JRHS 2015; 15(1): 22-27

journal homepage: www.umsha.ac.ir/jrhs

Original article

Calculating Population Attributable Fraction for Cardiovascular Risk Factors Using Different Methods in a Population Based Cohort Study

Sayyedeh Sara Azimi (MSc)^a , Davood Khalili (MD, PhD)b* , Farzad Hadaegh (MD, PhD)^b , Parvin Yavari (MD,PhD)^c , Yadollah Mehrabi (PhD)^d , Fereidoun Azizi (MD, PhD)^e

- *^a Department of Epidemiology, School of Public Health, Shahid Beheshti University of Medical Sciences and Center for Non-communicable disease Control, Deputy of Health, Ministry of health and medical Education, Tehran, Iran*
- *^b Prevention of Metabolic Disorders Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran*
- *^c Department of Community Medicine, School of medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran*
- *^d Department of Epidemiology, School of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran*
- *^e Endocrine Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

ARTICLE INFORMATION **ABSTRACT**

Article history: Received: 11 August 2014 Revised: 22 October 2014 Accepted: 08 November 2014 Available online: 24 November 2014

Keywords:

Risk Factor Cardiovascular Disease Relative Risk Population Attributable Fraction Methods

** Correspondence Davood Khalili (MD, PhD) Tel: +98 21 22432500 Fax: +98 21 22416264 E-mail: dkhalili@endocrine.ac.ir*

Background: Population Attributable Fraction (PAF) is one of the most practical measures for estimating the burden of risk factors with some challenges in its calculation. Cardiovascular disease (CVD) is the first cause of death worldwide and the estimation of accurate PAFs for CVD risk factors is of great importance in conducting preventive strategies. Our aim was to estimate the PAFs of CVD risk factors via direct, i.e. based on regression models, and indirect, i.e. using related equations, methods.

Methods: Participants (3200 males and 4245 females aged ≥30 yr) without history of CVD were selected from the population-based cohort of Tehran Lipid and Glucose Study (TLGS). Hazard ratio (HR) and Odds ratio (OR) of conventional risk factors were calculated for CVD events after ten yr of follow-up. Levin's and Miettinen's equations were applied to indirectly estimate the PAFs and average PAF was directly derived from logistic regression model.

Results: The sum of PAFs resulted from indirect estimations reached to more than 100% (around 200% and 150% based on Levin's and Miettinen's formula respectively). The direct estimation attributed 80% and 86% of burden of CVD events to conventional risk factors in men and women respectively. The rank and pattern of PAFs of risk factors was somehow different among different methods.

Conclusions: Estimating priorities of risk factors may differ in different methods for calculating PAF. This study provides evidence on the more expediency of direct method over indirect ways when individual data is available through a population-based cohort.

Citation: Azimi SS, Khalili D, Hadaegh F, Yavari P, Mehrabi Y , Azizi F. Calculating Population Attributable Fraction for Cardiovascular Risk Factors Using Different Methods in a Population Based Cohort Study. J Res Health Sci. 2015; 15(1): 22-27.

Introduction

roportion of incidence of a disease in a general population attributed to a specific risk factor is defined as population attributable fraction (PAF) of that risk factor¹. This proportion of incidence would not occur if the factor were eliminated. By merging the power of risk factor, i.e. its relative risk, with its prevalence, PAF provides an estimation of the relation between risk factor and incidence of diseases at the community level²⁻⁴. Due to the existing methodological challenges, PAF has been less applied as a common public health measure to date; because, most of the approaches that have been declared for the calculation of PAF may yield unacceptable results. For instance, they may yield a more than 100% sum of all existing PAFs, which is implausible⁵. Accordingly, the need for novel approaches of calculating PAF has been addressed in some recent studies^{1, 2}. P

In Iran, CVD is the first cause of death and the burden attributed to the CVD has been ranked third among the national rankings of the diseases burden⁶⁻⁸. PAFs may aid policy makers in anticipating the potential impact of preventive strategies that target certain risk factors at the community level⁹; thus ranking the importance of risk factors based on their PAFs is a key point that may be influenced by applying different methods in calculation of PAF. In the current study, we estimated the PAFs based on indirect methods and a direct method using logistic regression, and ranked the importance of risk factors according to their prevalence, relative risk and their PAFs resulted from different methods.

