

METHODOLOGY

Open Access

Population attributable fraction: comparison of two mathematical procedures to estimate the annual attributable number of deaths

Bernard CK Choi^{1,2,3,4}

Abstract

Objective: The purpose of this paper was to compare two mathematical procedures to estimate the annual attributable number of deaths (the Allison et al procedure and the Mokdad et al procedure), and derive a new procedure that combines the best aspects of both procedures. The new procedure calculates attributable number of deaths along a continuum (i.e. for each unit of exposure), and allows for one or more neutral (neither exposed nor nonexposed) exposure categories.

Methods: Mathematical derivations and real datasets were used to demonstrate the theoretical relationship and practical differences between the two procedures. Results of the comparison were used to develop a new procedure that combines the best features of both.

Findings: The Allison procedure is complex because it directly estimates the number of attributable deaths. This necessitates calculation of probabilities of death. The Mokdad procedure is simpler because it estimates the number of attributable deaths indirectly through population attributable fractions. The probabilities of death cancel out in the numerator and denominator of the fractions. However, the Mokdad procedure is not applicable when a neutral exposure category exists.

Conclusion: By combining the innovation of the Allison procedure (allowing for a neutral category) and the simplicity of the Mokdad procedure (using population attributable fractions), this paper proposes a new procedure to calculate attributable numbers of death.

Background

There are two mathematical procedures to estimate the number of deaths attributable to a risk factor such as obesity, smoking or alcohol consumption. Number of attributable deaths is the number of deaths in a population that could be avoided if the effects of the risk factor were eliminated from the population. The two procedures are Allison et al [1] and Mokdad et al [2]. Both procedures are under the assumptions of no confounding and no effect modification. Both can be applied to risk factors with polytomous exposure categories.

The Allison procedure [1], originally developed for obesity attributable deaths, is rather complex. It involves 12 steps, uses hazard ratios, and requires calculating

hazard rates by using a mathematical process to solve for an unknown quantity. The Mokdad procedure [2] on the other hand, is simpler. It involves only 6 steps, uses relative risks, and does not require solving for any unknown quantity.

In general, a common belief is that the more complex the procedure, the more accurate the results. Allison et al further stipulated that their procedure accounts for “complications”, because it can estimate attributable deaths for body mass index (BMI) along a continuum (ie, for each unit of BMI), and can adjust for time using hazard ratio (HR) that the relative risks (RR) cannot achieve [1]. As a result, Mokdad et al used the Mokdad procedure to estimate the attributable numbers for tobacco and alcohol, and then they reverted to the more complex Allison procedure to estimate the attributable number for obesity.

Correspondence: Bernard_Choi@phac-aspc.gc.ca

¹Centre for Chronic Disease Prevention and Control, Public Health Agency of Canada (PHAC), Government of Canada, Ottawa, Ontario, Canada
Full list of author information is available at the end of the article

A detailed read of Allison et al's paper revealed that two steps in the Allison procedure are not well-documented. First, while the equation for the overall number of deaths attributable to obesity and overweight (ω) is given in their paper, the equation to calculate the number of deaths attributable to each individual BMI category is missing. It is therefore unclear how data in their table three can lead to results in their table four. Second, it is said in their paper that the hazard (λ) can be obtained by numerically solving a complex equation for λ . The actual method, however, is not given. For the less sophisticated users, the Allison procedure is not user-friendly.

There are several questions arising from looking at these two procedures: How are the Allison and Mokdad procedures mathematically and practically different? What are the best aspects of each procedure? Can the underlying equations be combined or modified to take advantage of the best aspects of both?

This paper compares the Allison and Mokdad procedures for the estimation of annual attributable number of deaths, both mathematically and using real data. The paper also "recovers" the missing Allison equation to calculate the individual number of deaths attributable to each BMI category, develops a similar and simpler equation using the logic of the Mokdad procedure, compares estimated number of attributable deaths under the HR and RR models, and looks at several options for numerically solving the equation for λ . This paper also proposes a modified Mokdad procedure that can achieve the same results as the Allison procedure.

