripts 1 and 0 stand for the presence and absence of exposure to each risk factor. This can also be written as $\mathrm{RR}_{11} = \mathrm{RR}_{10} + \mathrm{RR}_{01} - 1$. Any departure from this situation, that is, $\mathrm{RR}_{11} \neq \mathrm{RR}_{10} + \mathrm{RR}_{01} - 1$, might be described as showing the presence of synergy between the two risk factors [12], but the term "synergy" is usually used only for the case $\mathrm{RR}_{11} > \mathrm{RR}_{10} + \mathrm{RR}_{01} - 1$, whereas the case $\mathrm{RR}_{11} < \mathrm{RR}_{10} + \mathrm{RR}_{01} - 1$ is termed "antagonism." This distinction is of course important in the epidemiologic interpretation of the effects of the risk factors. However, as far as numerical concepts are concerned, there is no need for such distinction between "synergy" and "antagonism": the difference amounts to whether one considers subtraction a concept different from addition or if one accepts to define subtraction as the addition of a negative quantity. In the following, we understand synergy as a generic term encompassing both synergy (positive synergy) and antagonism (negative synergy).

Moreover, synergy describes the resultant of the effects of the risk factors and, thus, should not be confused with the concept of interaction in statistical models which differs according to the class of models used (e.g., additive, multiplicative, etc.) [12–14]. Indeed, suppose that we use a multiplicative model. If the true relationship existing between the two factors is indeed multiplicative, no interaction term is needed in the model to describe their relationship: there is technically no 'statistical interaction' between these two factors. However, there is synergy because $\mathrm{RR}_{11} \neq \mathrm{RR}_{10} + \mathrm{RR}_{01} - 1$.

One way to quantify synergy is to calculate the "coefficient of synergy" [12]—also known in the literature as the "attributable proportion due to interaction" [15]—defined as:

$$\mathrm{CS}_{XY} = \frac{\mathrm{RR}_{11} - (\mathrm{RR}_{10} + \mathrm{RR}_{01} - 1)}{\mathrm{RR}_{11}}. \quad (4)$$

In the case of suppression of exposure to the risk factors, it can be shown that CS_{XY} can be expressed in terms of individual adjusted and global PAFs (see Appendix at www.jclinepi.com). Moreover, it can be empirically checked that the sum of the adjusted PAFs is exactly equal to the PAF associated with the combined intervention when CS_{XY} is nil; that is, when the effects of the two risk factors are additive.

2.5. Taking the synergistic effects of nonmodifiable risk factors into account

The calculation of a PAF for a given intervention uses the probabilities of the disease conditionally on the levels of exposure between the observed and counterfactual distributions of exposure. Thus, an intervention is characterized

by the counterfactual distribution it leads to. The PAF then becomes a measure that quantifies the difference (in terms of proportion of disease) obtained by switching from an observed to a counterfactual distribution of exposure. The question of whether the intervention that leads to the counterfactual distribution of interest is possible, although of paramount importance in the interpretation of PAFs [16], does not fall within the scope of this work.

For example, let us consider a study on the risk factors associated with cardiovascular diseases and a statistical model that estimates the relative risks associated with several risk factors among which are age and blood pressure. Restricting the study to these two factors, one can calculate the population attributable fraction PAF_{SBP} of a pharmaceutical intervention to reduce blood pressure in the study population. But what would be the impact PAF_{SBP}^{Target} of the same intervention on SBP in another population characterized by a similar distribution of blood pressures at a younger age? At first glance, one would say that these two PAFs should be equal.

However, one must take into account the fact that the prevalence of cardiovascular diseases in the younger population is lower than that in the original study population because of the proper effect of age in the occurrence of such diseases and also of the effect of hypertension in elderly people which may differ in younger people.

This can be done by considering a hypothetical intervention that consists in shifting the age distribution from its current state in the study population to a counterfactual one. By this intervention, we obtain a new population, that we call the ''target population,'' sharing the characteristics of the study population except for the distribution of age. As seen previously, this allows us to compute PAF_{Age}, which is the PAF of the shift from the age distribution observed in the study group to the age distribution of the target population and the population attributable fraction $PAF_{SBP,Age}$ relative to the combined modification of age and blood pressure distributions.

