Public Health Nutrition: 6(8), 765–772

DOI: 10.1079/PHN2003500

Evaluating the potential health gain of the World Health Organization's recommendation concerning vegetable and fruit consumption

Kurt Hoffmann[1],*, Heiner Boeing[1], Jean-Luc Volatier[2] and Wulf Becker[3]
[1]Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbrücke, Arthur-Scheunert-Allee 114–116, D-14558 Bergholz-Rehbrücke, Germany: [2]French Food Safety Agency, Maisons Alfort, France: [3]Swedish National Food Administration, Uppsala, Sweden

Submitted 25 March 2003: Accepted 9 June 2003

Abstract

Objective: The World Health Organization (WHO) recommends a daily intake of at least 400 g of vegetables and fruit. The aim of this paper was to evaluate the public health benefit of meeting this WHO recommendation by applying a statistical method that combines estimated intake distributions and simulated intake changes.
Design and setting: The benefit of an increased consumption of vegetables and fruit was quantified by the preventable proportion of diseases. This proportion was estimated by a general formula derived in the paper that incorporates individual relative risks. Three different strategies of increasing usual intake were simulated and compared. The first strategy assumes that all individuals increase their intake by the same amount, the second assumes a constant increase among low consumers, and the third simulates individual increments necessary to meet the WHO recommendation. Calculations were made for three different scenarios with varying relative risks.
Results: The third simulation strategy turned out to be the most appropriate one to quantify the potential health gain of the current dietary recommendation. Applying this strategy to prevent cancer, the proportion of preventable cases was country-specific. Estimates for France and Sweden were 21.9% and 19.3%, respectively, which are somewhat lower than the non-specific figure published by the World Cancer Research Fund.
Conclusions: To improve estimates of the preventable proportion of diseases, the estimation formula presented here can be applied. Its application requires intake data to estimate the initial intake distribution in the population and to simulate adequate dietary changes.

Keywords
Diet
Recommendation
Vegetables
Fruit
Public health
Prevention
Cancer
Statistics

A diet rich in vegetables and fruit is generally considered to be protective against certain cancers and cardiovascular diseases[1–5]. The World Health Organization (WHO) recommends a daily intake of at least 400 g of vegetables and fruit per person[6]. The public health benefit of the recommendation can be evaluated by estimating the fraction of diseases or deaths that would not have occurred if the dietary recommendation of the WHO had been met in the past.

There are, however, some methodical difficulties in specifying the dietary change necessary to achieve the recommended target value of 400 g day^{-1}. In previous attempts[7–9] all individuals were supposed to increase their usual intake by the same amount to raise the population mean to 400 g day^{-1}. With The Netherlands as reference population, the increment was equated to 160 g day^{-1}. This strategy requires only a simple formula for estimating the proportion of preventable cases if a linear dose–response relationship is assumed across the entire range of intakes. However, the strategy has the serious weakness that some individuals remain low consumers whereas other individuals increase their intake unnecessarily. Actually, the strategy does not aim to achieve the recommended target value. Consequently, it does not simulate the actual effect of the WHO recommendation. Moreover, the simple estimation formula does not allow for the initial intake distribution and yields the same estimates for all countries. For example, it is not realistic to expect that the potential benefit of meeting the WHO recommendation by all individuals is the same in the UK and in Spain, considering that the current mean intake of vegetables and fruit is ~200 g day^{-1} in the UK and ~600 g day^{-1} in Spain[10].

Therefore, two other strategies that focus on low consumers, i.e. individuals who eat less than 400 g of vegetables and fruit per day, were simulated in the present

---

*Corresponding author: Email khoff@mail.dife.de paper. One of these strategies assumes a constant increase in intake, whereas the other simulates the least increment of intake for each low consumer necessary to achieve the target value of 400 g day^{-1}. Applying the latter simulation strategy allows quantification of the benefit that is achievable by the dietary recommendation provided that all low consumers are willing and able to increase their vegetable and fruit consumption. In this sense, the proposed strategy is suitable to evaluate the maximal or potential health gain of the dietary recommendation. Clearly, the potential health gain simulated by this strategy will differ between countries.