This study was carried out in the framework of the Tehran Lipid and Glucose Study (TLGS), which is a population, based prospective cohort study $10, 11$.

TRHS

Methods

Participants

TLGS as a population based cohort initiated at 1999- 2001^{10} .TLGS consists of two phases: the first phase of the TLGS was a cross-sectional study that assessed the prevalence of non-communicable diseases and the respective risk factors among participants. In the second longitudinal phase, the incidence of non-communicable diseases and their risk factors are evaluated. The detail of TLGS and its cardiovascular outcomes has been described elsewhere $10,12,13$.

Totally, 7445 (3200 males) participants aged more than 30 yr and without history of CVD in phase I were considered for measurement of prevalence of risk factors; 6630 subjects with at least one year of follow-up (median of follow up $=$ 9.14 yr) entered to Cox regression model to calculate HRs; finally 5868 participants with complete follow-up, up to Mar 2009, at a range of 8-10 yr, were enrolled for calculation of ORs and PAFs through logistic regression model.

All subjects provided informed consent before participation. The study was approved by ethics Committee of the university.

Data collection

Data regarding demographics, physical examination and the laboratory assessments were collected in the phase I of the $TLGS¹⁰$. Follow-ups were performed annually through making phone calls. Additionally, non-communicable health condition events that resulted in hospitalization or death were recorded and related data were collected by a trained physician using hospital records or if needed a home visit; these data were assessed in terms of clinical diagnosis by an outcome committee consisted of internist, cardiologist, endocrinologist, epidemiologist, the physician who collected the data, and other experts as needed¹³. In the current study, incident cases of coronary heart diseases, i.e. myocardial infarction (MI), unstable angina pectoris, and angiography proven CHD and CHD death, and cerebrovascular accidents were recorded as the CVD events.

Exposure status was defined through dichotomized risk factors according to world and national cut pointes $12, 14$ including: Hypertension, systolic blood pressure ≥140 mmHg or diastolic blood pressure ≥90mmHg and/or antihypertensive medication; hypercholesterolemia, cholesterol ≥240 mg/dl or anti-hyperlipidemia medication; diabetes, fasting plasma glucose \geq 126 mg/dl or 2-h post challenge plasma glucose ≥200 mg/dl and/or anti-diabetic medication; general obesity, BMI≥30; central obesity, waist/hip ratio ≥ 0.95 for mail and ≥ 0.90 for female; physical inactivity, rigorous physical activity less than one time a weak; tobacco use, past or current usage of any kind of tobacco; premature family history of CVD, myocardial infarction, stork or sudden death in male first-degree relevant $<$ 55 yr and in female first-degree relevant $<$ 65; high risk age, male ≥45 yr old and female≥55 yr old.

Data analysis

For indirect estimation of PAFs, two equations -Levin's and Miettinen's equations- were applied³. Accordingly, we used the prevalence of CVD risk factors among all subjects and subjects with CVD event for Levin's and Meittinen's equations, respectively. Crude ORs and HRs were estimated

for calculation of PAFs by Levin's and adjusted ORs and HRs for calculation of PAFs by Miettinen's equation; OR based on logistic regression among individuals with complete follow-up and HR based on Cox proportional hazard model among participants with at least 1 year follow-up. In calculating PAF, we selected conventional risk factors of CVD based on reference¹⁴: hypertension, hypercholesterolemia, diabetes, general obesity, central obesity, tobacco use, physical inactivity, high-risk age, family history of premature CVD. However, physical activity and general obesity were excluded from final analysis because they did not show any direct effect on CVD beyond their effect through other risk factors; thus considering them for PAF was nonsense. Smoking did not show any significant effect on CVD in women because it has very low prevalence, so it was excluded from PAF analysis.

