Collection of Biostatistics Research Archive
COBRA Preprint Series

Year 2008 *Paper* 40

A SIMPLE INDEX OF SMOKING

Abhaya Indrayan Dr.[*] Rajeev Kumar Mr.[†]
Shridhar Dwivedi Dr.[‡]

[*]Department of Biostatistics and Medical Informatics, UCMS, a.indrayan@gmail.com
[†]Department of Biostatistics and Medical Informative ,University College of Medical Sciences, rajeev.kumar.malhotra@gmail.com
[‡]Preventive Cardiology, Department of Medicine, UCMS, shridhar.dwivedi@gmail.com

A SIMPLE INDEX OF SMOKING

Abhaya Indrayan Dr., Rajeev Kumar Mr., and Shridhar Dwivedi Dr.

Abstract

Background: Cigarette smoking is implicated in a large number of diseases and other adverse health conditions. Among the dimensions of smoking are number of cigarettes smoked per day, duration of smoking, passive smoking, smoking of filter cigarettes, age at start, and duration elapsed since quitting by ex-smokers. The practice so far is to study most of these separately. We develop a simple index that integrates these dimensions of smoking into a single metric, and suggest that this index be developed further. Method: The index is developed under a series of natural assumptions. Broadly, these are (i) the burden of smoking monotonically increases with the cigarette-years but it is more severe in the beginning, (ii) start of smoking early in life is more burdensome than a late start, and (iii) the burden gradually reverses as the duration elapsed since cessation by ex-smokers increases. Result: The index so arrived is: S = (3 – a/15)*1/2*sqrt[sumof(pi*ni*xi) – 0.5] - y for S greater than equal to 0, and sumof(pi*ni*xi) greater than equal to 0.5; otherwise zero (use a =30 for a>30); where i = 1, 2, . . . , I, and I is the number of segments in life with different smoking pattern and a is the age at start of smoking, pi is the proportion of smoke inhaled in case of passive smoking (or adjustment for filter cigarettes or for other forms of smoking), xi is the number of cigarettes smoked for ni years, and y is the number of years elapsed since cessation by ex-smokers. Negative values of S are to be considered equal to zero. Examples are given that demonstrate the use of this index. Conclusion: Just as almost any other composite index, our index too could be good as a comprehensive measure of burden of smoking but not to study its individual dimensions. This measures the present burden in absolute sense and not the risk of smoking-related diseases. Like body-mass index, the smoking index may have good correlation with the risk of some diseases and poor for many others, depending upon the extent to which the risk of disease agrees to our postulations.

A SIMPLE INDEX OF SMOKING

Introduction

Tobacco use is the top cause of disability-years in the world [1]. For the year 2015, WHO projected 6.4 million deaths attributable to tobacco consumption, which would be 10.0% of all deaths [2]. Thus the menace is increasing, and deserves greater attention.

Predominant use of tobacco is in terms of cigarette smoking. Other forms such as pipe, cigar and oral intake are perhaps not as common. Smoking of cigarettes is implicated in a large number of diseases and other adverse health conditions. These range from lung cancer to subfecundity [3-11]. For precise delineation of the role of smoking, it is important that smoking is assessed much more comprehensively than done so far.

One problem that has been consistently faced in studying smoking and its effects is the exact quantification of burden of smoking. Among various dimensions of smoking are number of cigarettes smoked per day, filter and nonfilter cigarettes, duration of smoking, duration elapsed since quitting by ex-smokers, age at start and passive smoking (environmental tobacco smoking). In some cases, features such as depth of inhalation, and time of first cigarette after wake up have also been studied [12]. Sometimes number and size of puffs and butt length are also considered [13]. So far, each of these dimensions is assessed separately for its effect, except for quantity and duration combined as pack-years. Often some dimensions are ignored. Thus a holistic picture is not obtained. Given that it is such an important risk factor, smoking should be studied in more detail. It would be very convenient if an index is available that comprehensively measures different dimensions of smoking by a single metric. No index is available yet that can integrate different dimensions of smoking.

The objective of this communication is to present a simple index that can measure personal burden of smoking in a comprehensive manner. We propose to combine the current and past smoking in terms of duration and cigarettes smoked, active and passive smoking, smoking of filter and regular cigarettes, duration elapsed since cessation by ex-smokers, and the age at start. Other features such as depth of inhalation and time of first cigarette after wake up are not included for two reasons. First, that would make the index too complex that could adversely affect its adoption. Second, these features are rarely studied probably because their effect is not considered substantial. In a study on risk of various types of lung cancer cells [12] indeed these features have been found of not much significance, although in a study on myocardial infarction, depth of inhalation was found significant [14].

We assume that the burden of passive smoking can be measured as a certain proportion of burden of active smoking. Further, we assume that the burden of smoking of filter cigarettes can be expressed as proportion of that of regular cigarettes. To develop the index, we also assume the following. All these are explained later.

1. The dose of smoking can be measured in terms of cigarette-years.
2. The burden monotonically increases strictly as the dose of smoking increases. There is no beneficial effect nor is there a state that smoking is harmless (unless it is as mild as stated in condition 2). See curve-1 and curve-2 respectively in Figure 1 that we exclude from the purview of our index.
3. The burden of each additional cigarette-year of smoking is not as much as the previous cigarette-year. The rate of increase in burden declines with dose of smoking though the burden itself continues to rise. This implies, for example, that first 5 years of smoking is more burdensome than the additional burden by smoking for 5 more years after, say, smoking for 10 years is already done. The admissible shape is designated as Working Hypothesis-I in Figure 1. See curve-3 that we keep outside the purview of our index.
4. Start of smoking at early age causes more burden than a start late in life, and the relationship is linear (Figure 1). But after certain age, say 30 years, the age at start has the same influence as start

at age 30 years. This is our Working Hypothesis-II. We explain later that very few start smoking after the age of 30 years.

5. The cumulative burden of smoking gradually reverses in a linear fashion as the duration elapsed since cessation increases (Figure 1). This is our Working Hypothesis-III.