To estimate the percentage of cases preventable by quite flexible strategies, a general formula was derived and presented in this paper that utilises relative risks for individuals according to their simulated intake changes. Using this formula we compared the three strategies of increasing vegetables and fruit intake within a simulation study. We further applied the strategies to estimate the preventable proportion of cancer cases in France and Sweden. Here, data from national food consumption surveys were utilised to estimate the initial distribution of the total vegetable and fruit intake in both countries.

Methods

To quantify the effect of a prevention strategy we follow the concept of attributable risk introduced by Levin[11] and generalised by others[12–15]. Thus, we estimate the preventable proportion of diseases that would result if the intake of vegetables and fruit were increased pursuing a specified prevention strategy. In the literature[8,9], the preventable proportion (*PP*) of cases was calculated by the simple formula:

$$PP = 1 - RR,$$

where *RR* denotes the relative risk, which is supposed to be the same for all individuals. In practice, this formula is only valid for the special strategy that all individuals in the population increase their intake by the same amount and assumes that the relative risk associated with the constant increment is the same for all possible initial intakes. Because of these rigid assumptions the initial intake distribution has no importance, and, therefore, the estimated proportion of preventable cases is always the same for all countries and regions. To calculate the proportion of cases preventable by a more flexible prevention strategy, we have to admit individually different changes and, consequently, also individually different relative risks. For this general case, we derived the formula:

$$PP = 1 - \frac{1}{N}\sum_{i=1}^{N} RR_i$$

(see Appendix A). Here, N is the population size and RR_i denotes the relative risk associated with the recommended increase of vegetables and fruit intake for the ith individual. Obviously, the above formula is a straightforward generalisation of the one represented before. The relative risks RR_i will be different, in general. They not only depend on the intake change of the individual but also on the individual initial intake. Thus, the above formula combines information of the underlying intake distribution and the whole spectrum of individual intake changes. Consequently, differences in current usual intake between countries will be reflected by different figures of *PP*. The preventable proportion can be estimated by applying the same formula but estimating the relative risks and substituting the population size (N) by the size (n) of a sample that is representative and not too small.

We compared three different strategies of increasing the total vegetable and fruit intake, which are described in Table 1. The mathematical representations of the prevention strategies, called target value functions[16], are given in the right column. Strategy A, as proposed and explored by other authors[7–9], demands all individuals to increase their daily intake by 160 g, which corresponds to two standard portions. Strategy B is a modified version of A, that allows for the current WHO recommendation of a daily intake of at least 400 g. Here, only individuals who do not meet the recommendation are advised to eat two

Table 1 Disease prevention strategies aimed to increase the intake of vegetables and fruit

Prevention strategy	Target value function
A *Increase by a constant amount* All individuals increase their usual intake by the same amount of 160 g day^{-1}	$\tau(x) = x + c$ $c = 160$ g day^{-1}
B *Equal increase among low consumers* All individuals with usual intake beneath the target value of 400 g day^{-1} increase their intake by the same amount of 160 g day^{-1}	$\tau(x) = x + cI_{\{x<t\}}$ $t = 400$ g day^{-1} $c = 160$ g day^{-1}
C *Increase to a target value* All individuals with usual intake beneath the target value of 400 g day^{-1} increase their intake to the target value	$\tau(x) = \max(t,x)$ $t = 400$ g day^{-1}

$I_{\{x<t\}}$ is the indicator function, which is one if $x < t$ and zero otherwise.

additional standard portions daily. However, successful realisation of strategy B does not ensure that eventually all individuals eat at least 400 g of vegetables and fruit per day. Therefore, a third strategy, C, advising all low consumers to meet the WHO recommendation, was included in our considerations. The preventable proportion of cases calculated for strategy C can be interpreted as the fraction of diseases that would not have occurred if all individuals had met the WHO dietary recommendation in the past.

To explore the effectiveness and benefit of the strategies we carried out a simulation study. We simulated different intake distributions by adopting the shape of the French intake distribution estimated by the Enquête Individuelle et Nationale sur les Consommations Alimentaires (INCA survey)[17] and varying the mean intake. We assumed a linear relationship between the logarithm of relative risk and intake for intakes below a threshold, which can be interpreted as the optimal intake value. The slope chosen corresponds to a relative risk of 0.7 for increasing the intake by 160 g. The calculations were made for two alternative threshold values of 400 g and 500 g. Relative risk estimates for cancer sites based on the World Cancer Research Fund (WCRF)[8] were used.