Let $\Delta_{SBP,Age}$ be the difference $PAF_{SBP} + PAF_{Age} - PAF_{SBP,Age}$. Then, the corrected PAF that represents the impact of the intervention on hypertension in the target population is $PAF_{SBP}^{Target} = PAF_{SBP} - \Delta_{SBP,Age}$. This corrected PAF associated with one risk factor after the modification of another risk factor is what has been described as the ''sequential attributable fraction'' [11].

Note here that this corrected PAF, as defined above, gives a proportion of the preventable cases that depends on the number of cases observed in the study population. If one wants to compute the proportion of cases theoretically present in the younger population and preventable by a reduction of SBP, it would be necessary to divide PAF_{SBP}^{Target} by $(1 - PAF_{Age})$ to account for the smaller prevalence of the disease in the younger population. However, these estimates would be misleading because they would artificially increase the corrected PAF estimates (because the denominator is less than 1) with regard to PAF_{SBP} of the original study population.

In fact, from a public health point of view, the quantity of interest is the absolute difference in disease prevalence; that is, the numerator of the PAF. Multiplied by the size of the population, it gives an estimate of the number of cases that would be prevented by the intervention. In the present situation, multiplying PAF_{SBP}^{Target} by the number of cases observed in the study population would directly give an estimate of the number of cardiovascular disease cases that would be prevented in the target population. The use of this corrected PAF is illustrated in the following application in which PAFs associated with various interventions on SBP are estimated in populations with different age distributions.

2.6. Data source and statistical analysis

The initial data set included all 2,371 patients of the control group of the Systolic Hypertension in the Elderly Program [17,18], a study that assessed the effect of antihypertensive treatments on stroke and other cardiovascular diseases in subjects aged 60 years and older with isolated systolic hypertension (SBP, >160 mm Hg and diastolic blood pressure, <90 mm Hg) [19]. Considering only the control group allowed us to study the role of a high SBP independently from its modification by some therapeutic intervention. Patients' characteristics are shown in Table 1.

To illustrate the concepts of synergy and corrected PAF, we studied the impact of a reduction of SBP according to the age distribution of the target population. The calculation of PAFs associated with a modification of the SBP or age distributions depended, as mentioned earlier, on three elements.

First, the effect of SBP and age on the occurrence of stroke events during a 5-year period was assessed by the odds ratios derived from a multivariate logistic regression analysis. Potential risk factors were selected by univariate logistic regression and then combined in a multivariate regression model. The factors considered were age, sex, SBP, and histories of diabetes, smoking, stroke, or myocardial infarction. To avoid the regression dilution bias usually observed in such studies [20], it was the SBP taken during the first visit (1 month after inclusion in the study) that was considered rather than the baseline measurement.

Second, calculating the PAFs required the knowledge of the distribution of exposure to SBP and age for a fixed

Table 1. Characteristics of the study population

Characteristics	n = 2,371	Missing values
Stroke events	159	0
Mean age (interquartile range), yr	71.5 (66–76)	0
Sex (ratio M/F)	0.74	0
Stroke history	40	17
Myocardial infarction history	95	29
Smoking (current smokers)	305	0
Diabetes history	241	49
Mean systolic blood pressure (interquartile range), mm Hg	159.3 (148–169)	46

Abbreviations: M, male; F, female.

combination of the other covariates of the model. Instead of using the observed distribution of SBP and age—which may be too specific and hinder generalization of our results—it was decided to model that distribution. Moreover, the distribution modeling was restricted to the largest subgroup that corresponded to patients with no histories of diabetes or stroke. Finally, the distribution of age and SBP in this subgroup was modeled using a truncated normal—normal distribution, the parameters of which were estimated using a maximum likelihood approach.