Methods

Mathematical derivations from first principles from population attributable fraction (PAF), defined as the proportion of deaths in a population that can be attributed to the causal effects of a risk factor or set of factors, were used to demonstrate the relationship and differences of the Allison and Mokdad procedures. The missing mathematical equation to estimate the number of deaths attributable to each exposure category was derived for the Allison procedure. A similar equation was created for the Mokdad procedure. The two procedures were then "taken apart" and the logics behind the two procedures were examined and compared. Based on this, a new procedure (modified Mokdad) was developed combining the innovation of the Allison procedure and the logic of the Mokdad procedure. Finally, estimation methods under the hazard ratio and relative risk models were compared. Some options for solving for λ were described. Real datasets provided by Allison et al were used to illustrate the practical differences of the two procedures (Allison and the new procedure provided in this paper).

Results

Table 1 is a conversion table of the notations used in Allison et al, Mokdad et al, and this paper. The 12 steps in the Allison Procedure and the 6 steps in the Mokdad procedure to calculate the attributable number of deaths are summarized in Additional file 1, Appendix S1, using their original notations.

1. Mathematical proof that the Allison procedure and Mokdad procedure differ in a neutral exposure category (Q)

Based on Levin [3], and using notations in Table 2,

$$PAF = \frac{p - p_0}{p} \tag{1}$$

where PAF is population attributable fraction; p is the probability of death in the population, or $P(D)$; p_0 is the probability of death among the nonexposed, or $P(D|E_0)$. Then using equation T1 (Table 2),

$$\begin{aligned} PAF &= \frac{p - p_0}{p} = \frac{(1-f)p_0 + fR p_0 - p_0}{(1-f)p_0 + fR p_0} \\ &= \frac{p_0 - f p_0 + f R p_0 - p_0}{p_0 - f p_0 + f R p_0} = \frac{f R - f}{1 + f R - f} \\ PAF &= \frac{f(R-1)}{1+f(R-1)} \end{aligned} \tag{2}$$

Equation 2 is another frequently quoted form of PAF [3].

Extending the same methodology above for the case of a dichotomous exposure variable to the case of a polytomous exposure variable, and using equation T2 (Table 3),

$$PAF = \frac{p - p_0}{p} = \frac{(1 - \sum f_i)p_0 + \sum f_i R_i p_0 - p_0}{(1 - \sum f_i)p_0 + \sum f_i R_i p_0} \tag{3}$$

$$PAF = \frac{(1 - \sum f_i) + \sum f_i R_i - 1}{(1 - \sum f_i) + \sum f_i R_i} \tag{4}$$

Equation 4 is Mokdad et al [2] (see Additional file 1, Appendix S1, equation A3), as $1 - \sum f_i$ is P_0 (Table 1).

From equation 3,

$$\begin{aligned} PAF &= \frac{(1 - \sum f_i)p_0 + \sum f_i R_i p_0 - [(1 - \sum f_i)p_0 + \sum f_i p_0]}{(1 - \sum f_i)p_0 + \sum f_i R_i p_0} \\ &= 1 - \frac{(1 - \sum f_i)p_0 + \sum f_i p_0}{(1 - \sum f_i)p_0 + \sum f_i R_i p_0} \end{aligned}$$

But because $p = (1 - \sum f_i)p_0 + \sum f_i R_i p_0$, see equation T2 in Table 3, therefore

$$PAF = 1 - \frac{1}{p} [(1 - \sum f_i)p_0 + \sum f_i p_0] \tag{5}$$

Equation 5 is a modified form of Allison et al [1] (see Additional file 1, Appendix S1, equation A1), as shown

