Human Biology Review (ISSN 2277 4424) 5 (3) 2016
©Human Biology Review
Original scientific paper (Choudhury and Roy pp. 255-283)
www.humanbiologyjournal.com
Revised and Accepted on April 30, 2016

Population attributable fraction analysis of leading chronic diseases in India

A. Choudhury[1] and P. Roy[2]

Citation: Choudhury A and Roy P. 2016. Population attributable fraction analysis of leading chronic diseases in India. Human Biology Review 5 (3), 255-283.

[1]Amit Choudhury, Department of Statistics, Gauhati University, Guwahati 781014, India. Email:achoudhury@rediffmail.com

[2]Paramita Roy, Department of Statistics, Gauhati University, Guwahati 781014, India. Email:paramita_roy08@rediffmail.com

Corresponding author: Paramita Roy, Department of Statistics, Gauhati University, Guwahati 781014, India. Email:paramita_roy08@rediffmail.com

ABSTRACT

Chronic diseases and their associated risk factors are increasing in India. We aim to quantify the Population Attributable Fractions (PAF) of leading chronic diseases in India associated with significant modifiable risk factors. In calculating adjusted population attributable fraction, non modifiable risk factors are taken as confounders. Our findings highlight that an agenda to improve public health in India must include effective interventions to control tobacco use for cancer and heart disease prevention. There is also an urgent need to educate the general public to maintain proper BMI level thereby reducing diabetes burden in India. The analysis is based on a country wide large scale survey.

Keywords: Chronic disease, risk factors, logistic regression, population attributable fraction

1. INTRODUCTION

India is undergoing a rapid epidemiologic transition characterized by an increase in the prevalence of chronic diseases. In India deaths due to chronic diseases were 3.78 million in 1990 and are expected to reach 7.63 million in 2020 (Kumar et al. 2011). Several risk factors both modifiable and non modifiable contribute to such an alarming increase in the prevalence of chronic diseases. Out of the collection of risk factors, we define modifiable risk factors as risk factors which can be prevented and modified. Examples of modifiable risk factors include poor diet, lack of physical activity, smoking, excessive alcohol intake, obesity, high blood pressure etc. Non modifiable risk factors include age and gender, because there is nothing that can be done to change these things. Interventions are necessary to reduce the prevalence of these risk factors. This is because should their prevalence come down, it would enormously help in combating the increasing epidemic of chronic diseases. Health interventions involve prioritization, and thus the impact of each possible risk factors should be known. Being able to predict the impact of removing a particular risk factor on the risk of developing a disease is an important public health consideration.

A statistical concept that can be used to quantify the impact is the PAF. The PAF is commonly used in epidemiology to describe the proportion of disease that is due to a particular causal factor (Levin 1953). This metric incorporates the prevalence of the risk factor such that interventions that decrease common risk factor reduce disease more than intervention that eliminates uncommon risk factor (Katz 2006). The PAF has been widely used in the medical literature to investigate risk in the population (Bousser and Kittner 2000, Palli et al. 2001) because of the failure of other measures of association to take into account the number of cases in the population.

Studies conducted by Sugathan and Sankaranarayan (2008) and Laskar et al. (2010) in India have shown an increase in the prevalence of risk factors of chronic diseases. The increased prevalence of these risk factors is a likely consequence of the change in lifestyle pattern of the Indian population thereby leading to an alarming increase of chronic diseases. As modifiable risk factors are given primary importance in chronic disease prevention, so we shall concentrate on them. Therefore in this paper we wish to determine the proportion of disease that is attributable to a specific modifiable risk factor as given by PAF. In this paper, we shall concentrate on the modifiable risk factors of three principal chronic diseases in India viz. cardiovascular disease, diabetes and cancer. There are reasons for which we choose these three diseases. Cardiovascular disease (CVD) is the leading cause of death and disability worldwide. It is projected that by the year 2030, CVD-related mortality will rise up to 25 million, mainly from heart disease and stroke. CVD comprising of coronary heart (CHD) and cerebro-vascular diseases, are currently the leading cause of death globally, accounting for 21.9 per cent of total deaths, and are projected to increase to 26.3 per cent by 2030 (WHO 2008). In 2008, Gupta et al. reported that India alone is burdened with approximately 25% of cardiovascular-related deaths and would serve as a home to more than 50% of the patients with heart ailments worldwide within next 15 years. According to the International Diabetes Foundation 285 million people are diabetics worldwide, that will rise to 438 million by 2030 (Shaw 2010). In India alone, the prevalence of diabetes is expected to increase from 31.7 million in 2000 to 79.4 million in 2030 (Wild et al. 2004). Globally, cancer is one of the top leading causes of death. Ferlay et al. (2010) predicted a global burden of 20.3 million new cases by 2030 compared with an estimated 12.7 million cases in 2008, and a predicted 13.2 million cancer-related deaths worldwide by 2030, up from 7.6 million in 2008. The International Agency for Research on Cancer GLOBOCAN project has predicted that India's cancer burden will nearly double in the next 20 years, from slightly over a million new cases in 2012 to more than 1·7 million by 2035 (Ferlay et al. 2013).

Studies conducted by Yoon et al. (2006), Chan et al. (2009) and Lee et al. (2011) in Asian populations, particularly in Asian Indians, have highlighted the "metabolically obese" phenotype among normal weight individuals. This phenotype, characterized by greater abdominal obesity despite a normal BMI, less muscle mass, higher percentage of body fat and increased propensity for insulin resistance compared with the Western population, renders higher susceptibility for diabetes in Asian populations (Chandelia et al. 1999, Raji et al. 2001).