Levin's equation is depicted as below:

(1)
$$
PAF = \frac{P_t(RR_{unadj} - 1)}{1 + P_t(RR_{unadj} - 1)}
$$

As, P_t is the prevalence of risk factor in the study population and RR_{unadj} is the crude relative risk of that risk factor and Miettinen's equation as given below:

$$
(2) \ \ PAF = P_c(\frac{RR_{\text{adj}}\text{-}1}{RR_{\text{adj}}})
$$

As, P_c is the prevalence of risk factor among incident CVD cases and RR_{adj} is the adjusted relative risk of that risk factor.

Direct PAF calculation means obtaining PAFs directly from individuals' data using logistic regression. The idea of direct attributable fraction based on logistic regression was introduced by Bruzziet al. 15 and developed by Eide and Gefeller using sequential and average attributable fraction¹⁶. This idea was used practically by Rückinger et al.².

In this regard, it is necessary to consider each risk factor dichotomously. Then irrespective of real exposure for each individual, that factor is removed from the population by coding all observations as unexposed. The predicted probability of CVD event for each individual, with the assumption that there was no exposure to a certain risk factor, is:

(3)
$$
P_{ki} = \frac{1}{1 + \exp[-(\beta_0 + \sum_{j \neq i} \beta_j x_j)]}
$$

As, P_{ki} is representative of predicted probability of CVD event in individual number k, assuming no exposure to a specific risk factor (x_i) ; β_i indicates the regression coefficient of risk factors (x_i) , except risk factor number $i(x_i)$. Subsequently, the sum of all predicted probabilities for all individuals in the sample would be equal to the adjusted estimate of total CVD events, which is anticipated in the absence of that specific risk factor (x_i) . Afterward, PAF was estimated by subtraction of total predicted CVD events from total observed CVD events, divided by the number of total observed CVD events:

$$
(4) \ \ PAF = \frac{CVD_{\rm obs} \text{-}CVD_{\rm pred}}{\text{CVD}_{\rm obs}}
$$

24 *Population attributable fraction calculating methods*

As, CVD_{obs} is the total number of CVD events in the sample and CVD_{nred} is the predicted number of CVD events assuming absence of that specific risk factor.

Indirect methods do not consider the sequence in which variables influence each other. Therefore, individual PAFs can sum to more than 100% because it is assumed each risk factor is the first to be eliminated and there is no association between the risk factors ⁵. The dependence on the removal sequence can be addressed by calculating sequential PAF. The sequential PAF is based on successively removing each exposure from the analysis. This approach depends on the sequence of the exposures removal. The average PAF is an average of the attributable fractions over all the sequences for any risk factor. By removing risk factors in every possible order and using the average over all obtained PAFs, the estimation does not depend on the order sequence anymore 2 .

Calculations in direct method were done via the Macro recommended by Rückinger et al. 2 through the STATA software.

Considering the difference between males and females regarding prevalence of risk factors and their relative risk, all analyses were applied separately by gender.

Results

Baseline characteristic is shown in Table 1. Mean value (SD) of age was 46.3 (11.52) and 47.9 (12.81) in males and females respectively. Prevalence of all risk factors significantly varied by gender, except general obesity and physical inactivity; smoking was more prevalent in men than in women and other risk factors were more prevalent in women.

Table 1: The prevalence of cardiovascular diseases (CVD) risk factors in study population at baseline using chi-square test

^a Prevalence in total population and in subjects with CVD was used by Levin's and Mittinen's formula respectively

Totally, 320 men and 238 women experienced CVD during the study period. Table 2 shows the OR and HR for each risk factor indicating negligible differences between these effect measures. Table 3 shows the PAFs resulted from different methods. The results of indirect methods are reported based on using HR. The results obtained by

Miettinen's equation were lower than that by Levin's formula; however, the sums of PAFs calculated via both equations were more than 100% (around 200% for Levin and 150% for Miettinen). Direct method yielded to a cumulative PAF of 86% and 80% in males and females respectively.

Table 2: Crude and adjusted measures of odds ratio (OR) and hazard ratio (HR) for cardiovascular diseases (CVD) risk factors

^a Adjustment for other risk factors in the table.