Table 1 Conversion table of notations

Variable	Allison et al [1]	Mokdad et al [2]	This Paper
No. of deaths attributable	ω	ω	ω
Total no. of deaths in population	M	M	M
Total no. in population	N	N	N
Fraction of population nonexposed	P(R)	P_0	$f_0, 1 - \sum f_i, 1 - \sum f_i - f_q$
Fraction of population exposed	P(O)	$\sum P_i$	$f, \sum f_i$
Fraction of population exposed to an exposure category (i)		P_i	f_i
Fraction of population exposed to a neutral category (e.g., underweight)	P(Q)		f_q
Hazard ratio (HR) for an exposure category	h		h
Hazard ratio (HR) for a neutral category (e.g., underweight)	q		q
Probability of death in a year in population	P(D), M/N		$p, P(D)$
Hazard of death in the nonexposed	λ		λ
Conditional probability of death in a year in nonexposed	$P(D R), 1 - e^{-\lambda}$		$p_0, P(D E_0)$
Conditional probability of death in a year in various exposure categories	$P(D O), 1 - e^{-h\lambda}$		R_i, p_0
Conditional probability of death in a year in a neutral category (e.g., underweight)	$P(D Q), 1 - e^{-q\lambda}$		p_q
Relative risk	RR, P(D O)/P(D R)	RR _i	R_i
Population attributable fraction		PAF	PAF

below. From equations A2 (Additional file 1, Appendix S1) and 5, and given $p = M/N$,

$$\begin{aligned} \omega &= M \cdot \text{PAF} \\ &= M \left\{ 1 - \frac{1}{p} \left[(1 - \sum f_i) p_0 + \sum f_i p_0 \right] \right\} \\ &= M - N[(1 - \sum f_i) p_0 + \sum f_i p_0] \\ &= M - N(f_0 p_0 + \sum f_i p_0) \\ &= M - N[P(R)P(D|R) + P(O)P(D|R)] \end{aligned} \tag{6}$$

Comparing the Mokdad equation (6) with the Allison equation (A1), the Mokdad procedure allows only for a single nonexposed category E_0 and a number of exposure categories $E_1 \dots E_i \dots E_k$ (see Table 3), while the Allison procedure in addition allows for a neutral (i.e., neither nonexposed nor exposed) category, in this case the category Q (underweight) (see Table 4).

The methodology of Allison et al leads to a modified Levin equation as shown below.

The original Allison equation (Additional file 1, Appendix S1, equation A1) was written for a three

category exposure (R, reference group or the nonexposed E_0 ; O, overweight and obese groups; and Q, the underweight group) (Table 4) and is

$$\omega = M - N[P(R)P(D | R) + P(O)P(D | R) + P(Q)P(D | Q)]$$

where ω is number of deaths attributable to exposure categories $E_1 \dots E_i \dots E_k$ combined, M is total number of deaths, N is total number in population, P(R) is $f_0 = 1 - \sum f_i - f_q$, P(O) is $\sum f_i$, P(D|R) is p_0 .

From equation A1, and using the notations of this paper (Table 4),

$$\begin{aligned} \omega &= M - N(f_0 p_0 + \sum f_i p_0 + f_q p_q) \\ &= M - N[(1 - \sum f_i - f_q) p_0 + \sum f_i p_0 + f_q p_q] \\ &= M \left\{ 1 - \frac{1}{p} \left[(1 - \sum f_i - f_q) p_0 + \sum f_i p_0 + f_q p_q \right] \right\} \end{aligned} \tag{7}$$

given $p = M/N$. And since $\text{PAF} = \omega/M$ (equation A2),

$$\begin{aligned} \text{PAF} &= 1 - \frac{1}{p} \left[(1 - \sum f_i - f_q) p_0 + \sum f_i p_0 + f_q p_q \right] \\ &= \frac{p - [(1 - \sum f_i - f_q) p_0 + \sum f_i p_0 + f_q p_q]}{p} \\ &= \frac{p(1 - f_q) p_0 - f_q p_q}{p} \end{aligned} \tag{8}$$

Comparing equation 8 (derived from Allison et al) with Levin's original equation for PAF (equation 1) which is identical to the Mokdad procedure, the Allison procedure subtracts out a certain weighted proportion of deaths associated with the neutral category (the underweight Q) from the attributable deaths to the exposure (the overweight and obese). In other words, the Allison procedure allows for a neutral exposure category (neither nonexposed nor exposed), while the Mokdad procedure does not.