From the studies described above, it implies that even normal BMI persons are at risk of developing diabetes in India. Hence in addition to overweight/obese persons, the contribution of normal BMI persons to diabetes disease burden in India need to be quantified. Also the burden attributable to each risk factor varies considerably across regions (Brown et al. 2014). Whereas developed countries have made strides in prevention of some cancers, such as tumours caused by smoking, the incidence of these cancers are on the rise in developing countries (Kanavos 2006). Higher incidences of lung cancer occur where there is abundant indoor air pollution, such as in India as well as smoking (Rastogi et al. 2004). Thus in a nutshell, we may state that PAF of leading chronic diseases associated with significant modifiable risk factors need to be calculated in order to understand the impact of the risk factors on the chronic diseases, thereby setting prevention strategies for the risk factors which poses maximum disease burden.

In this paper, we have obtained PAF for chronic diseases such as diabetes, cancer and heart disease associated with significant modifiable risk factors based on a relatively large scale nationally representative survey in India.

2. MATERIALS AND METHODS.

In this paper, we have used data from the Indian Human Development Survey (IHDS), 2005. IHDS was jointly organized by researchers from the University of Maryland and the National Council of Applied Economic Research (NCAER) India. It is a nationally representative multi topic survey of 41,554 households in 1503 villages and 276 towns and cities across all states and union territories of India except Andaman Nicobar and Lakshadweep islands. It includes both individual and household level responses on various topics such as education, employment, health, fertility, and gender relations. Stratified sampling design was used for selecting the sample from all over the country.

In 2009, IHDS was the top most download survey for surveys outside of United States in respect of downloads from ICPSR (Interuniversity Consortium for Political and Social Research located at University of Michigan) archive of over 7000 studies.

IHDS (2005) has two major datasets - Individual dataset and Household dataset. Our analysis was carried out on the individual dataset which consists of 2,15,754 cases, each with 211 variables. The household dataset has 41,554 cases each with 937 variables. Two one-hour interviews in each household covered health, education, employment, economic status, marriage, fertility, gender relations, and social capital.

For the logistic analysis, the dependent variable was the chronic disease (Cancer, Diabetes or Heart), operationalized as a binary response variable – yes (having disease) and no (do not have disease). Independent variables in the logistic regression equation were modifiable risk factors. Estimate of the parameter (B), corresponding standard error, p values, odds ratio and 95% Odds ratios (OR) were calculated and estimates are presented with 95% confidence intervals (CI). p values of 0.05 or less (2-tailed) were considered statistically significant. Data were analyzed using SPSS.

For calculation of Crude PAF, we form Table 1.

Table 1. Cross tabulation of exposure and disease

Exposure	Disease (Yes)	Disease (No)
Present	A	b
Absent	C	d

Crude PAF is given by

$$\text{Crude PAF} = \frac{p_t(RR-1)}{1+p_t(RR-1)}$$

where p_t = prevalence of the risk factor in the population

$$= \frac{a+b}{a+b+c+d}$$

RR = unadjusted relative risk

$$=\frac{\left(\frac{a}{a+b}\right)}{\left(\frac{c}{c+d}\right)}$$

For calculation of adjusted PAF with confounder C we form the following two tables. The confounder C is divided into two parts C_1 and C_2. Then for each of C_1 and C_2, cross tabulation of exposure and disease have been done and presented in tables 2 and 3 respectively.

Table 2. Cross tabulation of exposure and disease for C_1

Exposure	Disease (Yes)	Disease (No)
Present	a_1	b_1
Absent	c_1	d_1

Table 3. Cross tabulation of exposure and disease for C_2

Exposure	Disease (Yes)	Disease (No)
Present	a_2	b_2
Absent	c_2	d_2

Adjusted PAF is given by

$$\text{Adjusted PAF}=\frac{p_d\left(RR_a-1\right)}{RR_a}$$

where p_d = prevalence of exposure among cases of disease

$$=\frac{a_1+a_2}{a_1+c_1+a_2+c_2}$$

$RR_a =$ Relative risk adjusted for confounding factor.

$$=\frac{w_1RR_{C_1}+w_2RR_{C_2}}{w_1+w_2}$$

$$w_1=a_1+b_1+c_1+d_1$$

$$w_2=a_2+b_2+c_2+d_2$$

$$RR_{C_1} = \frac{\left(\dfrac{a_1}{a_1 + b_1}\right)}{\left(\dfrac{c_1}{c_1 + d_1}\right)}$$

$$RR_{C_2} = \frac{\left(\dfrac{a_2}{a_2 + b_2}\right)}{\left(\dfrac{c_2}{c_2 + d_2}\right)}$$

2.1 Exclusion criterion.

Internal consistency checks have been carried out before carrying out analysis. From the IHDS individual database, the following have been excluded.

a) Analysis of chronic diseases such as diabetes, cancer and heart diseases is restricted to population aged 22 years or above as the prevalence of these diseases is almost negligible in ages less than 22 years. The analysis of risk factors is also limited to people aged 22 years or more (adults).
b) For marital status, all categories other than married, single, widowed and separated / divorced are excluded from analysis.
c) With regards to the diseases diabetes, cancer and heart, a few respondents are classified as 'cured' in the database. These are excluded from analysis.
d) In the IHDS data, individuals have responded one of "Never, Sometimes or Daily" when asked about their frequency of tobacco smoking. The second response is ambiguous and does not quantify frequency of tobacco consumption. We have thus excluded "Sometimes" category of tobacco smokers from our analysis.

2.2 Data cleaning.

Data cleaning has been done for height and weight data. Weight less than 35 kg or more than 150 kg are discarded from the variables 'weight1' and 'weight 2' of IHDS database. For cleaning height data, values less than 121.9 cm (i.e. 4 ft) are discarded from the variable height. In the context of India, these extremes appear improbable.

After exclusions and data cleaning from individual dataset of IHDS (2005), 116255 cases are analyzed.