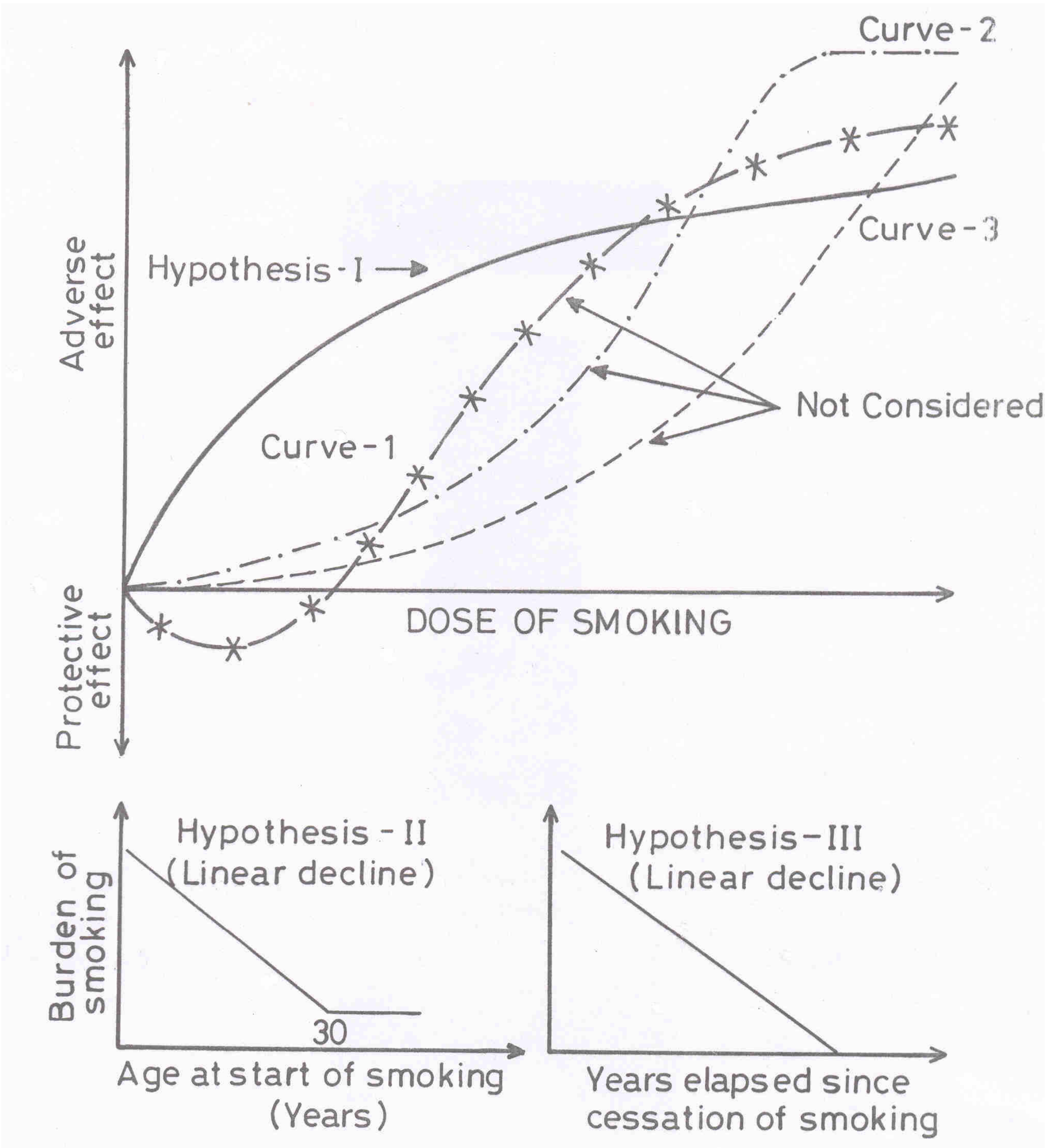


Figure 1. The hypothesized trend of smoking index and some examples of trend excluded from the purview of the index

These conditions are not all that restrictive as may seem. Some of these are conjectures and in the domain of epistemic uncertainties for which almost no data are available. We try to justify them as we proceed to develop the index.

It may be added as a word of caution that our index is not meant to measure the risk of any disease such as lung cancer. The index is an absolute quantity by itself that is intended to measure the current burden of cumulative smoking done by an individual. It may or may not relate well to the risk of smoking-related diseases. Thus this index has the same nature as body-mass index that relates well to the incidence of some diseases but not of so many other diseases.

Development of the index

Cigarettes smoked per day is the natural measure of the magnitude of smoking. This is generally categorized as (1-9), (10-19), (20-29), etc., for the purpose of reporting of the results. Some investigators divide it arbitrarily into light and heavy. For example, Baird and Wilcox [13] considered 20 or less cigarettes per day as light, and 21 or more heavy for studying fertility in women in the U.S. Petrauskaite et al. [15] used a cut-off of 7 cigarettes per day, and had categories 8-15 and 16+ also, although they did not call them as light, moderate and heavy. Their subjects were Lithuanian men residing near an industrial site, and the disease under investigation was lung cancer. Thus, variation exists in such categorization. As a side advantage, we hope to reduce, if not eliminate, such subjectivity by developing an index of smoking.

The next most commonly studied dimension of smoking is the duration of smoking. Ji et al. [16] divided it into (½-19), (20-29), (30-39) and (40+) years for the purpose of studying stomach carcinoma in China, whereas Hsing et al. [17] used 5-year categories beginning <25 years for prostate cancer in the U.S. veterans. Variation exists in this categorization also.

The cigarettes smoked per day and the duration of smoking are often combined into pack-years. The person-time is an epidemiologic tool that is commonly adopted when the duration of exposure varies from person to person. This however assumes that 'dose' of smoking one cigarette per day for 10 years is the same as the dose of 10 cigarettes per day for one year. This does not necessarily hold true for the hazard of smoking-related diseases. For example, it is known for lung cancer that duration is much more important than the number smoked per day [18]. But this assumption is implicit in all those studies that measure dose of smoking in terms of pack-years. MedLine search in Oct 2007 reveals 728 citations that used the term pack-years for cigarettes. Despite its drawbacks, pack-year is just about the most commonly used measure of the dose of smoking. For example, this measure has been used by Uchimoto et al. [19] for Type-2 diabetes mellitus in Japan and by Hellenbrand et al. [20] for Parkinson's disease in Germany. It is a simple and easy-to-comprehend measure. We use this as one of the important components in our proposed index also in a slightly modified form.

In our opinion, it is more convenient to measure cigarette-years than pack-years. First, history is predominantly taken in terms of cigarettes smoked per day rather than packs per day. Second, the size of pack may differ from country to country, time to time and brand to brand. There is a general agreement that a pack of cigarettes is defined to have 20 cigarettes but this can be overlooked by some researchers. Also, cigarette-years is not altogether in disuse. For example, Mao et al. [21] used this measure while reporting on unfavourable factors associated with low serum total cholesterol in Japanese population. Cigarette-years obviate the need to round off, for example, 25 cigarettes a day to one pack. The other approximation often done in investigating smoking is in eliciting cigarettes smoked as life-time average over the duration of smoking (see, e.g., Siemiatycki et al. [22]). This could also be termed as 'usual' daily consumption as done by Stucker et al. [23]. If 10 cigarettes a day are smoked for three years and 20 a day for seven years, the respondent may report the average as 15 though the actual average is 17. We therefore prefer to retain the exact number of cigarettes smoked per day, and use the duration for which specific number was smoked to calculate cigarette-years. This defines our starting point for building up the index of smoking in this communication. Our initial measure of burden of smoking is the dose of smoking in terms of cigarette-years:

$$S_1 = \sum n_i x_i ,$$

where x_i is the number of cigarettes smoked for n_i years (i = 1, 2, …, I), and I is the number of segments in life with different smoking pattern. The number I may not be high for most smokers. If same number of cigarettes are smoked through out then $I = 1$.

The cigarette-years S_1 is the cumulative dose for 'active' smokers. This can be adjusted for passive smokers. Passive smoking has been found to be a risk factor for lung cancer in Korea [24] and in India [25]. Olsen [11] found it to be a risk factor for subfecundity in Danish women. Depending upon the extent of exposure to smoke, a proportion p can possibly be estimated that quantifies the smoker's smoke inhaled by a person in the inhalation zone. This would be based on, for example, the cigarettes smoked by spouse when together, or similarly in workplace, and the duration of such exposure. The dose of smoking for a passive smoker is (pS_1), where S_1 now is the cigarette-years for which passive smoker is exposed. We could not locate any reference on the per cent inhaled by passive smokers but expect that the efforts will be made in future once our index catches attention. It has been estimated though that excess risk of lung cancer by passive smoking could be 15% [26] whereas the OR in active heavy smokers can go upto 18.3 [27]. Matsubara et al. [28] reported 96 g less weight on average of infants born to active smoking mothers in Japan compared to non-smoker parents, and 11 g less weight of infants born to non-smoking mother but smoking father. In this case the effect of passive smoking is nearly 11% of that of active smoking. For stroke, Bonita et al. [29] observed for New Zealand residents that OR in those exposed to environmental tobacco smoke is 1.82, and in active smokers is 6.33. In this case, excess risk due to passive smoking is nearly 15% of the excess due to active smoking. In carotid atherosclerosis, mean intimal-medial wall thickness, adjusted for confounders, among middle-aged U.S. residents was found 0.738 mm in passive smokers relative to

0.807 mm in active smokers and 0.706 mm in never smokers [30]. Considering the thickness in never smokers as baseline, this increase amounts to an effect of nearly 32% in passive smoking relative to active smoking. As in the case of all other effects, the effect of passive smoking also varies from disease to disease, and from population to population. Nonetheless, in view of these evidences, we are proposing that passive smoking be provisionally considered as 15% of active smoking till such time that a firm evidence emerges. If 15% burden is assumed, the dose of passive exposure to 10 cigarettes a day for 8 years amounts to the dose of 12 cigarette-years of active smoking. For smoker himself or herself, $p = 1$. The term we use for pS_1 is 'adjusted cigarette-years'.