Table 2 Distributions according to death and an exposure variable: Dichotomous exposure

	Nonexposed E_0	Exposed E	Total
Death D	$(1 - f) p_0$	$f R p_0$	p
No death \bar{D}	$(1 - f) (1 - p_0)$	$f (1 - R p_0)$	$1 - p$
Total	$1 - f$	f	1

Let

p = probability of death in population

f = fraction of population exposed

p_0 = probability of death in the nonexposed

R = relative risk

It follows that for a dichotomous exposure variable, based on the second row of the table,

$$p = (1 - f) p_0 + f R p_0 \tag{T1}$$

Table 3 Distributions according to death and an exposure variable: Polytomous exposure

	Nonexposed E ₀	Exposed Categories E ₁ ...E _i ...E _k	Total
Death D	(1 - Σf _i) p ₀	Σ f _i R _i p ₀	p
No death \bar{D}	(1 - Σf _i) (1-p ₀)	Σ f _i (1- R _i p ₀)	1-p
Total	1 - Σf	Σf _i	1

Let

f_i = fraction of population exposed to exposure category i

R_i = relative risk for exposure category i compared to non exposed

It follows that for a polytomous exposure variable

$$p = (1 - \sum f_i) p_0 + \sum f_i R_i p_0 \quad (T2)$$

2. Recovery of the Allison equation for number of deaths attributable to each exposure category

To derive the missing Allison equation, from equation A1,

$$\begin{aligned} \omega &= M-N [P(R)P(D|R)+P(O)P(D|R)+P(Q)P(D|Q)] \\ &= M-N [P(R)P(D|R)+\Sigma P(O_i)P(D|R)+P(Q)P(D|Q)] \\ &= N [P(R)P(D|R)+P(O)P(D|O)+P(Q)P(D|Q)] \\ &\quad - N [P(R)P(D|R)+P(O)P(D|R)+P(Q)P(D|Q)] \\ &= N [P(O)P(D|O)-P(O)P(D|R)] \\ &= N P(O) [P(D|O)-P(D|R)] \\ &= N \Sigma \{P(O_i) [P(D|O_i)-P(D|R)]\} \end{aligned} \quad (9)$$

Therefore, the missing equation is

$$\omega_i = N P(O_i) [P(D|O_i)-P(D|R)] \quad (10)$$

where ω_i is number of deaths attributable to each exposure category i.

Equations 9 and 10 are new equations we created for the Allison procedure. It is worth noting that the underweight (Q) category disappears in equations 9 and 10, as it does not play a role in the calculations of attributable numbers to obesity (O). However, equations 9 and 10 are still difficult to use because the probabilities of death in the exposed and reference groups can only be estimated through a complex process:

$$P(D|O_i) = 1 - e^{-h\lambda}, \quad P(D|R) = 1 - e^{-\lambda}.$$

The next section shows that the Mokdad procedure can be modified to do the same calculations as the Allison procedure, but more simply.

3. Development of an equation for number of deaths attributable to each exposure category for the modified Mokdad procedure

The Allison procedure directly estimates the number of attributable deaths, ω. The Mokdad procedure indirectly estimates ω by first estimating PAF. Using the logic of the Mokdad procedure, we develop a new, modified Mokdad procedure to estimate ω as follows:

From equation 10, and using notations in Table 4,

$$\begin{aligned} \omega_i &= N P(O_i) [P(D|O_i)-P(D|R)] \\ &= N P(O_i) [RR_i P(D|R)-P(D|R)] \\ &= N f_i [R_i p_0 - p_0] = N f_i (R_i - 1) p_0 \end{aligned}$$

$$\begin{aligned} M &= N p = N (f_0 p_0 + \sum f_i R_i p_0 + f_q p_q) \\ &= N (f_0 p_0 + \sum f_i R_i p_0 + f_q R_q p_0) \\ &= N p_0 (f_0 + \sum f_i R_i + f_q R_q) \end{aligned}$$

$$\begin{aligned} PAF_i &= \frac{\omega_i}{M} = \frac{N f_i (R_i - 1) p_0}{N p_0 (f_0 + \sum f_i R_i + f_q R_q)} \\ &= \frac{f_i (R_i - 1)}{f_0 + \sum f_i R_i + f_q R_q} \end{aligned}$$

Table 4 Consideration of a neutral exposure category

	Nonexposed E ₀ (e.g. normal weight, R)	Exposed Categories E ₁ ...E _i ...E _k (e.g. overweight and obese, O)	Neutral Category (e.g. underweight, Q)	Total
Death D	(1 - Σf _i -f _q) p ₀	Σf _i R _i p ₀	f _q p _q	p
No death \bar{D}	(1 - Σf _i - f _q) (1 - p ₀)	Σf _i (1- R _i p ₀)	f _q (1-p _q)	1 - p
Total	1 - Σf _i -f _q	Σf _i	f _q	1

Let

f₀ = fraction of population nonexposed = 1 -Σf_i - f_q

f_q = fraction of population underweight

It follows that when a neutral category is involved

$$\begin{aligned} p &= f_0 p_0 + \sum f_i R_i p_0 + f_q p_q \\ &= (1 - \sum f_i - f_q) p_0 + \sum f_i R_i p_0 + f_q p_q \end{aligned} \quad (T3)$$

Using Allison's notations, the above equation can also be expressed as

$$PAF_i = \frac{P(O_i)[RR_i - 1]}{P(R) + \sum P(O_i)RR_i + P(Q)RR_q} \quad (11)$$

It then follows that

$$\omega_i = M \cdot PAF_i = \frac{MP(O_i)[RR_i - 1]}{P(R) + \sum P(O_i)RR_i + P(Q)RR_q} \quad (12)$$

Equations 11 and 12 are new equations we created for exposure category-specific PAF and attributable number, respectively. Because the Mokdad procedure uses the PAF approach, the neutral category Q reappears in equations 11 and 12, because Q is part of the total population. P(Q) is easy to obtain from health surveys. Equation 12 is expected to yield identical results as those of equation 10 (Allison's), because it is derived from equation 10. Equation 12 is simpler to use as it needs only total deaths (M), fractions of exposure in each category ((P(O_i), P(R), P(Q)) which are readily available from health surveys and the relative risks (RR_i, RR_q). Additionally, equation 10 is difficult to use because it also requires total population (N), probability of death in each of the exposure categories (P(D|O_i)), and probability of death in the reference group (P(D|R)). The probabilities of death are difficult to obtain.

4. Difference in the estimated number of attributable deaths under the hazard ratio and the relative risk models

Relative risk (RR) is an estimate of hazard ratio (HR). HR is the ratio of hazard rates (instantaneous incidence rates) in the exposed to the nonexposed at a point in time [4]. RR is the ratio of average risks of disease or death in the exposed to the nonexposed over a period of time.

$$RR = \frac{1 - e^{-h\lambda}}{1 - e^{-\lambda}} \quad (13)$$

where h is hazard ratio (HR) and λ is hazard rate in nonexposed.

In theory, as pointed out by Allison et al, RR estimates "without adjustment for time can bias results (though the bias may be small)" [1]. The question is, in practice, does it matter whether RR or HR is used? Table 5 is a theoretical comparison of HR and RR using equation 13, based on hazard rates of 0.01 and 0.10, and HR of 1, 3, 5, 7. When hazard rate is low (e.g., 0.10 or below), HR and RR are close to each other. The lower the hazard rate (e.g., 0.01), the closer together the HR and RR. From the real data for Alameda County Health Study provided by Allison et al, when HR was 1.39, the RR was 1.38766116; when HR was 0.98, the RR was

0.98008466 (Table 5). There is no practical difference in the real setting in the estimated number of attributable deaths under the HR and RR models.