2.3 Construction of new variables.

Some variables needed for our analysis which are not available in the IHDS (2005) database have to be constructed in order to proceed with the analysis. For example BMI (Body Mass Index) is one of the risk factors of chronic disease. As this variable is not included in individual IHDS data, we have constructed the same. Information on familial food consumption such as edible oil, milk product and non veg (non vegetarian) are available in household data of IHDS (2005). Information of household size is also available in it. Using these, per capita edible oil consumption, per capita milk product consumption and per capita non veg consumption for each family have been computed. We have then exported these to the individual database of IHDS (2005). Then we have constructed quartiles for each of these diet related variables. To compare consumption of respondents consuming high quantities of edible oil, non veg etc with those consuming lower quantities of the same, we have constructed quartiles of the consumption and created quartile classes.

2.4 Recoding of variables.

We have categorised some of the scale variables and recategorised a few of the categorical variables as follows.

(i) BMI is coded as Underweight (BMI<18.5 kg/m ^{2}), Normal Weight (18.5 kg/m ^{2} <BMI<24.9 kg/m ^{2}) and Overweight (BMI>25 kg/m ^{2}).These categories are used worldwide.

(ii) Age is classified into two categories viz. between 22 to 40 years, above 40 years of age.

(iii) For each case, the IHDS database contains “education completed years”. This is reclassified with the following categories: Less than HS consisting of Illiterate, Primary (Class1to Class 5) and Secondary (Class 6 to Class 10) and HS and above consisting of HS/College (Class 11 to class 14) and Graduate (Class 15).

(iv) The IHDS database contains variable “highest male education” in the family. This is categorized similar to “education completed years”.

(v) For each case, the IHDS database contains another variable “highest female education” in the family. This is also categorized as in (iii) above

(vi) For each case, the IHDS database contains “highest adult education”. This is categorized as in (iii) above.

(vii) For each family in the survey, total edible oil consumption per month is reported. The same when divided by household size provided per capita consumption – each adult member of the household can hence be assumed to consume that amount of edible oil per month. Quartiles of the same are computed. Each respondent of the individual dataset are then classified to belong to a particular quartile class. The constructed variable "familial per capita edible oil consumption" is classified into four categories viz. less than 1st quartile, 1st-2nd, 2nd-3rd, 3rd-4th quartile. The quartile boundaries are less than 0.39999 litres per month, between 0.4000 litres to 0.5713999 litres per month, between 0.5714 litres to 0.83332999 litres per month and 0.8333 litres through highest per month.

(viii) Per capita milk product consumption per family is categorized in a manner similar to (vii) above. The quartile boundaries are less than 0.2221999 kg per month, between 0.2222 kg to 0.39999 kg per month, between 0.4000 kg to 0.85700009 kg per month, 0.8571 kg through highest per month. Milk product consumption includes butter, cream, curd, paneer, ghee etc.

(ix) Per capita non veg consumption is also classified into four categories similar to (vii) above. The quartile boundaries are less than 0.24999 kg per month, between 0.2500 kg to 0.49999 kg per month, between 0.5000 kg to 0.79999 kg per month, 0.8000 kg through highest per month. Non veg consumption includes consumption of meat and fish.

(x) Caste/ religion is categorized into five groups viz. (i) Hindu Brahmin/ High Caste, (ii) Hindu General, (iii) Hindu Other than general, (iv) Muslim and (v) Other religion. Other than general category consists of OBC (other backward classes), Dalit and Adivasi and Other religion consists of Sikh, Jain and Christian.

3. ANALYSIS AND DISCUSSION

3.1 The Risk Factors.

In our paper, we aim to quantify the proportion of disease that is attributable to a specific modifiable risk factor. For this purpose, we started by examining and quantifying the association between modifiable risk factors and chronic diseases of heart disease, diabetes and cancer. The technique used is logistic regression. The modifiable risk factors chosen are tobacco smoking, BMI, high BP, per capita edible oil consumption, per capita milk product consumption and per capita non veg consumption.

Logistic regression analyses were carried out separately for heart disease, diabetes and cancer with respect to different modifiable risk factors.

3.2 Logistic Regression analysis of heart disease

Table 4 presents logistic regression analysis of heart disease with respect to different modifiable risk factors.

Table 4: Results of Logistic Regression Analysis for heart disease

Variables	B	S.E	p- value	Odds Ratio	95% C.I. for Odds Ratio	
					Lower	Upper
Tobacco Smoke						
Daily	0.284	0.166	0.088**	1.329	0.959	1.841
Never®						
BMI						
Normalweight	0.155	0.223	0.487	1.168	0.754	1.808
Overweight	0.330	0.240	0.168	1.391	0.870	2.226
Underweight®						
High BP						
Yes	-0.062	0.081	0.446	0.940	0.802	1.102
No®						
Edible oil consumption						
1st-2nd quartile	-0.073	0.120	0.544	0.930	0.734	1.177
2nd-3rd quartile	0.076	0.113	0.502	1.079	0.864	1.348
3rd-4th quartile	0.447	0.105	0.000*	1.563	1.272	1.920
Less than 1st quartile®						
Milk product consumption						
1st-2nd quartile	0.157	0.188	0.403	1.170	0.810	1.691
2nd-3rd quartile	0.150	0.180	0.406	1.161	0.816	1.654
3rd-4th quartile	-0.061	0.181	0.736	0.941	0.659	1.343
Less than 1st quartile®						
Non veg consumption						
1st-2nd quartile	0.292	0.179	0.103	1.339	0.943	1.901
2nd-3rd quartile	0.135	0.188	0.475	1.144	0.791	1.656
3rd-4th quartile	0.509	0.172	0.003*	1.664	1.189	2.329