Results

The effectiveness of the strategies is indicated in Table 2. After applying strategy A or B, a proportion of individuals remained who did not meet the dietary recommendation. The percentage of remaining low consumers increased rapidly as the initial mean intake was decreased from 500 g to 200 g. In a population with a mean vegetable and fruit intake of 200 g day^{-1}, two-thirds of all initial low consumers failed to meet the recommendation after increasing their intake by two additional standard portions. By contrast, no low consumers remained after applying strategy C. Moreover, strategy A affected individuals who already ate 400 g of vegetables and fruit or more daily at the beginning. The proportion of affected high consumers was substantial for high initial mean intake. For example, 70.4% of all individuals in a population with an initial mean intake of 500 g day^{-1} already met the dietary recommendation, but increased their intake by 160 g day^{-1} following strategy A.

Table 2 Effectiveness of different strategies of increasing vegetable and fruit intake: results of a simulation study*

Initial mean intake (g day^{-1})	Remaining low consumers (%)†			Affected high consumers (%)		
	A‡	B	C	A	B	C
200	66.6	66.6	0	11.3	0	0
250	55.4	55.4	0	15.4	0	0
300	41.5	41.5	0	22.1	0	0
350	27.2	27.2	0	31.2	0	0
400	12.8	12.8	0	42.1	0	0
450	0.8	0.8	0	56.3	0	0
500	0	0	0	70.4	0	0

* Intake distributions were simulated by adopting the shape of the French intake distribution obtained in the Enquête Individuelle et Nationale sur les Consommations Alimentaires (INCA survey)[17] and varying the mean intake.

† Low consumers eat less than 400 g of vegetables and fruit daily, all other individuals are called high consumers.

‡ A – all individuals increase their intake by 160 g day^{-1}; B – only low consumers increase their intake by 160 g day^{-1}; C – low consumers increase their intake up to 400 g day^{-1}.

Table 3 gives the estimated proportion of cases preventable by the different strategies (see Appendix B). In the case of a threshold value of 400 g, strategy C was always the one with the highest benefit. Remembering that strategy C actually simulated the full potential health gain of the WHO recommendation, the other two strategies strongly underestimated the proportion of cases that are potentially preventable. For a population with initial mean intake of 200 g day^{-1}, strategies A and B prevented only 24.1% cases, whereas 35.3% of cases were potentially preventable as simulated by strategy C.

Assuming a daily intake of 500 g as the threshold value for beneficial effects, strategy C prevented the same proportion of cases. Strategy C also achieved higher health gain than the other two strategies as long as the initial mean intake was small. For increasing initial mean intake the proportion of cases preventable by strategy C decreased rapidly, whereas the decrease corresponding to strategies A and B was more moderate since they profited from intake changes above the recommended 400 g. Thus, strategies A and B overestimated the public health gain of the current dietary recommendation as long as the initial mean intake was larger than 300 g day^{-1}.

Besides simulating initial intake distributions, we estimated the distribution of the actual vegetable and

Table 3 Proportion of cases preventable by different strategies of increasing vegetable and fruit intake: results of a simulation study*

Initial mean intake (g day^{-1})	Preventable cases (%)					
	Threshold 400 g			Threshold 500 g		
	A†	B	C	A	B	C
200	24.1	24.1	35.3	27.2	26.4	35.3
250	21.8	21.8	29.2	25.9	25.0	29.2
300	18.7	18.7	23.0	24.1	22.8	23.0
350	15.1	15.1	17.1	21.8	20.0	17.1
400	11.1	11.1	11.6	18.7	16.5	11.6
450	6.9	6.9	6.9	15.1	12.3	6.9
500	3.3	3.3	3.3	11.1	8.1	3.3

* Intake distributions were simulated by adopting the shape of the French intake distribution obtained in the Enquête Individuelle et Nationale sur les Consommations Alimentaires (INCA survey)[17] and varying the mean intake. The relative risk associated with an increase of 160 g day^{-1} was set to 0.7, assuming a linear relationship between the logarithm of relative risk and intake for intakes below the threshold value.