Concerning modifiable risk factors the results of direct estimation revealed that, smoking (14.2%), hypertension (11.7%), diabetes (7.3%), hypercholesterolemia (6.9%) and central obesity (5.9%) in males and hypertension (19.3%), diabetes (18.8%), central obesity (9.9%) and hypercholesterolemia (7.9%) in females, had the highest rank of PAFs respectively (Table 3). This rank and the pattern of vigor of PAFs varied based on indirect methods in both genders (Figure 1).

Figure 2 shows the priority of modifiable risk factors based on their PAF, prevalence and adjusted HR in men and women respectively.

Table 3: Population attributable fraction (PAF) of cardiovascular diseases (CVD) risk factors, calculated by Levin's equation (using unadjusted hazard ratios in Cox regression), Miettinen's equation (using adjusted hazard ratios in Cox regression) and direct estimation

Hypercholesterolemia Central Obesity

■ Hypertension ■ Diabetes ■ Hypercholesterolemia ■ Central Obesity

Figure 1: The rank of population attributable fractions (PAFs) of CVD risk factors calculated by using different methods, by sex

Discussion

In the current study, the PAFs of CVD risk factors were calculated based on different methods in a population-based cohort of a Persian community. In this regard, the prevalence **Figure 2:** Ranking of modifiable CVD risk factors according to their prevalence (at baseline), adjusted hazard ratio and direct estimated population attributable fraction (PAF), by sex

rates, ORs and HRs of risk factors were determined and applied to derive PAFs via indirect methods and the direct average PAF was calculated based on regression model as

26 *Population attributable fraction calculating methods*

well. According to our results from the direct method, 86% and 80% of all CVD events could be attributed to the known risk factors in Tehranian males and females, respectively; however the sum of PAFs resulted from indirect methods exceeds 100% and the rank and pattern of PAFs changed.Our results indicated that the use of direct method could resolve the common problem of "sum of PAFs $\geq 100\%$ " which yielded by using Levin's and Miettinen's equations 2 . The major limitation of Levin's equation is that the confounding effect of other risk factors is ignored and the risk factors are presumed to be independent of each other 5 . Using the adjusted ratios in Miettinen's equation resolves the confounding problem; however, some other problems are remained, including interaction between risk factors and the fact that the frequencies of risk factors would not be constant after removing a certain risk factor from the community¹. Direct method leads us to results that are more reasonable because all risk factors are intended together and the sequence of risk factors elimination is considered as well. Consequently, the confounding effect of other risk factors is controlled and because any effect of risk factor is considered with and without the presence of others, the modification effect of other risk factors could be addressed as well.

The rank and pattern of PAFs regarding their magnitude were different among three methods. For instance, in men, based on the Levin's, central obesity is the 3rd rank but based on direct method and the Miettinen's it is the $6th$ one; this could be due to unadjusted effect of central obesity in Levin's formula. On the other hand, hypertension had the second rank according to the Levin's and Miettinen's but this rank was related to smoking in direct method; this could be explained by the sequence of removing of exposures from the community. If there was no interaction among risk factors, the sum of PAFs would not be above 100% by using adjusted RRs; consequently, the Miettinen's formula and direct method would lead to same results, however differences between these two methods in our analysis showed that there may be some interactions among risk factors. These interactions could be 2-way, 3-way, 4-way or more that are so hard to check.

To date, the direct method for the estimation of PAF has not been frequently applied. In a similar study performed in Germany, highest PAFs were reported for age>60 years, hypertension, smoking, hypercholesterolemia, male gender, low High Density Lipoprotein (HDL) and diabetes, respectively². Since that, the results were not separately calculated for males and females, the rankings of PAFs could not be compared with our study. Age is a surrogate of some unmeasured risk factors accumulate during the life e.g. air pollution, mental stress or oxidative exposures, thus it could explain a wide range of burden of CVD especially in men. In general, 46% and 56% of all CVD events were attributed to modifiable risk factors among our men and women respectively. It means that controlling the known conventional risk factors could prevent about half of burden of CVD in our population. It is noteworthy that comparing the results of studies reported the PAFs among different nations faces multiple limitations: varying age and sex distribution of the target populations, application of different definition of risk factors, different set of risk factors and confounders under consideration, and finally different methods for the calculation of PAFs.