Less than 1st quartile®						

* denotes significant at 5%

** denotes significant at 10%

® denotes reference category

3.3 Logistic Regression analysis of diabetes

Table 5: Results of Logistic Regression Analysis for diabetes

Variables	B	S.E	p- value	Odds Ratio	95% C.I. for Odds Ratio	
					Lower	Upper
Tobacco Smoke						
Daily	-0.163	0.128	0.205	0.850	0.661	1.093
Never®						
BMI						
Normal weight	1.282	0.375	0.001*	3.603	1.728	7.514
Overweight	2.105	0.375	0.000*	8.205	3.931	17.123
Underweight®						
High BP						
Yes	0.548	0.060	0.000*	1.730	1.539	1.945
No®						
Edible oil consumption						
1st-2nd quartile	0.631	0.112	0.000*	1.879	1.509	2.341
2nd-3rd quartile	0.911	0.106	0.000*	2.488	2.020	3.064
3rd-4th quartile	1.026	0.104	0.000*	2.789	2.275	3.418
Less than 1st quartile®						
Milk product consumption						
1st-2nd quartile	0.047	0.171	0.784	1.048	0.750	1.464
2nd-3rd quartile	0.332	0.155	0.033*	1.394	1.028	1.889
3rd-4th quartile	0.324	0.152	0.033*	1.382	1.027	1.860
Less than 1st quartile®						
Non veg consumption						
1st-2nd quartile	0.618	0.168	0.000*	1.854	1.335	2.576

2nd-3rd quartile	0.691	0.170	0.000*	1.996	1.429	2.786
3rd-4th quartile	1.167	0.159	0.000*	3.213	2.353	4.387
Less than 1st quartile®						

* denotes significant at 5%

® denotes reference category

3.4 Logistic Regression analysis of cancer

Table 6: Results of Logistic Regression Analysis for Cancer

Variables	B	S.E	p- value	Odds Ratio	95% C.I. for Odds Ratio	
					Lower	Upper
Tobacco Smoke						
Daily	0.761	0.415	0.067**	2.139	0.948	4.827
Never®						
BMI						
Not Underweight	-0.429	0.580	0.459	0.651	0.209	2.029
Underweight®						
High BP						
Yes	-0.769	0.280	0.006*	0.463	0.267	0.803
No®						
Edible oil consumption						
1st-2nd quartile	0.170	0.312	0.585	1.185	0.643	2.185
2nd-3rd quartile	-0.099	0.322	0.759	0.906	0.482	1.703
3rd-4th quartile	-0.091	0.315	0.772	0.913	0.492	1.692
Less than 1st quartile®						
Milk product consumption						
1st-2nd quartile	0.003	0.560	0.996	1.003	0.335	3.003
2nd-3rd quartile	0.318	0.496	0.521	1.375	0.520	3.635
3rd-4th quartile	-0.037	0.521	0.943	0.964	0.347	2.673
Less than 1st quartile®						

Non veg consumption						
1^{st}-2^{nd} quartile	-0.508	0.489	0.299	0.602	0.231	1.569
2^{nd}-3^{rd} quartile	-0.238	0.475	0.616	0.788	0.311	1.999
3^{rd}-4^{th} quartile	-1.282	0.603	0.546	0.719	0.285	1.905
Less than 1^{st} quartile®						

* denotes significant at 5%

** denotes significant at 10%

® denotes reference category

From table 4, we find that tobacco smoking, edible oil consumption and non veg consumption are significant risk factor of heart disease. Table 5 reveals that except smoking, all other modifiable risk factors are significant for diabetes. From table 6, we find that tobacco smoking and high BP are significant risk factors of cancer.

3.5 **Population Attributable Fraction (**PAF)**.**

Having obtained the significant modifiable risk factors for each disease, next we proceed to determine the PAF for each disease given a significant risk factor

a) without confounder
b) with non modifiable risk factor as confounder

In our study, the non modifiable risk factors chosen are age, urbanisation, education completed years, highest male education, highest female education, highest adult education, sex, marital status, caste/religion.

Table 7. Crude and Adjusted PAFs (age as confounder)

Confounder	Chronic disease	Significant risk factor	Crude RR	Adjusted RR	Crude PAF	Adjusted PAF
Age	Heart disease	Tobacco Smoking	1.30138	1.32862	0.14495	0.15496
		Edible oil consumption (risk category is 3^{rd}-4^{th} quartile)	1.47881	1.15811	0.12065	0.05087
		Non veg	1.35297	1.26567	0.09612	0.07733

		consumption (risk category is 3^{rd}-4^{th} quartile)				
	Diabetes	Overweight and Normal	4.66342	4.19150	0.75103	0.72795
		BP (Yes)	1.57817	1.41851	0.13576	0.10933
		Edible oil consumption (risk categories are 1^{st}-2^{nd}, 2^{nd}-3^{rd},3^{rd}-4^{th} quartile)	2.15966	2.43709	0.47427	0.51435
		Milk product consumption (risk categories are 2^{nd}-3^{rd},3^{rd}-4^{th} quartile)	1.29131	1.17084	0.13722	0.08733
		Non veg consumption (risk categories are 1^{st}-2^{nd}, 2^{nd}-3^{rd},3^{rd}-4^{th} quartile)	2.09634	1.75247	0.47142	0.38704
	Cancer	Tobacco Smoking	2.11914	2.18798	0.38728	0.39817
		BP (Yes)	0.46649	0.42326	NA [f]	NA [f]

[f] The PAF cannot be applied to estimations of relative risk <1.

From Table 7, we find that smoking is the most important factor for reducing risk of having heart disease in our population. The fraction of developing heart disease attributable to smoking equals 15.50% after accounting for confounder age, whereas PAF adjusted for age in case of highest quartile of edible oil and non veg are 5.09% and 7.73% respectively. Although adjusting for age of smokers has very little impact on the PAF of heart disease associated with smoking, a moderate impact can be seen in non veg consumption, while high impact can be seen in edible oil consumption. In case of cancer, smoking has the greatest impact on the population attributable to 39.82% of burden associated with cancer after accounting for confounder age. BMI is the diabetes risk factor with the greatest impact on the population attributable to 75.10% of diabetes burden, the figure reduces to 72.80% after accounting for confounder age. Next to BMI comes edible oil consumption attributing 51.44% accounting for confounder age. However the most prominent role of age as confounder is revealed in case of non veg consumption where crude PAF figure is 47.14% and age adjusted PAF is 38.70%. If we do not adjust for confounding by age, we mistakenly conclude that people consuming highest quartile of non veg diet are 2.10 times more likely (crude RR) to develop diabetes than persons who consume less non veg diet, they are actually 1.75 times more likely (adjusted RR) to develop diabetes. To correctly estimate PAF for persons consuming maximum non veg diet, adjusted RR should be used in the adjusted PAF formula to estimate that 38.70% of cases are attributable to maximum consumption of non veg diet. Crude PAF incorrectly estimate that 47.14% of diabetes cases are attributable to maximum consumption of non veg diet.