The constant p can also be used as an adjustment for smoking filter cigarettes. Compared with regular cigarettes, filter cigarettes have been found to have reduced effect on lung cancer histology[31] although not on coronary heart disease [32]. Lee [33] has estimated that the risk of lung cancer by smoking filter cigarettes is about two-thirds of that of regular cigarettes. If three filter cigarettes are considered equivalent in toxicity to two nonfilter cigarettes then $p = 0.67$. In fact, if the history in sufficient detail is available, it is possible to incorporate separate components for passive smoking of regular cigarettes, passive smoking of filter cigarettes, active smoking of regular cigarettes and active smoking of filter cigarettes. In that case, $(pS_1) = (p_a S_a) + (p_b S_b) + (p_c S_c) + (p_d S_d)$, where the four components relate, respectively, to the four types of smoking. For active smoking of regular cigarette, $p_c = 1$. Mathematically, it is more accurate to write it as $\sum p_i n_i x_i$.

Similar equivalence could be conjectured for cigars and pipes. For example, Hellenbrand et al.[20] considered 5 cigars and 2.5 pipes equivalent to a pack of 20 cigarettes each. One *bidi*, so common in India and other South Asian countries, is considered equal to one cigarette [34]. Any other equivalence can be considered. Perhaps similar equivalence can be established for smokeless tobacco also such as snuff, chew tobacco and betel quid. If so, the scope of our index would considerably enlarge.

Small amount of smoking for a short duration is generally considered harmless. West et al. [9] used the limit of five packs over a life time to study the risk of nuclear cataracts in the U.S. fisherman. For the purpose of gastric ulcer in Australia, Mcintosh et al. [35] defined smoker as the one who smoked at least a cigarette a day for at least six months. We also consider the one who has smoked less than one-half of a cigarette-year in the whole life as good as a nonsmoker. Similarly, first one-half of a cigarette-year can be excluded for those who smoke more. Thus our first index is

$$S_2 = \begin{cases} \sum p_i n_i x_i - 0.5, & \sum p_i n_i x_i \geq 0.5, \\ 0 & \text{otherwise.} \end{cases}$$

The cumulative cigarette-years, as modified in S_2, may range from zero to 1000 or higher. We postulated in Working Hypothesis-I that the cumulative burden of smoking is not linear but is severe in the beginning that tends to slow down as the dose increases. Field et al. [4] found no significant rise in odds ratio (OR) for breast cancer in the U.S. as the pack-years increased from one to more than 40. But Uchimoto et al. [19] found Type-2 diabetes mellitus related to pack-years in a dose dependent manner in a middle-aged Japanese population. Whether this relationship is linear or curvilinear is a moot question. Our review suggests that curvilinear relation is more commonly seen than the linear. For example, OR for stomach cancer in Chinese men [16] showed a rise from 1.05 for 10-19 pack-years to only 1.68 for 40+ pack years. In Japanese self-defense male officials [38] the OR for signoid colon carcinomas were 1.0, 2.1, 2.8 and 3.5 respectively for 0, 1-399, 400-499 and 500+ cigarette-years respectively. Thus it is not linear. Moderation generally applied for reduced rate of increase for advancing dose is the logarithm. But that would be too severe in this case as it would moderate 100 cigarette-years to two and 1000 cigarette-years to only three. We searched for a simple function of S_2 that may take a value of nearly 5 for 100 cigarette-years and nearly 15 for 1000 cigarette years. The values 5 and 15 are our subjective assessment of the years that must elapse after quitting for the burden of such smoking to disappear in many cases. Further details are given later. The function given below has this feature.

$$S_3 = ½ \sqrt{S_2}$$

$$= ½\sqrt{\sum p_i n_i x_i \ - 0.5}\,, \ \sum p_i n_i x_i \geq 0.5.$$

The minimum value of S_3 is zero and the maximum for extremely heavy smokers can go upto 25 when 50 cigarettes are smoked every day for 50 years!

Next, we incorporate an adjustment for the age at which smoking commenced. Hsing et al. [17] stated results for prostate cancer in the U.S. veterans beginning from less than 15 years for age at start of smoking and ending up with age 25 or more years. Ji et al. [16] started at less than 20 years of age and finished with 30+ in their results for studying stomach cancer in China. Although Vineis et al. [37] reported that the relative risk of bladder cancer in Italian subjects did not change with age at starting to smoke but there is otherwise overwhelming evidence that smoking exposure in early age is a risk factor. First, smoking in early age predicts longer duration of smoking, heavier daily consumption, increased nicotine dependence and less chances of quitting [38]. All these would be in our model any way. Age at start has relevance for our index only if it is an independent risk factor. Hirao et al. [39] observed increased prevalence of loss of heterozygosity on chromosome 3 at 3p21 with earlier age of smoking initiation among squamous cell carcinoma cases in the U.S. In former smokers in the U.S. Wiencke et al. [40] reported that age at smoking initiation was inversely associated with DNA adduct levels. A strong evidence comes from the study by Hegmann et al. [41] who found that men who began to smoke before age 20 had a substantially higher risk of developing lung cancer compared with those who started late, and this is after controlling for age, sex and amount of tobacco exposure. Larsson et al. [42] concluded for Swedish subjects that childhood exposure to environmental tobacco smoke is associated with an increased prevalence of asthma among adult never-smokers, especially in nonatopic subjects.

We assume for the purpose of our index that the net burden of early start regresses with age if start is before the age 30 years, and remains at the same level thereafter. The figure in Giovino et al. [43] indicates that nearly all smokers in the U.S. start smoking by the age 30 years. The mean and SD provided by Lando et al. [44] for a population of working adults in the U.S. show, under Gaussian assumption, that less than 2% start smoking after the age of 27 years. Situation in other countries could be different but it seems safe to assume that very few start smoking after the age of 30 years. Passive smoking can start in childhood through parents, even as a fetus. We won't be able to incorporate fetal smoking but arbitrarily assume that the burden of smoking, when started in infant period, is three times the burden when started at age 30 years or later, and the relationship is linear (Working Hypothesis–II). We explain this hypothesis in the next paragraph. This is the net effect of age at start on the burden of smoking independent of dose of smoking. This adjustment would mean that the start of smoking at age 15 years is twice as much a burden as starting at age 30 years. This has special significance for passive smoking since that may indeed start in childhood [45]. This assumption means, for example, that it would take 27 cigarette-years beginning at age 40 years to cause same burden as only 10 cigarette-years beginning at age 5 years. Though age also advances with duration but we assume that age *per se* remains an independent risk factor such as obesity is for coronary diseases and diet pattern is for some cancers.

Indeed not much evidence is available in support of our assumption regarding the exact magnitude of effect of early smoking postulated in Working Hypothesis–II. Hegmann et al. [41] reported nearly twice as much risk of developing lung cancer in the U.S. men when smoking started before the age of 20 years compared with those started at age 20 or older. Other diseases may have different pattern. We assume twice burden when age at start is 15 years relative to start at age 30 years or older. Absence of empirical evidence causes epistemic uncertainty that in our opinion can be filled up for the time being with some thing that otherwise looks rational. Working Hypothesis-II is an expression of this approach. Future evidence may confirm or refute this assumption. In the case of the latter, the exact form of the index can be changed, although broad framework can still remain the same. With this, our index now becomes

$$S_4 = \begin{cases} (3 - a/15)S_3 & \text{for } a < 30 \\ S_3 & \text{for } a \geq 30, \end{cases}$$

where a is the age in years at start of smoking.