Third, entering the PAF calculations was the choice of the exposure distribution modifications or equivalently the choice of the counterfactual distributions. Three interventions on SBP were studied; they corresponded to increasingly important reductions in blood pressure levels: the corresponding counterfactual SBP distributions were considered marginally normal, with means 150, 140, and 130 mm Hg, respectively, and the same variance as that of the study population. Besides, various counterfactual age distributions were considered: 60-year truncated normal distributions with means 65 or 60 years and with either the same standard deviation as that of the study population or a standard deviation of 5 years (i.e., a population with fewer very elderly patients). For each counterfactual distribution of age and SBP, PAF_{Age} and PAF_{SBP} were estimated as well as $PAF_{Age,SBP}$, the PAF associated with the combined intervention. Then, for each intervention on SBP, it was possible to compute the PAFs corrected for the considered age distributions. Confidence intervals of PAFs were obtained through Monte Carlo simulations. All analyses were performed with R statistical software version 2.9.1 (R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, available at http://www.r-project.org).

3. Results

The effects of the studied risk factors as given by the univariate logistic regression analysis and their effects as estimated by the final multivariate model are given in Table 2. In the subgroup of patients with no histories of diabetes or stroke (1,997 patients), the bivariate distribution of age and SBP was modeled as a truncated normal—normal probability density function whose parameters were a mean of 69.8 years (rounded to 70 years in the tables) and a standard deviation of 8.1 years (rounded to 8 years in the tables) for age, a mean of 158.7 mm Hg (rounded to 160 mm Hg in the tables) and a standard deviation of 16.5 mm Hg for SBP, and a correlation coefficient of 0.07.

Table 3 presents the PAFs obtained with various counterfactual bivariate distributions of age and SBP. The first line and first column correspond to the classical PAF estimates for SBP and age, respectively. A drop of the mean age of the study population from its observed value to 65 years would reduce the proportion of strokes by 14.2%, whereas a reduction of the mean SBP from its observed value to 140 mm Hg would lead to a reduction of 18%. A combination of these two interventions would lead to a 29.8% reduction only, that is, less than the sum of the PAFs relative to each intervention.

Table 4 presents the PAFs associated with different changes in the mean of the distribution of SBPs for the considered age distributions after removal of the synergistic effect. When examining columns, we can see that, whatever the age distribution, the PAF increases with the extent of SBP reduction. For example, if the goal of the intervention on SBP is to achieve a mean SBP of 150 mm Hg in a population with an age distribution centered around 60 years, this will result in a 6.9% reduction of stroke events. On the other extreme, if the treatment aims to achieve a mean SBP of 130 mm Hg in the same population, the corresponding reduction of stroke events will be of 20.5%.

Examining the rows shows that the PAFs associated with a given reduction of SBP decrease with the mean age of the population. Actually, if an 18% reduction of stroke events could be achieved through an intervention that drops the mean SBP to 140 mm Hg in a population whose age distribution is centered around 70 years, the same intervention would prevent only 14% of stroke events in another population whose age distribution is centered around 60 years. PAFs are also estimated in target populations characterized by a narrower standard deviation of the age distribution than those observed in the study population: it results once more in a reduction of the impact of the intervention on SBP; this can be seen through the differences in PAFs between corresponding columns in Table 4.

Table 2. Results of the univariate and multivariate logistic regression analyses of potential risk factors for stroke

Variables	Odds ratios (95% CI)	
	Univariate logistic regression	Multivariate logistic regression
Age (1 yr)	1.055*** (1.031, 1.080)	1.054*** (1.029, 1.080)
Sex (females, 0; males, 1)	1.12 (0.81, 1.55)	—
Stroke history (no, 0; yes, 1)	3.02** (1.31, 6.93)	3.05* (1.30, 7.15)
Myocardial infarction history (no, 0; yes, 1)	0.45 (0.14, 1.43)	—
Smoking (no, 0; yes, 1)	1.42 (0.92, 2.19)	—
Diabetes history (no, 0; yes, 1)	1.81** (1.16, 2.83)	1.94** (1.24, 3.05)
Systolic blood pressure (1 mm Hg)	1.014** (1.004, 1.024)	1.011* (1.002, 1.021)

Abbreviation: CI, confidence interval.
Levels of significance: * <0.05; ** $<10^{-2}$; *** $<10^{-4}$.