Table 8. Crude and Adjusted PAFs (urbanisation as confounder)

Confounder	Chronic disease	Significant risk factor	Crude RR	Adjusted RR	Crude PAF	Adjusted PAF
Urbanisation	Heart disease	Tobacco Smoking	1.30138	1.27884	0.14495	0.13661
		Edible oil consumption (risk category is 3rd-4th quartile)	1.47881	1.39929	0.12065	0.10633
		Non veg consumption (risk category is 3rd-4th quartile)	1.35297	1.37284	0.09612	0.10006
	Diabetes	Overweight and	4.66342	4.08285	0.75103	0.72188

		Normal				
		BP (Yes)	1.57817	1.79024	0.13576	0.16358
		Edible oil consumption (risk categories are 1st-2nd, 2nd-3rd,3rd-4th quartile)	2.15966	1.99929	0.47427	0.44147
		Milk product consumption (risk categories are 2nd-3rd,3rd-4th quartile)	1.29131	1.17389	0.13722	0.09010
		Non veg consumption (risk categories are 1st-2nd, 2nd-3rd,3rd-4th quartile)	2.09634	3.17445	0.47142	0.61745
	Cancer	Tobacco Smoking	2.11914	2.17123	0.38728	0.39558
		BP (Yes)	0.46649	0.47495	NA [f]	NA [f]

[f] The PAF cannot be applied to estimations of relative risk <1.

Urbanisation play less important role as confounder in case of significant risk factors associated with heart disease. Adjusting for urbanisation of smokers had little impact on the PAF of heart disease associated with smoking. Similar trend has been observed in respect of the other two significant risk factors which show less differences between crude and adjusted PAFs. The PAF of diabetes attributable to BMI (obese and normal) outweighs that of other risk factors and has the highest PAF value (adjusted for urbanisation). Urbanisation adjusted PAF value of diabetes associated with non veg diet is also high (adjusted PAF= 61.75%). High difference between crude and adjusted PAFs, adjusted for urbanisation suggests that urbanisation plays a prominent role as confounder in case of non veg consumption. The most important risk factor for reducing risk of cancer is tobacco smoking attributable to 39.56% of the disease burden with confounder urbanisation whereas crude PAF estimation is 38.73%. Hence the role of urbanisation as confounder is negligible in this case.

Table 9. Crude and Adjusted PAFs (education completed years as confounder)

Confounder	Chronic disease	Significant risk factor	Crude RR	Adjusted RR	Crude PAF	Adjusted PAF
Education (in completed years)	Heart disease	Tobacco Smoking	1.30138	1.37205	0.14495	0.17091
		Edible oil consumption (risk category is 3rd-4th quartile)	1.47881	1.16847	0.12065	0.05384
		Non veg consumption (risk category is 3rd-4th quartile)	1.35297	1.25329	0.09612	0.07375
	Diabetes	Overweight and Normal	4.66342	4.51662	0.75103	0.74380
		BP (Yes)	1.57817	1.61654	0.13576	0.14421
		Edible oil consumption (risk categories are 1st-2nd, 2nd-3rd,3rd-4th quartile)	2.15966	2.02859	0.47427	0.44783
		Milk product consumption (risk categories are 2nd-3rd,3rd-4th quartile)	1.29131	1.27716	0.13722	0.13159
		Non veg	2.09634	2.04150	0.47142	0.46015

		consumption (risk categories are 1^{st}-2^{nd}, 2^{nd}-3^{rd},3^{rd}-4^{th} quartile)				
	Cancer	Tobacco Smoking	2.11914	1.96443	0.38728	0.35287
		BP (Yes)	0.46649	0.48321	NA [f]	NA [f]

[f] The PAF cannot be applied to estimations of relative risk <1.

Table 10. Crude and Adjusted PAFs (highest male education as confounder)

Confounder	Chronic disease	Significant risk factor	Crude RR	Adjusted RR	Crude PAF	Adjusted PAF
Highest male education	Heart disease	Tobacco Smoking	1.30138	1.36712	0.14495	0.17134
		Edible oil consumption (risk category is 3^{rd}-4^{th} quartile)	1.47881	1.38651	0.12065	0.09892
		Non veg consumption (risk category is 3^{rd}-4^{th} quartile)	1.35297	1.25371	0.09612	0.07342
	Diabetes	Overweight and Normal	4.66342	4.47471	0.75103	0.74416
		BP (Yes)	1.57817	1.66577	0.13576	0.14642
		Edible oil consumption (risk categories are 1^{st}-2^{nd}, 2^{nd}-3^{rd},3^{rd}-4^{th}	2.15966	1.91594	0.47427	0.42056

		quartile)				
		Milk product consumption (risk categories are 2^{nd}-3^{rd},3^{rd}-4^{th} quartile)	1.29131	1.29258	0.13722	0.13547
		Non veg consumption (risk categories are 1^{st}-2^{nd}, 2^{nd}-3^{rd},3^{rd}-4^{th} quartile)	2.09634	2.13831	0.47142	0.47794
	Cancer	Tobacco Smoking	2.11914	2.13699	0.38728	0.39017
		BP (Yes)	0.46649	0.44369	NA [f]	NA [f]

[f] The PAF cannot be applied to estimations of relative risk <1.