5. Options for numerically solving an equation for the hazard of death in the nonexposed (λ)

There are commercial packages available for solving an equation for an unknown quantity; packages such as MATHEMATICA and MAPLE [5]. However, for these packages there is a steep learning curve for beginners, and packages can be quite expensive [6]. We looked into two simpler non-commercial options which one can easily program at no cost.

The first option is Newton's method (Additional file 1, Appendix S2). Applying Newton's method to the Alameda County Health Study data, provided in Allison et al's table three, gave an estimated λ of 0.008651. The second option is Taylor series (Additional file 1, Appendix S3). Applying Taylor series to the same data gave the same estimated λ of 0.008651. The two options gave virtually the same answers, with an error margin of less than 0.000001.

6. Comparison of the Allison procedure and modified Mokdad procedure with real datasets

We used the real dataset from the Alameda County Health Study provided by Allison et al [1] to compare the results using the Allison procedure and the modified Mokdad procedure, under both the hazard ratio (HR) and the relative risk (RR) models (Additional file 1, Appendix S4).

From our Additional file 1, Appendix S4, it can be seen that the results using the Allison procedure and the modified Mokdad procedure, under the hazard ratio (HR) and the relative risk (RR) models, are very similar to each other. The Allison procedure is a HR approach and the Mokdad procedure is a RR approach. Therefore the results of the Mokdad procedure using RR are closer to the Allison procedure than the Mokdad procedure using HR. However, the Mokdad procedure using HR to approximate RR provides good enough estimates of attributable number of deaths, and it avoids the use of equation 13 which involves estimation of RR that involves complex estimation of λ.

Discussion

The procedures recommended by Allison et al [1] and Mokdad et al [2] can both be applied to estimate the number, as well as fraction, of a single outcome (such as death) attributable to a risk factor (such as increased body mass index, BMI) that is polytomous (e.g., overweight, obese, and even stratified by BMI unit). Although not specifically mentioned in the two original articles [1,2], both procedures can be applied to one or

Table 5 Comparison of hazard ratio (HR) and relative risk (RR) using the notations of Allison et al [1]

Hazard rate in nonexposed	Hazard ratio	Probability of death in nonexposed	Probability of death in exposed	Relative risk
λ	HR = h	$P(D R) = 1 - e^{-\lambda}$	$P(D O) = 1 - e^{-h\lambda}$	$RR = \frac{1 - e^{-h\lambda}}{1 - e^{-\lambda}}$
Theoretical comparison				
0.01	1	0.00995	0.00995	1.00
	3		0.02955	2.97
	5		0.04877	4.90
	7		0.06761	6.79
0.10	1	0.09516	0.09516	1.00
	3		0.25918	2.72
	5		0.39346	4.13
	7		0.50341	5.29
Real example from table 3 of Allison et al [1]				
0.008651	1.39	0.00861368777	0.0119528799	1.38766116
0.008651	0.98	0.00861368777	0.0084421433	0.98008466

more risk factor combinations (such as BMI and smoking) as long as the risk factor combinations are expressed in independent (i.e., nonoverlapping) exposure categories. Furthermore, both procedures are under the assumptions of no confounding and no effect modification by the risk factors of interest and other covariates (such as age or sex).

The Allison procedure can be applied to the situation when there is a nonexposed category, one or more exposure categories, and one or more neutral (neither nonexposed nor exposed) categories. Allowance of a neutral exposure category is a benefit of the Allison procedure from a causal inference perspective, because in reality the population cannot always be dichotomized into nonexposed and exposed. The Mokdad procedure cannot allow for a neutral category. This paper proposes a modified Mokdad procedure that can achieve the same results as the Allison procedure, but through a simpler way.