Table 3. Estimated population attributable fractions (percentage and 95% confidence intervals) on the occurrence of stroke for different counterfactual exposure distributions to both factors

	Mean age = 70 yr		Mean age = 65 yr		Mean age = 60 yr	
Mean SBP (mm Hg)	**SD = 8 yr**	**SD = 5 yr**	**SD = 8 yr**	**SD = 5 yr**	**SD = 8 yr**	**SD = 5 yr**
160	0	11.85 (5.88, 18.52)	14.15 (7.47, 20.52)	25.30 (13.75, 36.29)	23.65 (12.51, 33.71)	33.92 (19.06, 47.34)
150	8.80 (1.62, 15.63)	19.55 (10.96, 27.95)	21.78 (13.05, 29.61)	32.01 (20.33, 42.72)	30.51 (19.35, 40.54)	39.90 (25.61, 52.59)
140	18.04 (3.27, 31.00)	27.91 (13.60, 40.10)	29.76 (16.37, 41.48)	39.03 (24.00, 51.16)	37.63 (23.28, 49.41)	46.12 (30.41, 58.96)
130	26.35 (4.35, 43.18)	35.32 (15.88, 50.76)	36.97 (17.46, 51.70)	45.33 (26.45, 59.30)	44.13 (25.54, 58.19)	51.75 (33.30, 65.30)

Abbreviations: SBP, systolic blood pressure; SD, standard deviation.

Table 4. Corrected population attributable fractions (percentage and 95% confidence intervals) for different interventions on systolic blood pressure (SBP) corrected for the age distribution of the targeted population

	Mean age = 70 yr		Mean age = 65 yr		Mean age = 60 yr	
Mean SBP (mm Hg)	**SD = 8 yr**	**SD = 5 yr**	**SD = 8 yr**	**SD = 5 yr**	**SD = 8 yr**	**SD = 5 yr**
150	8.80 (1.62, 15.63)	7.69 (1.32, 14.02)	7.63 (1.07, 13.76)	6.70 (1.07, 12.29)	6.86 (1.20, 12.64)	5.97 (0.77, 11.18)
140	18.04 (3.27, 31.00)	16.06 (2.49, 27.49)	15.62 (2.48, 27.19)	13.72 (1.82, 24.21)	13.98 (2.37, 24.91)	12.20 (1.93, 22.42)
130	26.35 (4.35, 43.18)	23.47 (4.12, 38.36)	22.82 (3.55, 38.14)	20.02 (3.26, 34.23)	20.48 (3.27, 34.79)	17.83 (2.84, 31.20)

Abbreviation: SD, standard deviation.

4. Discussion

The prevention of a disease may rely on public health interventions aimed at reducing the exposure to one or several risk factors. The impact of these interventions in a given population can be assessed by PAF estimates. However, when considering a transposition of the results of a given intervention to another population, it is necessary to take into account the differences between populations in terms of prevalence of exposure and disease incidence [21] and potential synergies between risk factors. This is based on the observation that, in some cases, a disease may be caused by the combined actions of several risk factors so that the prevention of any of them would be sufficient to prevent the disease.

One way to address this issue is to split the PAF by considering sequential interventions on the risk factors [11,22−25]. For example, Eide and Gefeller [11] proposed to remove sequentially the exposures to each risk factor: the sequential attributable fraction of the *i*th risk factor was defined as the difference between the PAF obtained by removing all first i factors and that obtained by eliminating only the first $i-1$ factors. It is then possible to estimate an average attributable fraction for each factor, which is the mean PAF obtained over all possible removal orderings. More recently, Mason and Tu [25] proposed a new partitioning strategy that takes into account the theoretical relationships between the factors; that is, the fact that they interact or are correlated. However, all these methods are based on a conception of PAFs as a means of apportioning the total risk of the disease between a set of causes and thus seek additivity of PAFs. This causal interpretation of PAFs should in our opinion be considered with much caution and we prefer the ''preventable fraction'' interpretation, although it also depends on a set of hypotheses that are seldom considered, let alone verified [16].