Table 11. Crude and Adjusted PAFs (highest female education as confounder)

Confounder	Chronic disease	Significant risk factor	Crude RR	Adjusted RR	Crude PAF	Adjusted PAF
Highest female education	Heart disease	Tobacco Smoking	1.30138	1.30825	0.14495	0.14798
		Edible oil consumption (risk category is 3^{rd}-4^{th} quartile)	1.47881	1.45179	0.12065	0.11536
		Non veg consumption (risk category is 3^{rd}-4^{th} quartile)	1.35297	1.28008	0.09612	0.08023

	Diabetes	Overweight and Normal	4.66342	22.70843	0.75103	0.95030
		BP (Yes)	1.57817	1.62283	0.13576	0.14218
		Edible oil consumption (risk categories are 1st-2nd, 2nd-3rd,3rd-4th quartile)	2.15966	2.10913	0.47427	0.46445
		Milk product consumption (risk categories are 2nd-3rd,3rd-4th quartile)	1.29131	1.29514	0.13722	0.13840
		Non veg consumption (risk categories are 1st-2nd, 2nd-3rd,3rd-4th quartile)	2.09634	2.15302	0.47142	0.30662
	Cancer	Tobacco Smoking	2.11914	1.90478	0.38728	0. 33710
		BP (Yes)	0.46649	0.48945	NA [f]	NA [f]

[f] The PAF cannot be applied to estimations of relative risk <1.

Table 12. Crude and Adjusted PAFs (highest adult education as confounder)

Confounder	Chronic disease	Significant risk factor	Crude RR	Adjusted RR	Crude PAF	Adjusted PAF
Highest adult education	Heart disease	Tobacco Smoking	1.30138	1.38227	0.14495	0.17326
		Edible oil	1.47881	1.40853	0.12065	0.10807

		consumption (risk category is 3rd-4th quartile)				
		Non veg consumption (risk category is 3rd-4th quartile)	1.35297	1.24642	0.09612	0.07284
	Diabetes	Overweight and Normal	4.66342	4.30515	0.75103	0.73397
		BP (Yes)	1.57817	.75021	0.13576	NA [f]
		Edible oil consumption (risk categories are 1st-2nd, 2nd-3rd,3rd-4th quartile)	2.15966	1.89462	0.47427	0.41706
		Milk product consumption (risk categories are 2nd-3rd,3rd-4th quartile)	1.29131	1.30346	0.13722	0.14162
		Non veg consumption (risk categories are 1st-2nd, 2nd-3rd,3rd-4th quartile)	2.09634	2.14476	0.47142	0.48113
	Cancer	Tobacco Smoking	2.11914	2.27247	0.38728	0.41063
		BP (Yes)	0.46649	0.46646	NA [f]	NA [f]

[f] The PAF cannot be applied to estimations of relative risk <1.

For education related variables such as highest male education, highest female education and highest adult education acting as confounders separately, tobacco smoking appears to be the most important factor for reducing risk of heart diseases, next comes edible oil and non veg consumption. The trend is somewhat different in case when education completed years is taken as the confounder. In this case the position of edible oil and non veg are interchanged while tobacco smoking attains the topmost position. Among all education related variables, the least important role as confounder is played by highest female education because of least differences between crude and adjusted PAFs of heart disease associated with smoking. The fraction of developing diabetes attributable to BMI equals 74.38%, 74.42%, 95.03% and 74.40% respectively in case education completed years, highest male education, highest female education and highest adult education are taken as confounders , but in case of education related variables except highest female education, other education related variables had little impact on the PAF of diabetes associated with BMI. BMI plays the most prominent position of reducing risk of diabetes after accounting for education related variables. Next to BMI comes non veg consumption for all education related variables except highest female education where edible oil consumption occupies the second place. However highest female education has high impact on the PAF of diabetes associated with non veg consumption. Tobacco smoking is the most important factor for reducing burden of cancer after accounting for education related variables as confounders.

Table 13. Crude and Adjusted PAFs (sex as confounder)

Confounder	Chronic disease	Significant risk factor	Crude RR	Adjusted RR	Crude PAF	Adjusted PAF
Sex	Heart disease	Tobacco Smoking	1.30138	1.36211	0.14495	0.16655
		Edible oil consumption (risk category is 3^{rd}-4^{th} quartile)	1.47881	1.48877	0.12065	0.12233
		Non veg consumption (risk category is 3^{rd}-4^{th} quartile)	1.35297	1.38078	0.09612	0.10160

	Diabetes	Overweight and Normal	4.66342	4.48289	0.75103	0.70975
		BP (Yes)	1.57817	1.60860	0.13576	0.14020
		Edible oil consumption (risk categories are 1^{st}-2^{nd}, 2^{nd}-3^{rd},3^{rd}-4^{th} quartile)	2.15966	2.29207	0.47427	0.49790
		Milk product consumption (risk categories are 2^{nd}-3^{rd},3^{rd}-4^{th} quartile)	1.29131	1.29430	0.13722	0.13798
		Non veg consumption (risk categories are 1^{st}-2^{nd}, 2^{nd}-3^{rd},3^{rd}-4^{th} quartile)	2.09634	2.12134	0.47142	0.47649
	Cancer	Tobacco Smoking	2.11914	4.08130	0.38728	0.55365
		BP (Yes)	0.46649	0.47445	NA [f]	NA [f]

[f] The PAF cannot be applied to estimations of relative risk <1.

For reducing the risk of developing heart disease, tobacco smoking is the most important factor, accounting for 16.66% of the disease burden after controlling for confounder sex. Edible oil and non veg consumption occupies second and third place contributing to 12.23% and 10.16% of the disease burden, but in all cases considered here impact of sex as confounder is low. The PAF values of diabetes attributable to BMI, edible oil and non veg accounting for confounder sex are high. Sex adjusted PAF for diabetes attributable to BMI is 70.98% whereas crude PAF value is 75.10%. However sex as confounder has little impact for PAF of diabetes associated with blood pressure, milk product consumption and non veg consumption. Tobacco smoking has the greatest impact for reducing risk of cancer attributing for 55.37% of the cancer burden after accounting for confounder sex. This figure reduces to 38.73% in case of crude estimation. If we do not adjust for confounding by sex, we mistakenly conclude that tobacco smokers are 2.12 times more likely to develop cancer than non smokers, they are actually 4.08 times more likely to develop cancer. To correctly estimate PAF for smokers, adjusted RR should be used in the adjusted PAF formula to estimate that 55.37% of cases are attributable to tobacco smoking. Crude PAF incorrectly estimate that 38.73% of the cases are attributable to smoking.