† A – all individuals increase their intake by 160 g day^{-1}; B – only low consumers increase their intake by 160 g day^{-1}; C – low consumers increase their intake up to 400 g day^{-1}.

Table 4 Estimated usual intake distribution* for vegetable and fruit† intake (g day^{-1}) in France and Sweden based on food consumption surveys‡

	Percentile								
Country	5th	10th	25th	50th	75th	90th	95th	Arithmetic mean	Standard deviation
France	67.6	84.2	139.4	228.8	340.3	467.2	551.4	256.6	154.1
Sweden	58.0	77.8	126.8	201.6	296.3	403.3	500.3	229.9	143.1

* The usual intake distribution was estimated by the simplified Nusser method[19].

† Excluding fruit juice.

‡ The French survey was conducted in 1998/9, with 1474 adults[17]. The Swedish survey was conducted in 1997/8, with 1211 adults[18].

fruit intake in France and Sweden using data from national food consumption surveys (Table 4). The French survey INCA was conducted in 1998/9 with 1474 adults[17]. The Swedish dietary survey conducted in 1997/8 gathered data of 1211 adults[18]. Note that the survey data were based on short-term measurements (7-day records); however, the distribution of usual dietary intake defined as the long-term daily average was needed. We therefore applied the simplified Nusser method, which is a statistical procedure developed for this purpose and recommended by the EFCOSUM (European Food Consumption Survey Method) Group[19]. We see from Table 4 that the mean vegetable and fruit intake in France is somewhat higher than in Sweden, but in both countries the mean intake is markedly less than the target value of 400 g day^{-1} recommended by the WHO[6]. Thus, there is the need to increase the vegetable and fruit intake in the French and Swedish populations. Therefore, we applied the three simulation strategies described in Table 1 to change the usual intake distribution in both countries.

Tables 5 and 6 present preventable proportions for overall cancer and those of specific sites predicted for France and Sweden. Analogous to other studies[7–9] we considered three different scenarios – 'conservative', 'best guess' and 'optimistic' – and adopted the site-specific relative risks for each scenario from the WCRF[8]. 'Best guess' estimates use the mid-point of relative risks estimated by different authors; 'optimistic' estimates use relative risks from Block *et al.*[1] or Margetts *et al.*[4]; and 'conservative' estimates assume that smoking and drinking alcohol should first be discounted. In contrast to previous papers, we assumed a linear relationship between intake of vegetables and fruit and the logarithm of cancer risk only for daily intakes below a threshold of 500 g. This assumption is somewhat weaker than the overall linearity since it excludes prevention effects for increasing very high initial intakes.

Because of the threshold value used, the estimated proportions of preventable cancer cases in Tables 5 and 6 were somewhat lower than the values given by others[8]. Considering the 'best guess' scenario, the WCRF predicted a proportion of 22.7% by applying strategy A, whereas the corresponding values in France (Table 5) and Sweden (Table 6) were 20.6% and 17.0%, respectively. However,

Table 5 Preventable proportion of total cancer cases in France by strategies of increasing usual vegetable and fruit intake

		Scenario											
		Conservative				Best guess				Optimistic			
Cancer site	Proportion*	RR†	A‡	B	C	RR	A	B	C	RR	A	B	C
Mouth/pharynx	6.0	0.50	2.58	2.48	2.68	0.45	2.85	2.73	2.90	0.40	3.11	2.98	3.12
Larynx	1.8	0.50	0.78	0.74	0.80	0.45	0.85	0.82	0.87	0.40	0.93	0.89	0.94
Oesophagus	2.2	0.50	0.95	0.91	0.98	0.45	1.04	1.00	1.06	0.40	1.14	1.09	1.14
Lung	10.4	0.55	4.03	3.86	4.24	0.55	4.03	3.86	4.24	0.45	4.94	4.73	5.03
Stomach	3.1	0.50	1.34	1.28	1.38	0.45	1.47	1.41	1.50	0.40	1.61	1.54	1.61
Pancreas	1.7	0.70	0.44	0.42	0.48	0.60	0.58	0.56	0.63	0.35	0.96	0.91	0.94
Colon/rectum	13.4	0.70	3.44	3.31	3.80	0.60	4.60	4.42	4.93	0.50	5.77	5.54	5.98
Breast	14.5	1.00	0.00	0.00	0.00	0.85	1.86	1.79	2.14	0.75	3.10	2.99	3.48
Ovary	1.7	1.00	0.00	0.00	0.00	0.85	0.22	0.21	0.25	0.55	0.66	0.63	0.69
Cervix	1.6	1.00	0.00	0.00	0.00	0.85	0.21	0.20	0.24	0.50	0.69	0.66	0.71
Prostate	11.0	1.00	0.00	0.00	0.00	0.90	0.94	0.91	1.10	0.75	2.35	2.27	2.64
Bladder	5.6	0.70	1.44	1.39	1.59	0.60	1.92	1.85	2.06	0.50	2.41	2.31	2.50
Other	27.0	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Total	100		15.0	14.4	16.0		20.6	19.8	21.9		27.7	26.5	28.8