The PAF being primarily a measure of reduction of disease cases by reducing the exposure to the risk factors, some authors raised the question of the pertinence of computing PAFs for nonmodifiable risk factors [26]. Although devoid of practical signification, these PAFs can, however, be used to extrapolate the estimated impact of an intervention on a modifiable risk factor in a given population to another population (referred to as the target population) differing by the distribution of the exposure level to one or several nonmodifiable synergistic risk factors.

The present approach provides an interpretation of sequential PAFs in terms of PAFs corrected for synergistic effects. Moreover, a different interpretation of the PAF is proposed: it can be seen as the differential proportion of cases observed between populations differing by their exposures to several risk factors. With this concept, it is legitimate to consider PAFs for nonmodifiable risk factors.

Dealing with continuous risk factor distributions may be considered as a useless complication. However, Altman and Royston [27] drew the attention to the fact that categorization of continuous variables, although helpful to label individuals, was unnecessary for statistical analysis and the source of several problems, mainly a potentially important loss of information.

For simplicity, the interventions considered in the present illustration were limited to reductions of the means of the distribution of SBP, a choice based on the assumption that the treatment has the same effect in all subjects whatever their initial SBP level. However, the proposed method is not limited to such situations and can be straightforwardly transposed to other modifications of the exposure distribution [28]. For example, it would be interesting to study an intervention on SBP that would depend on the initial level of hypertension of the patients.

Besides, the present results emphasize the link between the PAF estimates and the particular intervention these estimates rely on. Actually, if the reverse—and commonly used—definition of the PAF as the proportion of cases attributable to the exposure to a risk factor seems to make sense when considering a suppression of that exposure, it is more difficult to give it a practical meaning within the context of exposure modification because there are as many PAF estimates as possible modifications of the exposure distribution. Some authors have dealt with that issue in the case of SBP reduction by using an 'ideal' counterfactual distribution [29], but the definition of that distribution is still controversial [30].

Note finally that the interpretation of PAF estimates relies on the hypothesis of a causal relationship [31,32] between a risk factor and the event of interest. However, as pointed out by Walter [4], causality is not sufficient by itself to allow a direct interpretation of PAF estimates. The example of blood pressure treatment may illustrate this point: first, high SBP is deemed related to stroke through mechanical action on blood vessels, but it may also reflect other pathological processes that SBP reduction cannot entirely reverse; besides, antihypertensive drugs may act by other ways than reducing SBP [33].

To conclude, the proposed method may be used by public health practitioners to transpose the potential benefits of interventions as estimated on a particular population to populations with different exposure distributions of synergistic risk factors.

Acknowledgments

The authors thank Jacques Estève for his valuable comments on this study and Jean Iwaz for his thorough editing of the article.

Appendix

Supplementary material

Supplementary data related to this article can be found online at doi:10.1016/j.jclinepi.2012.11.004.

References

[1] Rose G. Sick individuals and sick populations. Int J Epidemiol 1985;14:32−8.

[2] Levin ML. The occurrence of lung cancer in man. Acta Unio Int Contra Cancrum 1953;9:531−41.

[3] Miettinen OS. Proportion of disease caused or prevented by a given exposure trait or intervention. Am J Epidemiol 1974;99:325−32.

[4] Walter SD. The estimation and interpretation of attributable risk in health research. Biometrics 1976;32:829−49.

[5] Walter SD. Prevention for multifactorial diseases. Am J Epidemiol 1980;112:409−16.

[6] Drescher K, Becher H. Estimating the generalized impact fraction from case-control data. Biometrics 1997;53:1170−6.