Table 14. Crude and Adjusted PAFs (marital status as confounder)

Confounder	Chronic disease	Significant risk factor	Crude RR	Adjusted RR	Crude PAF	Adjusted PAF
Marital Status	Heart disease	Tobacco Smoking	1.30138	1.26555	0.14495	0.13146
		Edible oil consumption (risk category is 3^{rd}-4^{th} quartile)	1.47881	1.53346	0.12065	0.13096
		Non veg consumption (risk category is 3^{rd}-4^{th} quartile)	1.35297	1.35911	0.09612	0.09715
	Diabetes	Overweight and Normal	4.66342	4.55008	0.75103	0.74187
		BP (Yes)	1.57817	1.58343	0.13576	0.13669
		Edible oil consumption (risk categories are 1^{st}-2^{nd}, 2^{nd}-3^{rd},3^{rd}-4^{th} quartile)	2.15966	2.17176	0.47427	0.47663
		Milk product	1.29131	1.30773	0.13722	0.14349

		consumption (risk categories are 2^{nd}-3^{rd},3^{rd}-4^{th} quartile)				
		Non veg consumption (risk categories are 1^{st}-2^{nd}, 2^{nd}-3^{rd},3^{rd}-4^{th} quartile)	2.09634	2.10022	0.47142	0.47268
	Cancer	Tobacco Smoking	2.11914	2.24253	0.38728	0.40632
		BP (Yes)	0.46649	0.44812	NA [f]	NA [f]

[f] The PAF cannot be applied to estimations of relative risk <1.

Table 15. Crude and Adjusted PAFs (caste as confounder)

Confounder	Chronic disease	Significant risk factor	Crude RR	Adjusted RR	Crude PAF	Adjusted PAF
Caste	Heart disease	Tobacco Smoking	1.30138	1.38746	0.14495	0.17496
		Edible oil consumption (risk category is 3^{rd}-4^{th} quartile)	1.47881	1.46680	0.12065	0.11858
		Non veg consumption (risk category is 3^{rd}-4^{th} quartile)	1.35297	1.34924	0.09612	0.09536
	Diabetes	Overweight and Normal	4.66342	4.52916	0.75103	0.74495
		BP (Yes)	1.57817	1.65419	0.13576	0.14655
		Edible oil	2.15966	2.07695	0.47427	0.45799

		consumption (risk categories are 1st-2nd, 2nd-3rd,3rd-4th quartile)				
		Milk product consumption (risk categories are 2nd-3rd,3rd-4th quartile)	1.29131	1.24684	0.13722	0.12042
		Non veg consumption (risk categories are 1st-2nd, 2nd-3rd,3rd-4th quartile)	2.09634	2.06766	0.47142	0.46545
	Cancer	Tobacco Smoking	2.11914	2.13085	0.38728	0.38918
		BP (Yes)	0.46649	0.46457	NA [f]	NA [f]

[f] The PAF cannot be applied to estimations of relative risk <1.

Marital status play minor role as confounder in case of heart disease associated with significant risk factors. When marital status is taken as the confounder, tobacco smoking and edible oil are the most important factors for reducing risk of heart disease. BMI accounts for 74.19% diabetes burden and tobacco smoking accounts for 40.63% of cancer burden after accounting for confounder marital status. Tobacco smoking is the most important factor for reducing risk of heart disease and cancer after controlling for confounder caste. The PAF of diabetes associated with BMI is 74.50% after controlling for confounder caste.

The values of PAF of heart disease attributed to smoking remain highest in case of confounders of any types. BMI is the most important risk factor of diabetes and the PAF of diabetes associated with BMI is highest in presence of all confounders. BMI is the strongest predictor of diabetes is supported by the fact that the occurrence of type 2 diabetes is mainly due to the ongoing obesity epidemic (Wild et al. 2004). In case of cancer, the value of PAF associated with smoking is highest adjusting for confounders of any types. The value of PAF associated with tobacco smoking in case of cancer is highest when adjusted for confounder sex thereby implying that sex appears to be the strongest confounder associated with tobacco smoking in case of cancer.

According to WHO (2012), about 44% of the diabetes burden are attributable to BMI (overweight and obesity). In our study, BMI constitutes of normal and overweight/obese. Hence our estimate of 75% of diabetes burden attributable to BMI cannot be comparable to such world wide figure. Smoking is a potent risk factor not only for coronary artery disease but also for stroke, the PAF of smoking for CVD is high, at 30% (Matriniuk et al. 2006). Our estimate is only 14% because our data does not contain stroke and computation has been done on heart disease only. According to our analysis PAF of cancer associated with smoking is highest. The tobacco hazard, although clearly linked to the development of lung cancer, also causes an increased risk of several other cancers, notably oral, larynx, pharynx, oesophagus, stomach, liver, pancreas, kidney, bladder, uterine cervix cancers, and myeloid leukaemia (Holmes 2008). Indeed, the World Health Organization (WHO) estimates that 40% of all cancers diagnosed today could have been prevented, partly by maintaining healthy diet, promoting physical activity, and preventing infections that may cause cancer, but largely through tobacco control (WHO 2009). A Brazilian study also assessed the population attributable risk between smoking and developing some types of cancers and reported that the total elimination of smoking would reduce the risk of oesophageal cancer by 54%, of lung cancer cases by 71%, and of cancer of larynx by 86% (Menezes et al. 2002). Our estimate cannot be compared to that of Menezes et al. (2002) because they have obtained PAF of different types of cancer associated with smoking.