After accumulating all the negative aspects of smoking, our final adjustment is for the benefit of quitting smoking. While residual burden of smoking can remain for many years after cessation but it may decline as the duration elapsed increases. Stubbe et al. [46] reported increase in high density lipoprotein concentrations within six weeks of stopping by heavy smokers in Sweden. This may not be so dramatic for risk of diseases. Peto et al. [47] in a recent paper concluded that the former smokers had only a fraction of the lung cancer rate of continuing smokers, and this fraction fall steeply with time since stopping. In case of heart diseases, the risk reduces rapidly immediately after cessation and at a slower rate thereafter [48]. van Domburg et al. [49] estimated that benefit of survival for quitters in cases of coronary artery bypass surgery increased from 3% at five years to 14% at 15 years. The review of Critchley and Capewell [50] observed 36% reduction in risk of mortality in patients with coronary heart disease after quitting smoking although they were not able to assess how quickly does it happen. The mean length of follow-up in the studies reviewed by them was between 3 and 7 years. Speizer et al. [51] observed that the risk of lung cancer in middle-aged U.S. women rapidly decreased with discontinuation of smoking but took 15 years to fall to about the same level as risk for women who never smoked. Bueno de Mesquita [5] reported similar finding for exocrine pancreatic cancer. The pattern is different for different diseases. But these benefits partially explain why we earlier searched for a function that becomes zero if the duration elapsed is 15 years or more in case of heavy smoking and 5 years in case of moderate smoking. Our Working Hypothesis-III is that the burden would gradually reverse linearly as the duration elapsed since quitting increases until it settles down to no burden. It would disappear after a very long time if the accumulated burden of smoking is heavy. With this adjustment, the index finally becomes

$$S = \begin{cases} S_4 - y, & \text{for } y < S_4 \\ 0, & \text{for } y \geq S_4, \end{cases}$$

where y is the number of years elapsed since cessation. In terms of original data on smoking, this can be written as

$$S = \begin{cases} (3 - a/15)\ ½\sqrt{\sum p_i n_i x_i - 0.5} - y; \text{ (use } a = 30 \text{ for } a \geq 30), & \sum p_i n_i x_i \geq 0.5,\ y < S_4 \\ 0 & \text{otherwise} \end{cases}$$

where a = age in years at start of smoking,

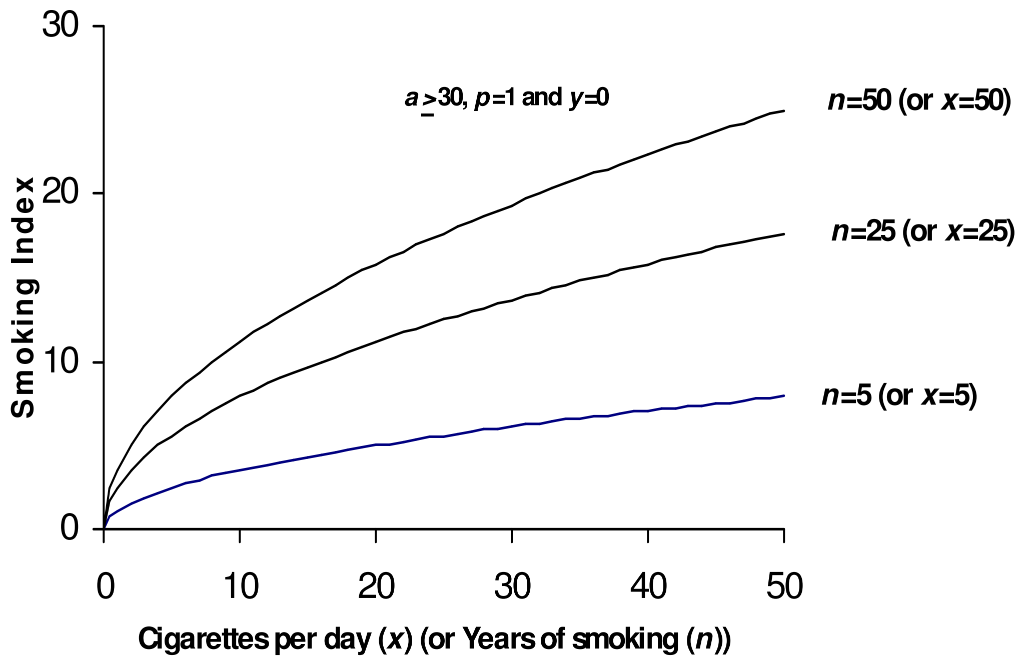


Figure 2 Trend of smoking index on cigarettes per day or duration of smoking (see text for notations)

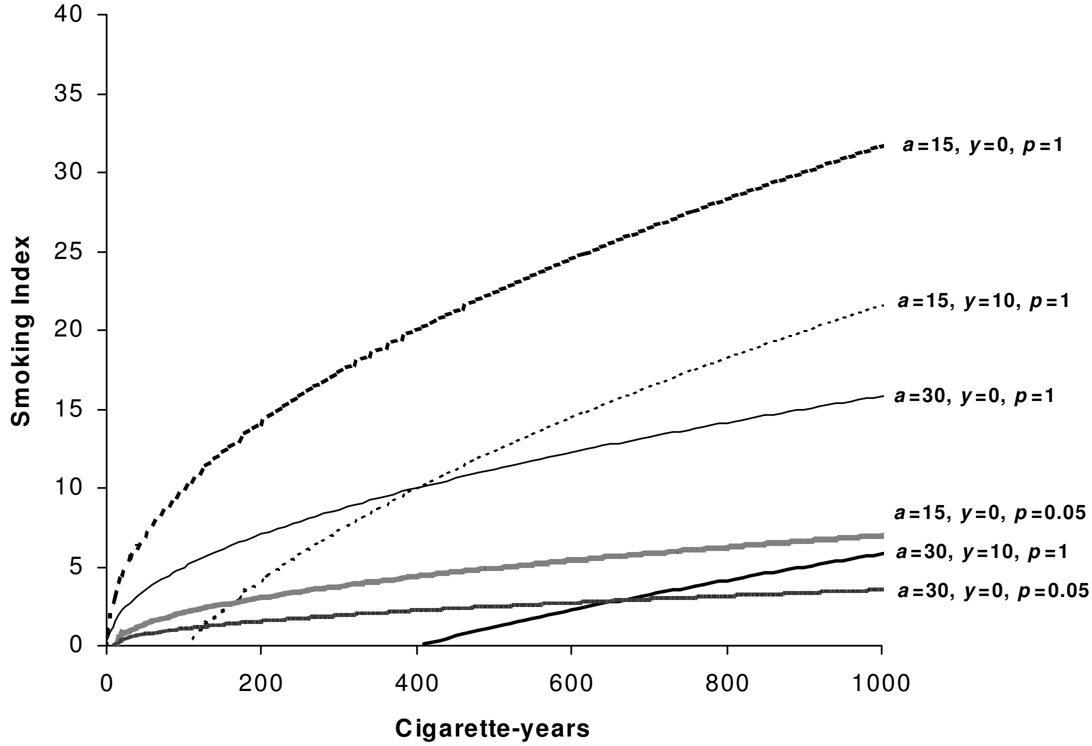


Figure 3. Trend of smoking index for different values of various smoking parameters (see text for notations)

p_i = proportion of smoke inhaled by passive smokers, or regular cigarette equivalent to one filter cigarette when filter cigarettes are smoked (or similar equivalence for other forms of smoking),

x_i = number of cigarettes smoked (or exposure in case of passive smoking) for n_i years, (i = 1, 2, …, I; where I is the number of segments with smoking of different numbers or type of cigarettes),

and y = years elapsed since stopped by ex-smokers.

According to this index, if the smoking started at age 30 years and 400 cigarette-years are smoked then it would take nearly 10 years for the burden to vanish. But if the same smoking is done starting at age 15 years, it would take nearly 20 years. Thus, this index gives large weight to smoking early in life, and is based on the assumption that the burden by smoking in childhood is much more than in the adulthood. Note that our restrictive conditions rule out the possibility of a negative value of the index. The minimum value is zero.

Figures 2 illustrates the trend in this index with increasing duration of smoking or increasing cigarettes smoked per day. Both get the same weightage in line with the concept of pack-years. Trends of the index for different values of various parameters are shown in Figure 3. Note the trend of smoking index as the cigarette-years increase, and see how this declines for ex-smokers and increases for those who start smoking at early age.