The Allison procedure involves twelve steps, while the Mokdad procedure involves only six steps (Additional file 1, Appendix S1). The reason why the Allison procedure involves more steps is because it attempts to directly estimate the attributable number of deaths (equation A1), and this necessitates the estimation of the probabilities of death in the nonexposed, various exposure and the neutral categories. This in turn necessitates the calculation of the hazard rate in the nonexposed, λ , which requires substantial mathematical skills. The Mokdad procedure, on the other hand, first calculates the population attributable fraction (equation A3), and then obtain the attributable

number of deaths by multiplying the PAF with the total number of deaths (equation A2). Our paper (equations 1-4) shows that in the derivation of the equation for PAF in the Mokdad procedure, the probabilities of death cancel out each other in the numerator and denominator, leaving only fractions of exposure and relative risks as necessary input parameters for the estimation of PAF. This greatly simplifies the calculation process in the Mokdad procedure.

The Mokdad procedure, however, breaks down if a neutral category (such as underweight) that is neither nonexposed (such as normal weight) nor exposed (such as overweight and obese) exists. Also, while the Mokdad procedure can calculate the overall number of deaths attributable to a risk factor with multiple exposure categories, it does not calculate the number attributable to each individual exposure category. The Allison procedure, on the other hand, can estimate the individual exposure category attributable numbers (although the exact equation was not given in Allison et al [1].)

By combining the innovation of the Allison procedure [1] (i.e., allowing for a neutral category which is neither nonexposed nor exposed) and the simplicity of the Mokdad procedure [2] (i.e., calculating attributable numbers indirectly through population attributable fractions), this paper proposes a new procedure (modified Mokdad) to calculate population attributable fractions and numbers (Table 6). This paper also “recovers” the missing equation in Allison et al that calculates the individual exposure category attributable numbers (equation 10), and develops a similar equation using the Mokdad et al approach (equation 12). Furthermore, this paper

Table 6 Eight steps in our proposed new procedure (modified Mokdad) to calculate number of deaths attributable to a risk factor with multiple exposure categories, allowing for one or more neutral categories

Let ω be the number of deaths attributable to a risk factor (e.g. overweight and obese); and ω_i be the number of deaths attributable to a specific exposure category i (e.g. overweight) of the risk factor.

$$\omega = \sum \omega_i \tag{T4}$$

$$\omega_i = M \cdot \text{PAF}_i \tag{T5}$$

Using the notations in Table 2,

$$\text{PAF}_i = \frac{f_i(R_i - 1)}{f_0 + \sum f_i R_i + f_q R_q} \tag{11}$$

Using the notations of Allison et al, equation 11 can be expressed as

$$\text{PAF}_i = \frac{P(O_i)[RR_i - 1]}{P(R) + \sum P(O_i)RR_i + P(Q)RR_q} \tag{11}$$

Using the notations of Mokdad et al, and adding P_q (fraction of population underweight) and RR_q (relative risk for underweight), equation 11 becomes

$$\text{PAF}_i = \frac{P_i[RR_i - 1]}{P_0 + \sum P_i RR_i + P_q RR_q} \tag{11}$$

Eight steps to estimate ω :

Step 1: Obtain M , total no. of deaths in the population during 1 y (from death records).

Step 2: Obtain f_0 [or $P(R)$ or P_0], percentage of individuals in the population nonexposed (e.g. normal weight; never smokers) (from health surveys).

Step 3: Obtain f_i [or $P(O_i)$ or P_i], percentage in separate exposure categories of the risk factor (e.g. overweight or obese; current occasional smokers or current daily smokers) (from health surveys).

Step 4: Obtain f_q [or $P(Q)$ or P_q], percentage in a neutral category (e.g. underweight; former smokers) (from health surveys).

Step 5: Obtain R_i [or RR_i], relative risk of death for each separate exposure category (e.g. former smokers, current smokers) relative to none (e.g. never smokers) (from prospective cohort studies). When R_i is not available and when the event (e.g. death) is rare, relative risk can be approximated by HR (ie, h and q).