4. Conclusion

Our study quantified the theoretical reduction in chronic diseases such as heart disease, diabetes and cancer with the elimination of significant modifiable risk factors in presence as well as in the absence of non modifiable risk factors as confounders. Findings suggest that tobacco smoking is the most important factor for reducing risk of heart disease and cancer. Moreover we find that the PAF of diabetes attributable to BMI (normal weight and overweight) outweighs other risk factors. Hence the findings of the study highlight that an agenda to improve public health in India must include effective interventions to control tobacco use for cancer and heart disease prevention. There is an urgent need to educate the general public to maintain proper BMI level thereby reducing diabetes burden in India.

Our paper is not without limitation. Benichou (1998) pointed out that the PAF for an individual risk factor assumes that all other risk factors have been held constant, i.e. the elimination of the PAF risk factor has no effect on the distribution of other risk factors. For a range of chronic diseases and cancers, this condition is unlikely to hold true. However, the PAF may give a good estimate for potential risk factor reduction and may be helpful for setting priorities.

REFERENCES

1. Benichou J.1998. Attributable risk. In: Armitage P, Colton T, editors. Encyclopaedia of biostatistics. Wiley, Chichester. p 216-219.
2. Bousser MG & Kittner SJ. 2000. Oral contraceptives and stroke. Cephalalgia, 20, 183–189.
3. Brown WJ, Pavey T and Bauman AE. 2014. Comparing population attributable risks for heart disease across the adult life span in women. Br J Sports Med, 1-8.
4. Chan JC, Malik V, Jia W, Kadowaki T, Yajnik CS, Yoon KH, Hu FB. 2009. Diabetes in Asia: epidemiology, risk factors, and pathophysiology. JAMA, 301, 2129–2140.
5. Chandalia M, Abate N, Garg A, Stray-Gundersen J, Grundy SM.1999. Relationship between generalized and upper body obesity to insulin resistance in Asian Indian men. J Clin Endocrinol Metab, 84: 2329–2335.
6. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer; 127: 2893–917.
7. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. 2013. Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, France: International Agency for Research on Cancer. GLOBOCAN 2012 v1.0, 2013. Available from: http://globocan.iarc.fr, accessed on day/month/year (accessed Dec 23, 2013).
8. Gupta R, Joshi P, Mohan V, Reddy KS and Yusuf S. 2008. Epidemiology and causation of coronary heart disease and stroke in India. Heart, 94: 16-26.
9. Holmes JH. 2008. Critical issues in burn care. Journal of Burn Care & Research, 29: pp. S180–S187.

10. Kanavas P. 2006. The rising burden of cancer in the developing world. Annals of Oncology, 17 (Supplement 8): viii15–viii23.
11. Katz MH. 2006. Study design and statistical analysis: a practical guide for clinicians, Cambridge University Press, New York.
12. Kumar NP, Shankaregowda HS, Revathy R. 2011. An assessment of preventable risk factors for chronic non communicable diseases in an adult population. Asian J Epidemiol, 4(1): 9-16.
13. Laskar A, Sharma N and Bhagat N. 2010. Lifestyle disease risk factors in a North Indian community in Delhi. Indian J Community Med, 35 (3): 426-28.
14. Lee JW, Brancati FL, Yeh HC.2011. Trends in the prevalence of type 2 diabetes in Asians versus whites: results from the United States National Health Interview Survey, 1997-2008. Diabetes Care, 34:353–357.
15. Levin ML. 1953. The occurrence of lung cancer in man. Acta Union International Contra Cancrum, 9: 531–541.
16. Matriniuk ALC, Lee CMY, Lam TH, Huxley R, Suh I, Jamrozik K, Gu DF, Woodward M. 2006. The fraction of ischaemic heart disease and stroke attributable to smoking in the WHO Western Pacific and South-East Asian regions. Tob Control, 15:181–188.
17. Menezes AM, Horta BL, Oliveira AL, Kaufmann RA, Suquia R, Diniz A.2002. Attributed risk to smoking for lung cancer, laryngeal cancer and esophageal cancer. Rev Saude Publica, 36:129-134.
18. Palli D, Russo A and Decarli A. 2001. Dietary patterns, nutrient intake and gastric cancer in a high-risk area of Italy. Cancer Causes and Control, 12: 163–172.
19. Raji A, Seely EW, Arky RA and Simonson DC. 2001. Body fat distribution and insulin resistance in healthy Asian Indians and Caucasians. J Clin Endocrinol Metab., 86:5366–5371.
20. Rastogi T, Hildesheim A and Sinha R. 2004. Opportunities for cancer epidemiology in developing countries. Nat Rev Cancer, 4: 909–917.
21. Shaw J, Sicree R, Zimmet P. 2010. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract, 87:4–14.
22. Sugathan TN and Sankaranarayan K. 2008. Behavirol risk factors for non communicable diseases among adults in Kerala, India. Indian J Med Res, 127: 555-563.

23. World Health Organization.2008. World Health Statistics. Department of Measurement & Health Information Systems of the Information, Evidence and Research Cluster. Geneva: WHO Press; 2008. p. 29-31.
24. World Health Organization. 2009. 2008–2013 Action Plan for the Global Strategy for the Prevention and Control of of Noncommunicable Diseases, World Health Organization, Geneva, Switzerland.
25. World health Organization. 2012. Obesity and Overweight. Retrieved September 29, 2012, from http://www.who.int/mediacentre/factsheets/fs311/en/
26. Wild S, Roglic G, Green A, Sicree R and King H. 2004. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care, 27:1047–53.
27. Yoon KH, Lee JH, Kim JW, Cho JH, Choi YH, Ko SH, Zimmet P and Son HY. 2006. Epidemic obesity and type 2 diabetes in Asia. Lancet, 368:1681–1688