Finally we compared the priority of risk factors regarding their prevalence, relative risk (according to adjusted HR) and PAF (according to direct method) (Figure 2) and as it was expected, the priorities differed based on these measurements. Regarding causality, the most important risk factors in men were hypertension and diabetes, while concerning prevention in community level; smoking was the first priority because of its high prevalence in male population. It shows that PAF as a surrogate of relative risk and prevalence could be a useful effect measures at community level if it is calculated in a valid way.

The main limitations of using direct method in calculation of PAF is that it cannot be estimated by using survival models. However, it was of strength of our study that completely followed up 80% of participants and HRs did not noticeably differ from ORs. On the other hand, the estimated PAFs provide a somewhat theoretical concept on the attribution of risk factors to the outcome of interest. In fact, total elimination of risk factors by conducting preventive strategies in the community level is practically impossible and the direct methods should be developed to consider more practical indices like Potential Impact Fraction (PIF) instead of $PAF^{9,1}$. PIF defined as the fractional reduction of a disease resulting from changing the current level of a risk factor to other modified levels, not necessarily to the not-exposed level that consider in PAF; because lowering the level of risk factors, especially by using community-based interventions is not hundred percent effective¹⁸. For instance, Karami et al. applied the PIF of diabetes to update the burden of CVDs in Iran by a feasible minimizing $risk^8$. However, PAF is the common used and appropriate tool for priority setting in reducing incidence of diseases, as previously applied for CHD and CVD in the TLGS $^{19, 20}$. However, in interpreting and using of this measure, this point must be considered.

As a limitation of our study, we considered individuals who take medication for hypertension as high risk and hypertensive subjects. In our population, thirty-six percent of individuals with hypertension are using antihypertensive medication and of these about 40% have normal blood pressure i.e. lower than 15% of hypertensive patients 21 . Thus, because many hypertensive patients who take medicine do not have a controlled hypertension, we preferred not to consider them as normal subjects. On the other hand, reanalyzing the data considering them as normal subjects showed lower OR and PAF for hypertension. There was the same pattern for hypercholesterolemia and diabetes (data available on request). These points show the importance of those who take medicine but are still at risk.

When the sum of PAFs exceeds 100%, the PAF belongs to any risk factor does not show an actual amount of outcome which could be prevented by eliminating that risk factor. Accordingly, from theoretical point of view, it is clear that the direct method, based on actual data, is better than indirect methods in terms of controlling confounding and estimating cumulative PAF less than 100%; consequently, we reflected a preference for direct method and using averaged PAF. However, we do not know the truth and the results from actual data are not the grounds of the superiority of the direct method. The authors think that the truth will come from randomized controlled trials by controlling risk factors sequentially over all the sequences for any risk factor, but believe that in reality, it is about impossible to achieve.

Conclusions

This study provides evidence on the more expediency of direct method over indirect ways –Levin's and Miettinen's equations- for the estimation of PAF when individual data is available through a population-based cohort. We showed that indirect and direct methods for calculating PAFs result in different rankings and patterns of PAFs and as a result indicate different priorities in policymaking and we should be aware of faults yielded from the routine indirect methods. Based on direct methods we found that controlling traditional modifiable risk factors could result in controlling 50% of burden of CVD in our population; this estimation could not be taken by using indirect methods because the sum of PAFs exceeds 100%.

Acknowledgements

This work was derived from Sayyedeh Sara Azimi's MSc thesis in Department of Epidemiology, School of Public Health, Shahid Beheshti University of Medical Sciences and was supported by Prevention of Metabolic Disorders Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences.

Conflict of interest statement

The authors declare no conflict of interest.