[7] Eide GE, Heuch I. Attributable fractions: fundamental concepts and their visualization. Stat Methods Med Res 2001;10:159−93.

[8] Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol 1985;122:904−14.

[9] Benichou J. A review of adjusted estimators of attributable risk. Stat Methods Med Res 2001;10:195−216.

[10] Rowe AK, Powell KE, Flanders WD. Why population attributable fractions can sum to more than one. Am J Prev Med 2004;26: 243−9.

[11] Eide GE, Gefeller O. Sequential and average attributable fractions as aids in the selection of preventive strategies. J Clin Epidemiol 1995;48:645−55.

[12] Roy P, Estève J. Using relative risk models for estimating synergy between two risk factors. Stat Med 1998;17:1357−73.

[13] Walker AM. Proportion of disease attributable to the combined effect of two factors. Int J Epidemiol 1981;10:81−5.

[14] Knol MJ, van der Tweel I, Grobbee DE, Numans ME, Geerlings MI. Estimating interaction on an additive scale between continuous determinants in a logistic regression model. Int J Epidemiol 2007;36: 1111−8.

[15] Rothman KJ. Epidemiology: an introduction. New York: Oxford University Press; 2002.

[16] Levine B. What does the population attributable fraction mean? Prev Chronic Dis 2007;4:1−5.

[17] Borhani NO, Applegate WB, Cutler JA, et al. Systolic Hypertension in the Elderly Program (SHEP). Part 1: rationale and design. Hypertension 1991;17:II2−15.

[18] Gueyffier F, Boutitie F, Boissel JP, et al. INDANA: a meta-analysis on individual patient data in hypertension. Protocol and preliminary results. Therapie 1995;50:353−62.

[19] Garland C, Barrett-Connor E, Suarez L, Criqui MH. Isolated systolic hypertension and mortality after age 60 years. A prospective population-based study. Am J Epidemiol 1983;118:365−76.

[20] MacMahon S, Peto R, Cutler J, et al. Blood pressure, stroke, and coronary heart disease. Part 1. Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. Lancet 1990;335:765−74.

[21] Ellenberg JH. The extrapolation of attributable risk to new populations. Stat Med 1988;7:717−25.

[22] Land M, Vogel C, Gefeller O. Partitioning methods for multifactorial risk attribution. Stat Methods Med Res 2001;10:217−30.

[23] McElduff P, Attia J, Ewald B, Cockburn J, Heller R. Estimating the contribution of individual risk factors to disease in a person with more than one risk factor. J Clin Epidemiol 2002;55:588−92.

[24] Rabe C, Lehnert-Batar A, Gefeller O. Generalized approaches to partitioning the attributable risk of interacting risk factors can remedy existing pitfalls. J Clin Epidemiol 2007;60:461−8.

[25] Mason CA, Tu S. Partitioning the population attributable fraction for a sequential chain of effects. Epidemiol Perspect Innov 2008;5:5.

[26] Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. Am J Public Health 1998;88:15−9.

[27] Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ 2006;332:1080.

[28] Murray CJL, Lopez AD. On the comparable quantification of health risks: lessons from the Global Burden of Disease Study. Epidemiology 1999;10:594−605.

[29] Lawes CMM, Vander Hoorn S, Law MR, Elliott P, MacMahon S, Rodgers A. Blood pressure and the global burden of disease 2000. Part II: estimates of attributable burden. J Hypertens 2006;24:423−30.

[30] Gueyffier F, Wright J. The lower the better: does simplicity lead to absurdity? [Editorial Comment]. J Hypertens 2006;24:431−3.

[31] Parascandola M, Weed DL. Causation in epidemiology. J Epidemiol Community Health 2001;55:905−12.

[32] Maldonado G, Greenland S. Estimating causal effects. Int J Epidemiol 2002;31:422−9.

[33] Boissel JP, Gueyffier F, Boutitie F, Pocock S, Fagard R. Apparent effect on blood pressure is only partly responsible for the risk reduction due to antihypertensive treatments. Fundam Clin Pharmacol 2005;19:579−84.