* Proportion of cancer site incidence in relation to total cancer incidence using data of Globocan 2000[20].

† Relative risk associated with an increase of vegetables and fruit intake by 160 g day^{-1}, adopted from the World Cancer Research Fund[8] and estimated by three different scenarios (conservative, best guess, optimistic). Relative risks for individuals who increase their intake by other amounts were calculated by linear interpolation and assuming a threshold of 500 g day^{-1}.

‡ A – all individuals increase their intake by 160 g day^{-1}; B – only low consumers increase their intake by 160 g day^{-1}; C – low consumers increase their intake up to 400 g day^{-1}.

Table 6 Preventable proportion of total cancer cases in Sweden by strategies of increasing usual vegetable and fruit intake

		Scenario											
		Conservative				Best guess				Optimistic			
Cancer site	Proportion*	RR†	A‡	B	C	RR	A	B	C	RR	A	B	C
Mouth/pharynx	1.9	0.50	0.86	0.84	0.94	0.45	0.95	0.92	1.02	0.40	1.04	1.01	1.09
Larynx	0.5	0.50	0.23	0.22	0.25	0.45	0.25	0.24	0.27	0.40	0.27	0.27	0.29
Oesophagus	0.9	0.50	0.41	0.40	0.45	0.45	0.45	0.44	0.48	0.40	0.49	0.48	0.52
Lung	6.8	0.55	2.77	2.70	3.08	0.55	2.77	2.70	3.08	0.45	3.40	3.31	3.64
Stomach	2.9	0.50	1.32	1.28	1.44	0.45	1.45	1.41	1.55	0.40	1.58	1.54	1.67
Pancreas	2.4	0.70	0.65	0.64	0.76	0.60	0.87	0.85	0.98	0.35	1.42	1.38	1.47
Colon/rectum	12.7	0.70	3.45	3.36	4.02	0.60	4.60	4.49	5.19	0.50	5.76	5.62	6.29
Breast	14.7	1.00	0.00	0.00	0.00	0.85	1.99	1.94	2.43	0.75	3.32	3.24	3.93
Ovary	2.2	1.00	0.00	0.00	0.00	0.85	0.30	0.29	0.36	0.55	0.90	0.87	1.00
Cervix	1.3	1.00	0.00	0.00	0.00	0.85	0.18	0.17	0.21	0.50	0.59	0.57	0.64
Prostate	15.1	1.00	0.00	0.00	0.00	0.90	1.36	1.33	1.69	0.75	3.41	3.33	4.04
Bladder	5.0	0.70	1.36	1.32	1.58	0.60	1.81	1.77	2.04	0.50	2.27	2.21	2.48
Other	33.6	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Total	100		11.1	10.8	12.5		17.0	16.6	19.3		24.5	23.8	27.1

* Proportion of cancer site incidence in relation to total cancer incidence using data of Globocan 2000[20].
† Relative risk associated with an increase of vegetables and fruit intake by 160 g day^{-1}, adopted from the World Cancer Research Fund[8] and estimated by three different scenarios (conservative, best guess, optimistic). Relative risks for individuals who increase their intake by other amounts were calculated by linear interpolation and assuming a threshold of 500 g day^{-1}.
‡ A – all individuals increase their intake by 160 g day^{-1}; B – only low consumers increase their intake by 160 g day^{-1}; C – low consumers increase their intake up to 400 g day^{-1}.

taking the adequate simulation strategy C to evaluate the health gain of the WHO recommendation, the differences were smaller. Applying strategy C, the estimated reduction in overall cancer incidence was 21.9% in France and 19.3% in Sweden.