References

- **1.** Mason CA, Tu S. Partitioning the population attributable fraction for a sequential chain of effects. *Epidemiol Perspect Innov.* 2008;5:5.
- **2.** Rückinger S, von Kries R, Toschke AM. An illustration of and programs estimating attributable fractions in large scale surveys considering multiple risk factors. *BMC Med Res Methodol.* 2009;9:7.
- **3.** Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. *Am J Public Health.* 1998;88(1):15-19.
- **4.** Land M, Vogel C, Gefeller O. Partitioning methods for multifactorial risk attribution. *Stat Methods Med Res.* 2001;10(3):217-230.
- **5.** Rowe AK, Powell KE, Flanders WD. Why population attributable fractions can sum to more than one. *Am J Prev Med.* 2004;26(3):243-249.
- **6.** Jafari N, Abolhassani F, Naghavi M, Pourmalek F, Moradi Lakeh M, Kazemeini H, et al. National burden of disease and study in Iran. *Iran J Public Health.* 2009;38(Suppl1):71-73.
- **7.** Naghavi M, Abolhassani F, Pourmalek F, Moradi Lakeh M, Jafari N, Vaseghi S, et al. The burden of disease and injury in Iran. 2003. *Popul Health Metr.* 2009;7:9.
- **8.** Karami M, Khalili D, Eshrati B. Estimating the proportion of diabetes to the attributable burden of cardiovascular diseases in Iran. *Iran J Public Health.* 2012;41(8):50-55.
- **9.** Levine B. Peer Reviewed: What Does the Population Attributable Fraction Mean? *Prev Chronic Dis.* 2007;4(1):1-5.
- **10.** Azizi F, Rahmani M, Emami H, Mirmiran P, Hajipour R. Cardiovascular risk factors in an Iranian urban population: Tehran Lipid and Glucose Study (Phase 1). *Soz Praventivmed.* 2002;47(6):408-426.
- **11.** Azizi F, Ghanbarian A, Momenan AA, Hadaegh F, Mirmiran P, Hedayati M, et al. Prevention of non-communicable disease in a population in nutrition transition: Tehran Lipid and Glucose Study phase II. *Trials.* 2009;10:5.
- **12.** Hadaegh F, Zabetian A, Sarbakhsh P, Khalili D, James W, Azizi F. Appropriate cutoff values of anthropometric variables to predict cardiovascular outcomes: 7.6 years follow-up in an Iranian population. *Int J Obes.* 2009;33:1438-1445.
- **13.** Khalili D, Mosavi-Jarrahi AR, Eskandari F, Mousavi-Jarrahi Y, Hadaegh F, Mohagheghi MA, et al. Evaluation of Cause of Deaths' Validity Using Outcome Measures from a Prospective, Population Based Cohort Study in Tehran, Iran. *PLoS One.* 2012;7(2):e31427.
- **14.** Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *JAMA.* 2001;285(19):2486-2497.
- **15.** 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- 914.
- **16.** Eide GE, Gefeller O. Sequential and average attributable fractions as aids in the selection of preventive strategies. *J Clin Epidemiol.* 1995;48(5):645-655.
- **17.** Ezzati M, Lopez AD, Rodgers A, Murray CJL. *Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors.* Geneva: WHO; 2004.
- **18.** Azizi F, Mirmiran P, Momenan AA, Hadaegh F, Habibi Moeini A, Hosseini F, et al. The Effect of Community-Based Education for Lifestyle Intervention on the Prevalence of Metabolic Syndrome and Its Components: Tehran Lipid and Glucose Study. *Int J Endocrinol Metab.* 2013;11(3):145-153.
- **19.** Khalili D, Haj Sheikholeslami F, Bakhtiyari M, Azizi F, Momenan AA, Hadaegh F. The Incidence of Coronary Heart Disease and the Population Attributable Fraction of Its Risk Factors in Tehran: A 10-Year Population-Based Cohort Study. *PLoS One.* 2014;9(8):e105804.
- **20.** Azimi SS, Khalili D, Hadaegh F, Yavari P, Mehrabi Y, Azizi F. Direct Estimate of Population Attributable Fraction of Risk Factors for Cardiovascular Diseases: Tehran Glucose and Lipid Study. *Iran J Epidemiol.* 2012;7(4):9-18.
- **21.** Azizi F, Ghanbarian A, Madjid M, Rahmani M. Distribution of blood pressure and prevalence of hypertension in Tehran adult population: Tehran Lipid and Glucose Study (TLGS), 1999- 2000. *J Hum Hypertens.* 2002;16(5):305-312.