Computational example

Consider a male of age 46 years who quitted smoking 3 years ago following a coronary attack. He himself started smoking at the age of 18 years but was exposed to passive smoking throughout childhood. He estimates that his father was smoking nearly 6 cigarettes a day during the period when he and his father were together. He himself has smoked nearly 10 regular cigarettes a day for 3 years, 15 filter-cigarettes a day for the next 7 years, and 6 cigars a day for the next 15 years.

Assumptions -1. One filter cigarette = 2/3 regular cigarette

2. One cigar = 4 cigarettes

3. Passive smoking = 15% of active smoking

For active smoking ($p_i = 1$), $\sum n_i x_i$ = 10×3 + 2/3×15×7 + 4×6×15 = 460;

For passive smoking till the age of 18 years, $\sum p_i n_i x_i$ = 0.15×6×18 = 16.2;

Age at start, a = 18;

Years since quitting, y = 3.

Thus, smoking index for this person is

$$S = (3 - 18/15)\ \tfrac{1}{2}\sqrt{460 + 16.2 - 0.5} - 3$$

$$= 16.6.$$

Discussion

We have made a series of assumptions to keep the index simple. These can be changed as shown in the Appendix. In our opinion, any change in these assumption would be as arbitrary as the ones now proposed till such time that evidence in support or to the contrary emerges. The index actually is a metric of our assessment of the burden on any person at any point of time in life by smoking or exposure to smoking. But it is easy to call it an index of smoking.

The index suffers from the same demerits that almost all other composite indices do. It is likely to be good as a comprehensive metric for studying joint effect of different dimensions of smoking but is not appropriate if these dimensions are needed to be studied separately for etiologic reasons. For example, our index will not be very efficient in finding that start of smoking early in life is more of a burden or the large dose of smoking. Also, this index fails to segregate the effect of duration of smoking from the effect of the number of cigarettes. The effect of such individual dimensions of smoking in a multivariate setup can be studied although this is rarely done. Another problem with our index is its inability to take care of occasional smoking. While the volume of smoking can be calculated in terms of cigarette-years in this case also but the possible moderation in burden due to intermittent cessation [52] can not be studied by this index. Conventionally also this aspect is generally ignored.

The following advantages of the proposed index can be listed. These also are same as for almost any other index.

1. It comprehensively measures several dimensions of smoking, and can be used to study the effect of smoking in an integrated manner. Thus there is no need to study ex-smokers separately from current smokers, and passive smokers separately from active smokers, unless needed for etiologic studies. It can also include smoking of bidis, pipes and cigars that quite often do not get attention. The index combines the qualitative and quantitative dimensions into a single

metric, and allows for a more direct comparison of the burden of smoking across various categories of tobacco use.

2. The index can be used as a regular quantitative measure and it is not necessary to divide it into categories the way number of cigarettes and duration of smoking are almost invariably done. The subjectivity in categorization is entirely eliminated when the exact values of the index are used for analysis. When needed for a particular problem, categories such as mild, moderate and heavy burden of smoking can be formed by arbitrary divisions of the index values. We suggest 5 and 10 as the cut-points. This subjectivity would be much less than the sum total of separate subjectivities presently occurring in categorizing the number of cigarette per day, duration of smoking, age at start, and time elapsed since quitting by ex-smokers.

No index can work equally well for all diseases. A simple index such as BMI = wt/ht^2 correlates well with risk of angina [53] but poorly with risk of lung cancer in older women [54]. Yet it is considered a valid measure of obesity. The index of activities of daily living [55]. is supposed to be a valid measure but it may not correlate well with, say, tumor stage in a patient. Our index is not disease specific just as is pack-years. It is extremely unlikely that any index of this nature would correlate well with all smoking related health conditions. It should give excellent correlation with the net risk (due to smoking) of a disease that fulfills the conditions stated earlier. In situations where not, and these are far more in number, the index will have less correlation despite smoking being a known risk factor. This does not diminish the utility of the index but only indicates that the disease is not affected by smoking the way we postulate as burden. The index may encourage generation of better data on different aspects of smoking that are sometimes ignored at present.

This index is not the last word on this topic but this is the most comprehensive attempt made so far. It contains most but not all the information contained in different dimensions of smoking. Using all the various dimensions of smoking simultaneously would always be superior, although this approach is difficult to adopt. The index may have to be developed further once the assumptions are verified, and the difficulties in adopting the index are spelt out by the users.

The suggested proforma for eliciting smoking history required for this index is available with the first author (AI). This proforma is designed in easy-to-fill manner and to calculate the index for any present or past smoker. We realise that life long history of smoking may not be fully precise — there is always a chance of forgetting or of approximation. But that applies to some extent to pack-years also, which is now in common use. We expect that the history would be reasonably accurate to give fairly precise value of the index. Given the importance of smoking in many diseases, it may be worthwhile to spend extra time and effort to elicit accurate and complete history. Once this is available, the index can be easily calculated. The index has side advantages in terms of its potential to encourage further work. For example, the investigators would be encouraged to find how many filter cigarettes are required to produced same burden as 100 nonfilter (regular) cigarettes; what percentage of smoke is inhaled in case of passive smoking; how early start of smoking affects the course of the disease, etc. Once these questions are precisely answered the index can be accordingly modified.

Validation example

Our Department of Medicine runs a Preventive Cardiology Clinic twice a week in the afternoon. Besides present and past known cases of coronary artery diseases (CAD), many subjects who apprehend or anticipate heart problem of any type come to this clinic for advice. The subjects reporting in this clinic are routinely assessed for different risk factors such as obesity, smoking, alcohol intake and diet. A patient with stable or unstable angina, or with present or past history of myocardial infarction is labeled as a case of CAD.

Although smoking history has always been part of the assessment but after development of this index, smoking history is taken on a more elaborate format so that all aspects as required to compute the index could be covered. This exercise started in March 2007 and is still going on. The data we now use to illustrate the smoking index belongs to all the new consecutive subjects that visited this clinic from March 2007 to October 2007. A total of 254 subjects visited the clinic during this 8-month period. Smoking history was incomplete in 2 cases. Thus the present analysis is based on the information on 252 subjects. The age of the subjects ranges from 21 years to 90 years with heavy concentration in forties and fifties (40.1%). The percentage of females is 24.2. Nearly 16% were *bidi* smokers. This is roll of a plant leaf with tobacco inside and is a low-cost local variant of cigarette. Distribution of subjects by different smoking parameters is given in Table 1. Univariate odds ratios (ORs) are also mentioned in this table. These clearly indicate that odds of CAD increase with smoking and its different dimensions among attendees of this clinic.

Table 1. Smoking history of subjects without and with CAD among preventive cardiology clinic attendees

Particulars	Number of subjects		Univariate OR (ref: Never smoker)
	Without CAD	With CAD	
Total	140 (100.0)	112(100.0)	
Smoking status			
Never smoker	85 (59.9)	31 (27.7)	1.00
Current smoker	45 (31.7)	71 (63.3)	4.33
Past smoker	10 (7.0)	10 (8.9)	2.74
Ever smokers (n=136)	55 (38.7)	81(72.3)	4.04
Age at start (yrs)			
<20 years	5 (3.5)	17 (15.2)	9.32
20-30 years	26 (18.3)	44 (39.3)	4.64
30+ years	24 (16.9)	20 (17.9)	2.28
Type of smoking			
Cigarette-filter	6 (4.2)	6 (5.4)	2.74
Cigarette-nonfilter	37 (26.1)	46 (41.1)	3.41
Bidi	12 (8.5)	29 (25.9)	6.63
Cigarettes per day (average)			
<10	10 (7.0)	15 (13.4)	4.11
10-20	19 (13.4)	35 (31.3)	5.05
≥20	26 (18.3)	31 (27.7)	3.27

Duration of smoking (years)			
<10	24 (16.9)	4 (3.6)	0.46
10-20	17 (12.0)	30 (26.8)	4.84
≥20	14 (9.9)	47 (42.0)	9.21
Years since stopped			
0 (current smoker)	45 (31.7)	71 (63.4)	4.33
<5	5 (3.5)	8 (7.1)	4.39
≥5	5 (3.5)	2 (1.8)	1.10

In parentheses are percentage

Among many that could have been tried, we present results for three types of logistic model. These are labeled as Logistic-1, Logistic-2 and Logistic-3 in Table 2. The first is the conventional setup with age, duration-of-smoking, cigarettes-per-day, age-at-start and type-of -smoking (never smoker, filter cigarette, nonfilter cigarette and bidi) as the independent variables. All these variables except the last are considered continuous in each of the three models. Categories such as in Table 1 are arbitrary and can introduce bias. Therefore we decided against using any categories. The variables were entered in a sequential manner so that their utility in the presence of the previously entered variables can be evaluated. Analysis for partial contribution, which is the contribution in the presence of all the other variables, is also presented in Table 2 for all the models. Age, which is an independent risk factor, was entered first in each model. In Logistic-1, duration-of-smoking and cigarettes-per-day were entered together so that these can be compared with the conventional 'cigarette-years' that we study in Logistic-2. Next, we entered age-at-start and type-of-smoking. The other variable that we studied is 'years-since-stopped' but this could not be entered in Logistic-1 because of redundancy—years-since-stopped is a linear combination of age, duration-of-smoking and age-at-start.

