Environmental Health

LETTER TO THE EDITOR RESPONSE   Open Access

Estimation and application of population attributable fraction in ecological studies

Cheng-Kuan Lin[1]* and Szu-Ta Chen[2]

Abstract

Estimation of population attributable fraction (PAF) requires unbiased relative risk (RR) by using either Levin's or Miettinen's formula, on which decision depends on the available exposure information in reference group, not the types of studies. For ecological studies and studies with aggregated outcomes, once having unbiased RRs, Levin's and Miettinen's formulae would provide identical PAF estimates. PAF could also be applied to compare relative burdens of disease between countries across time, which is an additional information in consideration of country-level policies.

Keywords: Population attributable fraction, Ecological study

To the Editor

Population attributable fraction (PAF) is widely used to measure the disease burden attributable to a given risk factor. Concerns on improper estimation of PAF in ecological studies are raised [1] in consideration of potential ecological bias. However, if unbiased relative risks (RR) are available, estimation of PAF from ecological studies is feasible, either by using Levin's or Miettinen's formula.

In a recent published ecological study [2], for example, the PAF is referred to the proportion of subjects with lung cancer that would have not occurred if coal-fired power plants (the exposure of interest) were absent counterfactually, assuming the same probability of getting lung cancer in the exposed and unexposed groups with the remaining risk factors. Considering a randomly selected country i, the numbers of subjects with and without lung cancer in the exposed and unexposed groups can be presented as Table 1.

$$\text{PAF} = \frac{O_i - E_i}{O_i} = \frac{(a+c) - \left(X\frac{c}{Y} + c\right)}{a+c} = \frac{a - X\frac{c}{Y}}{a+c}$$

where O = observed numbers of subjects with lung cancer; E = expected numbers of subjects with lung cancer had everyone not been exposed in country i.

PAF could also be expressed as either Levin's or Miettinen's formula, respectively.

$$\text{PAF} = \frac{P_e \times (RR-1)}{P_e \times (RR-1)+1} \qquad \text{(Levin's formula) [3]}.$$

$$= P_c \times \frac{RR-1}{RR} \qquad \text{(Miettinen's formula) [4]}.$$

where.

$P_e = \frac{X}{X+Y}$, the proportion of subjects being exposed in the population;

$RR = \frac{\frac{a}{X}}{\frac{c}{Y}} = \frac{a \times Y}{X \times c}$, relative risk of lung cancer comparing the subjects among exposed and unexposed groups;

$P_c = \frac{a}{a+c}$, the proportion of exposed cases among lung cancer subjects.

Mathematically, the PAF calculated by using Levin's formula would be identical to that from Miettinen's formula, regardless whether $P_e = P_c = 1$ or not, as proved below.

Levin's formula:

* Correspondence: chl309@mail.harvard.edu
[1]Department of Environmental Health, Harvard T.H. Chan School of Public Health, 665 Huntington Avenue, Boston, MA 02115, USA
Full list of author information is available at the end of the article

Table 1 Numbers of subjects with and without disease in exposed and non-exposed groups

	With disease (Cases, lung cancer)	Without disease (reference)	Total
Exposed group	a	b	X
Unexposed group	c	d	Y

$$PAF = \frac{P_e \times (RR-1)}{P_e \times (RR-1) + 1} = \frac{\frac{X}{(X+Y)} \times \frac{(aY-Xc)}{Xc}}{\frac{X}{(X+Y)} \times \frac{(aY-Xc)}{Xc} + 1}$$

$$= \frac{\frac{XaY-X^2c}{(X+Y)Xc}}{\frac{X(aY-Xc)+(X+Y)c}{(X+Y)c}}$$

$$= \frac{X(aY-Xc)}{X(aY-Xc+Xc+Yc)} = \frac{aY-Xc}{aY+cY} = \frac{a-X\frac{c}{Y}}{a+c}$$