Step 6: Calculate PAF_i , population attributable fraction, using equation 11.

Step 7: Calculate ω_i , using equation T5.

Step 8: Calculate ω using equation T4.

extends the concept and equation of population attributable fraction, originally proposed by Levin [3], from a two-category risk factor (nonexposed, exposed) (equation 1) to a three-category risk factor (nonexposed, exposed, and neutral) (equation 8).

Both our proposed procedure (Table 6) and the Allison procedure allow for one or more neutral categories (such as underweight). The numbers of death associated with the neutral categories are excluded from the calculation of the number of death attributable to the risk factor under study. The Allison procedure involves twelve steps while the proposed procedure involves only eight steps. The proposed procedure, using the logic of the Mokdad procedure, does not require calculation of the probabilities of death in the various categories. Therefore no solving for the hazard λ is required. The proposed procedure is expected to produce similar results as the more complex Allison procedure. Slight discrepancies in the results, as shown in the real examples provided in this paper, are due to rounding errors

in the additional steps in the Allison procedure to estimate probabilities of death in various nonexposed, exposure and neutral categories, and solving for λ , the hazard in the nonexposed, all of which are not required in the proposed procedure. Discrepancies will also occur depending on whether relative risks or hazard ratios are used, but this is expected to be small when the event (e.g., death) is rare (section 4). If one insists to use the Allison procedure instead of the proposed procedure, this paper discusses a number of options for solving for λ which could be helpful (section 5).

Acknowledgements

The author would like to thank Sunita Narang, Justin Francis and Rita Zhang for statistical and data support for this paper. The author declares there is no conflict of interest. No funding or support was received for this study. The author has full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Additional material

Additional file 1: file containing Appendices S1-S4 [7-10].

Author details

¹Centre for Chronic Disease Prevention and Control, Public Health Agency of Canada (PHAC), Government of Canada, Ottawa, Ontario, Canada.
²Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada. ³Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. ⁴Shantou University Medical College, Shantou, China.

Competing interests

The authors declare that they have no competing interests.

Received: 10 September 2009 Accepted: 31 August 2010

Published: 31 August 2010

References

1. Allison DB, Fontaine KR, Manson JE, Stevens J, VanItallie TB: **Annual deaths attributable to obesity in the United States.** *JAMA* 1999, **282**:1530-1538.
2. Mokdad AH, Marks JS, Stroup DF, Gerberding JL: **Actual causes of death in the United States, 2000.** *JAMA* 2004, **291**:1238-1245.
3. Levin ML: **The occurrence of lung cancer in man.** *Acta Union International Contra Cancrum* 1953, **9**:531-541.
4. Last JM: *A Dictionary of Epidemiology* New York: Oxford University Press 1995.
5. Department of Applied Mathematics: **Mathematica, Maple, Matlab, IDL.** [<http://amath.colorado.edu/computing/mmm/>], Accessed July 23, 2010.
6. **bitwise magazine. Maple 10 v Mathematica 5.2.** [http://www.bitwisemag.com/copy/reviews/software/math/maple10_mathematica52.html], Accessed July 24, 2009.
7. Wikipedia: **Newton's method.** [http://en.wikipedia.org/wiki/Newton's_method], Accessed July 23, 2010.
8. WolframMathWorld: **Newton's method.** [<http://mathworld.wolfram.com/NewtonsMethod.html>], Accessed July 23, 2010.
9. Wikipedia: **Taylor series.** [http://en.wikipedia.org/wiki/Taylor_series], Accessed July 23, 2010.
10. WolframMathWorld: **Taylor series.** [<http://mathworld.wolfram.com/TaylorSeries.html>], Accessed July 23, 2010.

doi:10.1186/1742-5573-7-8

Cite this article as: Choi: Population attributable fraction: comparison of two mathematical procedures to estimate the annual attributable number of deaths. *Epidemiologic Perspectives & Innovations* 2010 **7**:8.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