Table 2. Logistic regression results (all factors continuous except type of smoking) Smokers + Non smokers (n=252)

Factors in the model	Model χ^2	Model *P*-value	Sequential OR	Sequential *P*-value	Percent correctly classified	Partial OR	Partial *P*-value
None	346.23						
Logistic-1							
Age	29.39	0.0000	1.0520	0.0000	61.11	0.5969	0.0007
Duration of smoking			1.0704	0.0001		2.0776	0.0000
+	67.81	0.0000			75.79		
Cig./day			1.0689	0.0005		1.0952	0.0009
Age at start			1.8509	0.0001			
+	31.96	0.0000			81.75		
Type of smoking						Same as for sequential	
Never smoker (ref.)			1.0000				
- Cig.– filter			0.4022	0.4615			
- Cig.– nonfilter			0.6720	0.6563			
- Bidi			0.9270	0.9407			
Logistic-2							
Age	29.39	0.0000	1.0520	0.0000	61.11	1.1342	0.0004
cig. years (S_1)	67.66	0.0000	1.0078	0.0000	77.78	1.008	0.0003
Age at start			0.9727	0.4556			
+							
Years since stopped			0.5019	0.0001			
+	36.37	0.0000			82.54	Same as for sequential	
Type of smoking			—	0.1538			
Never smoker (ref.)			1.0000	—			
- Cig.– filter			1.1646	0.8949			
- Cig.– nonfilter			2.1008	0.3318			
- Bidi			2.9842	0.2626			
Logistic-3							
Age	29.39	0.0000	1.0520	0.0000	61.11	1.1189	0.0010
Smoking index (S)	102.80	0.0000	1.4132	0.0000	83.73	1.5535	0.0006
Cig. years (S_1)	6.14	0.0132	0.9943	0.0134	83.73	0.9983	0.5737
Age at start			1.0125	0.7432			
+							
Years since stopped			0.7007	0.0674			
+	7.34	0.1939			83.73	Same as for sequential	
Type of smoking			—	0.5045			
Never smoker (ref.)			1.0000	—			
- Cig. – filter			0.9142	0.9386			
- Cig.– nonfilter			0.9462	0.9489			
- Bidi			1.6302	0.6330			

Similar sequential entering of variables was done in Logistic-2 and Logistic-3 as well. In Logistic-2, duration-of-smoking and cigarettes-per-day were replaced by cigarette-years. In this model, we could enter 'years-since-stopped' as an independent variable. In Logistic-3, we first entered our smoking index *S* after age, then cigarette-years, and then age-at-start, years-since-stopped and type-of-smoking. Those entered together are shown bracketed on left in Table 2. The corresponding odds-ratios are also shown.

While interpreting these results, keep in mind that CAD is affected by a large number of other factors (obesity, exercise, stress, diet, etc.) that have not been included in our models. Our purpose is to examine the utility of smoking index *S* relative to the conventional smoking indicators, particularly cigarette-years.

In this series of subjects, the results of Logistic-1 show that all the smoking variables have statistically significant ($P<0.05$) association with CAD except categories of type-of-smoking. Of a total chi-square of 346.23, the contribution of duration-of-smoking together with cigarettes-per-day is very high (67.81) relative to age (29.39) and age-at-start + type-of-smoking (31.96). Logistic model consisting of only age is able to correctly classify 61.11% subjects into CAD and non-CAD groups. This increased to 75.79% when duration-of-smoking and cigarettes-per-day are included in the model. Inclusion of age-at-start and type-of-smoking increased it further to 81.75%. Less than one ORs for filter and non-filter cigarettes compared to never smoker is worrying but they are statistically not significant.

Now compare these with the results obtained when duration-of-smoking and cigarettes-per-day are replaced by cigarette-years. These are given under Logistic-2 in Table 2. The chi-square contribution of this variable is 67.66 compared to 67.81 of duration-of-smoking + cigarettes-per-day. The percent correctly classified are also nearly the same —77.78% compared to 75.79% in the first model. Thus, cigarette-years has nearly the same efficiency in explaining CAD as jointly by duration-of-smoking and cigarette-per-day. Age-at-start, years-since-stopped and type-of-smoking together remains statistically significant (P<0.0001) even after cigarette-years and age are in the model. Their contribution is not small.

Finally we enter our smoking index S after age in Logistic-3 model. Several important things occur when this is done. First, age-at-start, years-since-stopped and type-of-smoking loose statistical significance (P=0.1939) when S and cigarette-years are already in the model. This could indicate that these three aspects are adequately taken care of in the index S. Note that cigarette-years does not contain these variables. Secondly the contribution of S to chi-square is 102.80, which is much more than 67.66 of cigarette-years in Logistic-2. The per cent correctly classified also increases to 83.73 from 77.78 in Logistic-2 and 75.79 in Logistic-1 at this stage. This might mean that S could be better in explaining CAD odds than cigarette-years or duration-of-smoking + cigarettes-per-day. What is worrying however is that cigarette-years is still statistically significant (P=0.0132) after S is in the model. This could mean that information in cigarette-years is not adequately accounted for in our index S. Nevertheless, note from Logistic-3 results that the percentage correctly classified does not improve any more by other parameters once smoking index is the model.

As far as this series of subject goes, it is clear that the index S incorporates most of the information contained in different parameters of smoking though not all. Considering all the dimensions of smoking in multivariate set-up is still better although this is rarely done. We certainly do not claim S to be the final answer. It however seems to deserve attention and further trials for improvement.

Application example

As an application for inferential purposes on relation of CAD with smoking, note from Figure 4 how the percentage with CAD increases with the value of the smoking index. Age is ignored in this figure. Discriminant function analysis shows that smoking index and age are able to correctly classify 80.1% smokers into CAD and non-CAD categories (Figure 5). This figure is truncated at S=25, and 7 cases are excluded. Both indicate for this series of subjects that CAD is substantially associated with smoking index. If OR is interpreted as relative risk, which can be done for rare diseases, we find from logistic model that a unit increase in smoking index is associated with an increased risk of CAD by a factor of 1.41 among the clinic attendees when age is already in the model. Together, age and smoking index are able to contribute nearly 38% of the Chi-square in these subjects.

A brief of this index is described by Indrayan[56].