Note that the preventable fraction of total cancer cases in France was larger than in Sweden although the initial vegetable and fruit consumption was higher in France than in Sweden. This result can be explained by the relative frequencies of cancer sites, which are very different between both countries. Whereas cancers of the mouth and pharynx, larynx, oesophagus and lung, which are more likely to be preventable by a diet rich in vegetables and fruits, have higher relative incidence rates in France, cancers weakly or not associated with vegetable and fruit intake occur relatively more in Sweden. For the comparison between the two countries, the differences in relative frequencies of cancer sites have a stronger effect on the overall prevention of cancer than the differences in the distribution of usual vegetable and fruit intake.

Discussion

We have presented a statistical method to evaluate the benefit of the current WHO recommendation to eat at least 400 g of vegetables and fruit daily. As an important feature, the method allows for individually different deficiencies of vegetable and fruit consumption and restricts the dietary changes necessary to meet the recommendation. Moreover, the method incorporates the initial intake distribution of the country or region. Because of these features, the method is more precise and flexible than the one applied hitherto. In particular, the statistical method is suitable to quantify the country-specific potential health benefit of the dietary recommendation and to compare the benefits for different countries.

Applying the method to France and Sweden results in estimated proportions of preventable cancer cases of 21.9% and 19.3%, respectively, which are not very different from the prognosis of 22.7% given by the WCRF[8]. However, this figure cannot be transferred to Southern Europe. Simulation results given in this paper (Table 3) suggest that European countries like Spain, Greece and Italy, where the mean total vegetable and fruit intake[10] is greater than 450 g day^{-1}, can prevent only less than 10% of cancer cases by increasing the consumption of vegetables and fruit.

The public health gain of the WHO recommendation estimated in the present paper must be interpreted as a potential one. We tacitly assumed that all individuals in a population, who are affected by a strategy, are willing and able to increase their vegetable and fruit intake. This assumption is not realistic. Commonly there will be a proportion p of non-responders. It can easily be shown that, to allow for non-responders, the proportion of preventable cases must be multiplied by $(1-p)$, as already mentioned by others[21]. For example, assuming that 40% of the French and Swedish population are non-responders, then the estimate of preventable cancer cases must be multiplied by 0.6 and, therefore, reduced to 13.1% and 11.6%, respectively. The factor $(1-p)$ can be interpreted as the response rate. In the more general case of stratum-specific response rates $(1-p_i)$, the proportion of preventable cases in stratum i must be multiplied by $(1-p_i)$ and the relative frequency of the stratum before summing the products over all strata.

To estimate the preventable proportion of diseases, an assumption concerning the dose–response relationship is needed. In previous publications[7–9], a linear relationship between the logarithm of relative risk and intake was assumed across the entire range of intakes. However, such an assumption implies that the relative risk tends to zero if the usual intake were to be increased indefinitely. An unrealistic implication of this linearity assumption would be that all cases might be prevented by increasing vegetable and fruit intake by sufficiently large amounts. This would contradict our knowledge about other important exposure pathways and risk factors. Therefore, other dose–response relationships should be considered. From the mathematical point of view, any non-linear function that is bounded and monotonically non-decreasing can be chosen, but no specific choice can be justified by biological knowledge at present. The simplest modification is to restrict the linearity assumption to intakes less than a threshold. The threshold can be interpreted as the optimal mean intake of vegetables and fruit. Clearly, we do not know this optimal amount and presume that it is equal to or greater than the pragmatic health target of 400 g day^{-1}. As can be seen from Table 3, the public health benefit of the current WHO recommendation evaluated by strategy C does not depend on the threshold provided that it exceeds the target value. However, if the optimal mean intake is much greater than 400 g day^{-1}, the public health benefit can be improved considerably by increasing the target value to the optimal one.