Miettinen's formula:

$$PAF = P_c \times \frac{RR-1}{RR} = \frac{\left(\frac{a}{a+c}\right) \times \left(\frac{aY-Xc}{Xc}\right)}{\frac{aY}{Xc}} = \frac{aY-Xc}{aY+cY}$$

$$= \frac{a-X\frac{c}{Y}}{a+c}$$

One major reason that researchers commonly use Levin's formula to estimate PAF in case-control studies and use Miettinen's formula in cohort studies, respectively, is because of data availability. In case-control studies, researchers have sufficient information of the four cells in Table 1. Whereas in cohort studies and studies with aggregated data (e.g., standardized incidence ratio (SIR) and standardized mortality ratio (SMR) studies), most researchers do not have the exposure information among subjects without diseases (i.e. b and d cells in Table 1), and thus P_e is not available. Once having sufficient information in observational studies, the use of either Miettinen's or Levin's formula would yield an identical PAF estimate.

Furthermore, although both requiring unbiased RRs for estimation of PAF in Levin's and Miettinen's formulae, methods to obtain the unbiased RRs are different according to study types, such as using confounding adjustment (mostly from regression models) in case-control studies and stratification (e.g., SMR stratified by age and sex) in observational studies [5]. Theoretically, unbiased RRs from either fully adjustment (i.e., no residual confounding) or fully stratification (i.e., no residual confounding within stratum) should be identical and feasible for PAF estimation [6]. World Health Organization and Institute for Health Metrics and

Evaluation applied a hybrid method, which age- and sex-stratified RRs were retrieved by prior meta-analysis or regression models, and summarized in standardized populations to estimate PAFs (i.e., global burden of diseases) [7]. Given fully adjusted and unbiased estimation, RRs derived from ecological studies and other studies with aggregated outcomes would be as valid as those from case-control studies, and therefore, are legit for PAF estimation.

Lastly, PAF estimates could be interpreted as relative strength of a relationship between exposure and disease, regardless of the nature of association or causation [8], and subsequently be applied to compare relative burden of diseases across countries/populations. In Lin's study [2], for example, relative burden of lung cancer between countries across time contribute valuable information in consideration of country-level policies.

In conclusion, valid PAF estimation could be achieved from both Levin's and Miettinen's formulae with sufficient information in different types of studies with unbiased RR estimation, regardless stratification or adjustment. Comparison of PAFs between countries across time might provide additional information, along with the point estimate per se.

Author details
[1]Department of Environmental Health, Harvard T.H. Chan School of Public Health, 665 Huntington Avenue, Boston, MA 02115, USA. [2]Department of

Epidemiology, Harvard T.H. Chan School of Public Health, 667 Huntington Avenue, Boston, MA 02115, USA.

References

1. Khosravi, A. & Mansournia, M.A., Issues with incorrect computing of population attributable fraction (PAF) in a global perspective on coal-fired power plants and burden of lung cancer, Environmental Health, 2019, https://doi.org/10.1186/s12940-019-0490-6
2. Lin CK, et al. A global perspective on coal-fired power plants and burden of lung cancer. Environ Health. 2019;18(1):9.
3. Levin ML. The occurrence of lung cancer in man. Acta Unio Int Contra Cancrum. 1953;9(3):531–41.
4. Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. Am J Epidemiol. 1974;99(5):325–32.
5. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. Am J Public Health. 1998;88(1):15–9.
6. Morgenstern H. Ecologic studies in epidemiology: concepts, principles, and methods. Annu Rev Public Health. 1995;16:61–81.
7. Murray CJL, L.A., The *Global Burden of Disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in* 1990 *and projected to 2020*. 1996, Cambridge, MA USA: Harvard University Press.
8. Flegal KM, Panagiotou OA, Graubard BI. Estimating population attributable fractions to quantify the health burden of obesity. Ann Epidemiol. 2015; 25(3):201–7.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.