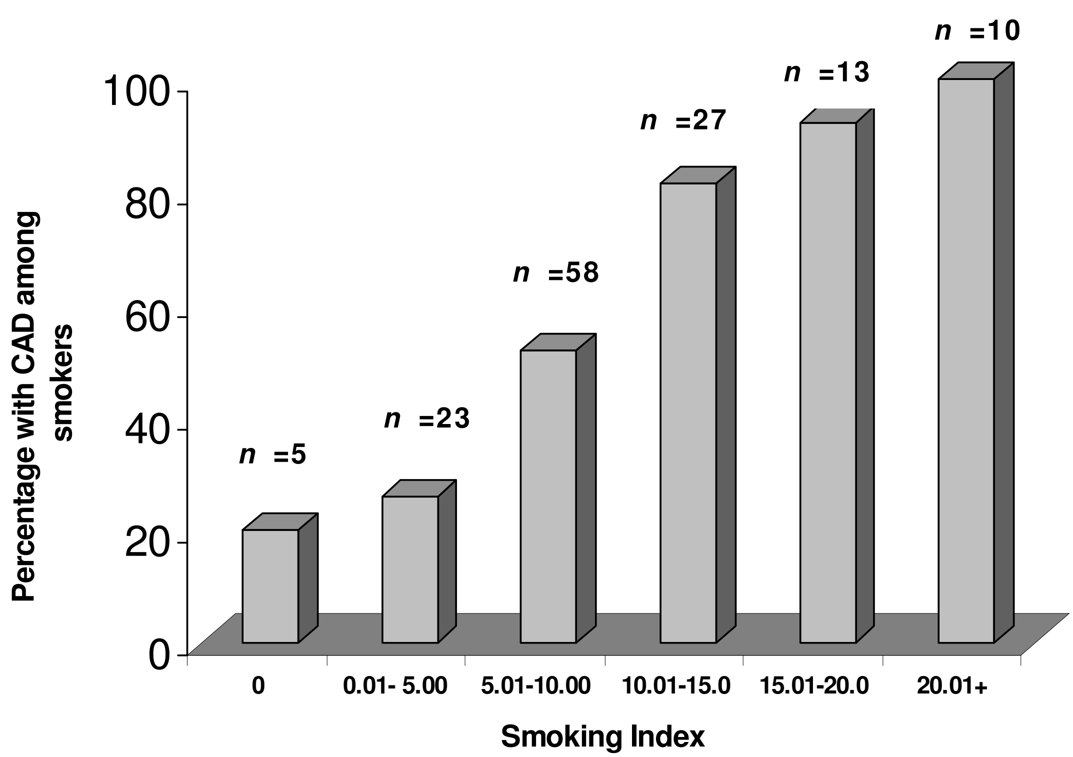
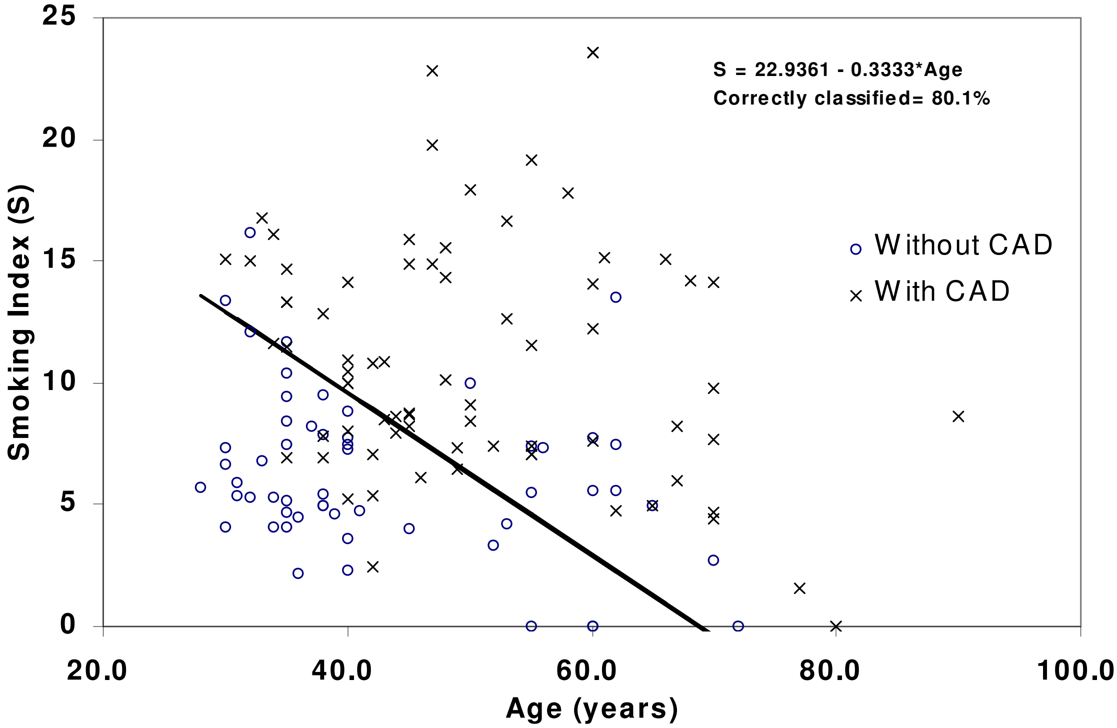


Figure 4 Relation of smoking index with CAD among preventive cardiology clinic



attendees

Figure 5 Best discriminating line for CAD among smokers attending preventive cardiology clinic

Appendix

One general form of smoking index that can get rid of some assumptions is the following:

$$S = f(a)\, g(pS_1) - h(y).$$

The function $f(a)$ delineates the role of age at start of smoking, which is (3-a/15) in our index. The function $g(pS_1)$ measures the burden due to dose of smoking. We assume it to be $½\sqrt{\sum p_i n_i x_i} - 0.5$ for the purpose of our index. The function $h(y)$ is for relief in burden by the duration elapsed since quitting by ex-smokers. We took it to be simply y itself. Various other forms of these functions can be suggested. In place of linearity assumed in Working Hypothesis-II and Working Hypothesis-III, an exponential or parabolic decline can be considered. But that would make the index too complex for adoption.

Key Messages

- Since smoking is predominant risk factor for many diseases and deaths, there is a need to measure it more comprehensively than done so far.
- We propose an index of smoking that combines the effect of age at start, amount of smoking, duration of smoking, passive smoking, filter smoking, and duration elapsed since quitting, into a single metric.
- The application of the index is demonstrated, and it is suggested that this be developed further.

REFERENCES

1. Murray CJL, Lopez AD. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Boston: Harvard School of Public Health, 1996: 311.
2. WHO. The World Health Statistics 2007. Geneva: World health Organization 2007: p12.
3. Haldorsen T, Grimsrud TK. Cohort analysis of cigarette smoking and lung cancer incidence among Norwegian women. Int J Epidemiol 1999; 28:1032-1036.
4. Field NA, Baptiste MS, Nasca PC, Metzger BB. Cigarette smoking and breast cancer. Int J Epidemiol 1992; 21:842-848.
5. Bueno de Mesquita HB, Maisonneuve P, Moerman CJ, Runia S, Boyle P. Life-time history of smoking and exocrine carcinoma of the pancreas: a population-based case-control study in The Netherlands. Int J Cancer 1991; 49:816-822.

6. Prescott E, Hippe M, Schnohr P, Hein HO, Vertbo J. Smoking and risk of myocardial infarction in women and men: longitudinal population study. BMJ 1998; 316:1043-1047.
7. Piper DW, Nasiry R, Mcintosh J, Shy CM, Pierce J, Byth K. Smoking, alcohol, analgesics and chronic duodenal ulcer: a controlled study of habits before first symptoms and before diagnosis. Scand J Gastroenterol 1984; 19:1015-1021.
8. Mikhailidis DP, Papadakis JA, Ganotakis ES. Smoking, diabetes and hyperlipidaemia. J Royal Soc Health 1998; 118:91-93.
9. West S, Munoz B, Emmett EA, Taylor HR. Cigarette smoking and risk of nuclear cataracts. Arch Ophthalmol 1989; 107:1166-1169.
10. Evans HJ, Fletcher J, Torrance M, Hargreave TB. Sperm abnormalities and cigarette smoking. Lancet 1981; i(8221):627-629.
11. Olsen J. Cigarette smoking, tea and coffee drinking, and subfecundity. Am J Epidemiol 1991; 133:734-739.
12. Morabia A, Wynder EL. Cigarette smoking and lung cancer cell types. Cancer 1991; 68:2074-2078.
13. Baired DD, Wilcox AJ. Cigarette smoking associated with delayed conception. JAMA 1985; 253:2979-2983.
14. Prescott E, Hippe M, Schnohe P, Ole H, Vestbo J. Smoking and risk of myocardial infarction in women and men: longitudinal population study. BMJ 1998;316:1043-1047.
15. Petrauskaite R, Pershagen G, Gurevicius R. Lung cancer near an industrial site in Lithuania with major emissions of airway irritants. Int J Cancer 2002; 99:106-111.