The proposed method can be applied to other kinds of exposure. Its application requires estimates of relative risks and exposure distributions. Since the validity of relative risks is presumably world-wide, whereas an exposure distribution refers to a country or region, two different data sources should be used. Relative risks should be estimated from meta-analyses of epidemiological studies summarising the available evidence. As long as well-accepted overall estimates are not available, different estimates covering the range of possible risks should be used as done in the present paper. On the other hand, epidemiological studies are not a suitable data basis to estimate exposure distributions since the study sample is seldom representative. Rather, national surveys should be preferred for estimating exposure distributions because of the mostly random selection of individuals and the large sample size. This procedure should ensure that the exposure distribution in the sample is similar to that in the population. Moreover, national surveys will be often be repeated periodically and allow real changes in exposure distribution to be explored. The presented statistical method can also be applied to estimate the proportion of prevented cases attributable to actual exposure changes.

Unfortunately, no explicit formula is available to calculate confidence intervals for the proportion of preventable cases estimated on the basis of two data sources. The difficulty is in incorporating the inaccuracy of relative risk estimates as well as random errors in estimating intake distributions. The consideration of three different scenarios with varying relative risks, as done in the present paper, reflected only the inaccuracy of risk estimates. A possible approach to determine confidence intervals that allow for all estimation errors and are not based on explicit formulas is to use statistical re-sampling methods like bootstrap. However, we did not apply such an approach because of the high effort required and difficulties in reconstructing the results.

The preventable proportion estimated in this paper quantifies the proportion of currently diseased individuals who would not have become diseased if a well-defined modified exposure distribution had existed in the past. This proportion generally differs from the proportion of diseased individuals that can be avoided in the future if the current exposure distribution was to be changed to the modified one. Murray and Lopez[22] use the terms ‘attributable’ and ‘avoidable’ burden of a disease to differentiate between past- and future-directed views. However, if all risk factors, with the exception of the exposure of interest, do not change markedly with time, the avoidable proportion of diseased individuals in the future will be approximately equal to the preventable proportion calculated in the present paper.

But, what is meant by future? For many diseases, the beneficial or detrimental effect of modifying the exposure distribution may not be expected immediately but may take many years. For example, migrant studies suggest that the delay between changing diet and the reduced or augmented emergence of colorectal cancer may be 10–20 years[8]. Thus, disease prevention strategies must allow for delay of the full impact after changing the exposure distribution.

Another important point is that an increase in vegetable and fruit consumption does not necessarily prevent cancer, but can also partly stand for a delay of onset. Consequently, if the incidence of a disease increases with age and if the age distribution in the population changes simultaneously as a result of an increased life expectancy, the overall effect of a prevention measure can be reduced or even vanish. In such cases, the preventable proportion should at first be calculated for age groups. Then, the adjusted overall effect of prevention can be estimated by the weighted average of the age-specific preventable proportions, with weights chosen as percentages of cases in age groups at the beginning of the prevention measure.

In conclusion, the statistical method presented here is a useful tool to quantify the public health benefit of a recommendation. Choosing the current exposure distribution as a starting point, individual exposure changes necessary to meet the recommendation must be simulated. The method allows the improvement of current estimates of the preventable fraction of diseases in most countries and regions. It combines data from different sources by applying a quite general statistical estimation formula.

References

1 Block G, Patterson B, Subar A. Fruit, vegetables, and cancer prevention: a review of the epidemiological evidence. *Nutrition and Cancer* 1992; **18**: 1–29.

2 Steinmetz KA, Potter JD. Vegetables, fruit, and cancer I. Epidemiology. *Cancer Causes & Control* 1991; **2**: 325–57.

3 Steinmetz KA, Potter JD. Vegetables, fruit, and cancer II. Mechanism. *Cancer Causes & Control* 1991; **2**: 427–42.

4 Margetts BM, Thompson R, Duffy S, on behalf of the Nutritional Epidemiology Working Group on Diet and Cancer. *A Review of the Epidemiological Literature linking Fruit and Vegetable Consumption to Risk of Cancer*, 1994.

5 Ness AR, Powles JW. Fruit and vegetables, and cardiovascular disease: a review. *International Journal of Epidemiology* 1997; **26**: 1–13.

6 World Health Organization (WHO). *Diet, Nutrition and the Prevention of Chronic Diseases: Report of WHO Study Group*. WHO Technical Series Report No. 797. Geneva: WHO, 1990.