16. Ji B, Chow WH, Yang G, et al. The influence of cigarette smoking, alcohol, and green tea consumption on the risk of carcinoma of the cardia and distal stomach in Shanghai, China. Cancer 1996; 77:2449-2457.

17. Hsing AW, McLaughlin JK, Hrubec Z, Blot WJ, Fraumeni JF Jr. Tobacco use and prostate cancer: 26-year follow-up of US veterans. Am J Epidemiol 1991; 133: 437-441.

18. Doll R, Peto R. Cigarette smoking and bronchial carcinoma: dose and time relationship among regular smokers and lifelong non-smokers. J Epidemiol Community Health 1978; 32:303-313.

19. Uchimoto S, Tsumura K, Hayashi T, et al. Impact of cigarette smoking on the incidence of Type 2 diabetes mellitus in middle-aged Japanese men: the Osaka Health Survey. Diabet Med 1999; 16:951-955.

20. Hellenbrand W, Seidler A, Robra B, et al. Smoking and Parkinson's disease: a case-control study in Germany. Int J Epidemiol 1997; 26:328-339.

21. Mao X, Okamura T, Choudhury SR, et al. What unfavorable factors are associated with low serum total cholesterol in a Japanese population? J Epidemiol 2002; 12:271-279.

22. Siemiatycki J, Krewski D, Franco E, Kaiserman M. Associations between cigarette smoking and each of 21 types of cancer: a multi-site case-control study. Int J Epidemiol 1995; 24:504-514.

23. Stucker I, de Waziers I, Cenee S, et al. GSTM1, smoking and lung cancer: a case-control study. Int J Epidemiol 1999; 28:829-835.

24. Jee SH, Ohrr H, Kim IS. Effects of husbands' smoking on the incidence of lung cancer in Korean women. Int J Epidemiol 1999; 28: 824-828.

25. Rapiti E, Jindal SK, Gupta D, Boffetta P. Passive smoker and lung cancer in Chandigarh, India. Lung Cancer 1999; 23:183-189.

26. Copas JB, Shi JQ, Reanalysis of epidemiological evidence on lung cancer and passive smoking. BMJ 2000; 320;417-418.

27. Khuder SA. Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. Lung Cancer 2001; 31:139-148.

28. Matsubara F, Kida M, Tamakoshi A, Wakai K, Kawamura T, Ohno Y. Maternal active and passive smoking and fetal growth: a prospective study in Nagoya, Japan. J Epidemiol 2002; 10:335-343.

29. Bonita R, Duncan J, Truelsen T, Jackson RT, Beaglehole R. Passive smoking as well as active smoking increases the risk of acute stroke. Tob Control 1999; 8:156-160.

30. Diez-Roux AV, Nieto FJ, Comstock GW, Howard G, Szklo M. The relationship of active and passive smoking to carotid atherosclerosis 12-14 years later. Prev Med 1995; 24:48-55.

31. Stellman SD, Muscat JE, Hoffmann D, Wynder EL. Impact of filter cigarette smoking on lung cancer histology. Prev Med 1997; 26:451-456.

32. Castelli WP, Garrison RJ, Dawber TR, McNamara PM, Feinleib M, Kannel WB. The filter cigarette and coronary heart disease: the Framingham study. Lancet 1981; 2(8238):109-113.

33. Lee PN. Lung cancer and type of cigarette smoked. Inhal Toxicol 2001;13:951-976.

34. WHO. Tobacco or health: a global status report. Geneva: World Health Organization 1997: 20.

35. Mcintosh JH, Byth K, Piper DW. Environmental factors in aetiology of chronic gastric ulcer: a case-control study of exposure variables before the first symptoms. Gut 1985; 26:789-798.

36. Todorki I, Kono S, Shinchi K, et al. Relationship of cigarette smoking, alcohol use, and dietary habits with sigmoid colon adenomas. Ann Epidemiol 1995; 5:478-483.

37. Vineis P, Esteve J, Terracini B. Bladder cancer and smoking in males: types of cigarettes, age at start, effect of stopping and interaction with occupation. Int J Cancer 1984; 34;165-170.

38. Breslau N, Peterson EL. Smoking cessation in young adults: age at initiation of cigarette smoking and other suspected influences. Am J Public Health 1996; 86:214-220.

39. Hirao T, Nelson HH, Ashok TD, et al. Tobacco smoke-induced DNA damage and an early age of smoking initiation induce chromosome loss at 3p21 in lung cancer. Cancer Res 2001; 61:612-615.

40. Wiencke JK, Thurston SW, Kelsey KT, et al. Early age at smoking initiation and tobacco carcinogen DNA damage in the lung. J Natl Cancer Inst 1999; 91:614-619

41. Hegmann KT, Fraser AM, Keaney RP, et al. The effect of age at smoking initiation on lung cancer risk. Epidemiology 1993; 4:444-448.

42. Larsson ML, Frisk M, Hallstrom J, Kiviloog J, Lundback B. Environmental tobacco smoke exposure during childhood is associated with increased prevalence of asthma in adults. Chest 2001; 120:711-717.

43. Giovino GA, Henningfield JE, Tomar SL, Escobedo LG, Slade J. Epidemiology of tobacco use and dependence.s Epidemiologic Reviews 1995; 17:48-65.

44. Lando HA, Thai DT, Murray DM, et al. Age of initiation, smoking patterns, and risk in a population of working adults. Prev Med 1999;29;590-598.

45. Gergen PJ, Fowler JA, Maurer KR, Davis WW, Overpeck MD. The burden of environmental tobacco smoke exposure on the respiratory health of children 2 months through 5 years of age

in the United States: Third National Health and Nutrition Examination Survey, 1988 to 1994. Pediatrics 1998;101;E8.

46. Stubbe I, Eskilsson J, Nilsson-Ehle P. High-density lipoprotein concentrations increase after stopping smoking. BMJ 1982; 284:1511-1513.

47. Peto R, Darby S, Deo H, Silcocks P, Whitlay E, Doll R. Smoking, smoking cessation and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. BMJ 2000; 321(7257):323-329.

48. Landmark K. [Smoking and coronary heart disease] [Norwegian]. Tidsskr Nor Laegeforen 2001; 121:1710-1712.

49. van Domburg RT, Meeter K, van Berkel DF, Veldkamp RF, van Herwerden LA, Bogers AJ. Smoking cessation reduces mortality after coronary artery bypass surgery: a 20-year follow-up study. J Am Coll Cardiol 2000; 36:878-883.

50. Critchley JA, Capwell S. Mortality risk reduction associated with smoking cessation in patients with coronary heart disease: A systematic review. JAMA 2003; 290:86-97.

51. Speizer FE, Colditz GA, Hunter DJ, Rosner B, Hennekens C. Prospective study of smoking, antioxidant intake, and lung cancer in middle-aged women (USA). Cancer Causes Control 1999; 10:475-482.

52. Lindstrom M, Ostergren PO. Intermittent and daily smokers: two different socioeconomic patterns, and diverging influence of social participation. Tob Control 2001; 10:258-266.

53. Beard CM, Orencia A, Kottke T, Ballard DJ. Body mass index and the initial manifestation of coronary heart disease in women aged 40-59 years. Int J Epidemiol 1992; 21:656-664.

54. Drinkard CR, Sellers TA, Potter JD, et al. Association of body mass index and body fat distribution with risk of lung cancer in older women. Am J Epidemiol 1995; 142:600-607.

55. The WHOQOL Group. What quality of life? World Health Forum 1996; 17:354-356.

56. Indrayan A. Medical Biostatistics, Second Edition. Chapman & Hall/CRC Press, 2008: 228-230.