7 Jansen MCJF, van't Veer P, Kok FK. *Fruits and Vegetables in Chronic Disease Prevention*. Wageningen: Landbouwuniversiteit, 1995.

8 World Cancer Research Fund (WCRF)/American Institute for Cancer Research (AICR). *Food, Nutrition and the Prevention of Cancer: A Global Perspective*. Washington, DC: WCRF/AICR, 1997.

9 Van't Veer P, Jansen MCJF, Klerk M, Kok FJ. Fruits and vegetables in the prevention of cancer and cardiovascular disease. *Public Health Nutrition* 2000; **3**: 103–7.

10 Joffe M, Robertson A. The potential contribution of increased vegetable and fruit consumption to health gain in the European Union. *Public Health Nutrition* 2001; **4**: 893–901.

11 Levin ML. The occurrence of lung cancer in man. *Acta Unio International contra Cancrum* 1953; **9**: 531–41.

12 Walter SD. Prevention for multifactorial diseases. *American Journal of Epidemiology* 1980; **112**: 409–16.

13 Morgenstern H, Bursic ES. A method for using epidemiologic data to estimate the potential impact of an intervention on the health status of a target population. *Journal of Community Health* 1982; **7**: 292–309.

14 Wahrendorf J. An estimate of the proportion of colo-rectal and stomach cancers which might be prevented by certain changes in dietary habits. *International Journal of Cancer* 1987; **40**: 625–8.

15 Drescher K, Becher H. Estimating the generalized impact fraction from case–control data. *Biometrics* 1997; **53**: 1170–6.

16 Graham P. Bayesian inference for a generalized population attributable fraction: the impact of early vitamin A levels on chronic lung disease in very low birthweight infants. *Statistics in Medicine* 2000; **19**: 937–56.

17 Volatier JL. *Enquête Individuelle et Nationale sur les Consommations Alimentaires*. Cachan: Editions TEC et DOC Lavoisier, 2000 [in French].

18 Becker W. *Riksmaten 1997–98. Dietary Habits and Nutrient Intake in Sweden 1997–98. The Second National Food Consumption Survey*. Uppsala: Livsmedelsverket, 2002 [in Swedish].

19 Hoffmann K, Boeing H, Dufour A, Volatier JL, Telman J, Virtanen M, *et al*. Estimating the distribution of usual dietary intake by short-term measurements. *European Journal of Clinical Nutrition* 2002; **56**(Suppl. 2): S53–62.

20 Ferley J, Bray F, Pisani P, Parkin DM, eds. *Globocan 2000. Cancer Incidence, Mortality and Prevalence Worldwide, Version 1.0*. International Agency for Research on Cancer (IARC) Cancer Base No. 5. Lyon: IARC Press, 2001 (http://www-dep.iarc.fr/globocan).

21 Nakayama T. Under-reporting of attributable risk and reporting of the risk ratio in epidemiologic literature. *Epidemiology* 2000; **11**: 366–7.

22 Murray CJL, Lopez AD. On the comparable quantification of health risks: lessons from the global burden of disease study. *Epidemiology* 1999; **10**: 594–605.

Appendix A – A general formula for the preventable proportion of cases

Consider the ith individual who changed their usual intake, in reaction to a prevention strategy, from x_i to x_i^*. As a consequence, the individual conditional probability for disease D has been changed from $P(D|X=x_i)$ to $P(D|X=x_i^*)$. The term

$$PP_i = \frac{P(D|X=x_i) - P(D|X=x_i^*)}{P(D|X=x_i)}$$

can be interpreted as the probability that the disease was prevented by an individual intake change from x_i to x_i^*. Thus, selecting randomly an individual in a population of size N, the probability that the disease D was prevented can be calculated by the total probability formula, and therefore is equal to:

$$PP = \frac{1}{N}\sum_{i=1}^{N} PP_i.$$

PP is the proportion of cases prevented by the whole spectrum of individual intake changes. Now, PP_i can be written as $1 - RR_i$, where RR_i denotes the relative risk $P(D|X=x_i^*)/P(D|X=x_i)$ associated with the dietary intake change of the ith individual. Therefore, PP can be calculated by the formula:

$$PP = \frac{1}{N}\sum_{i=1}^{N}(1 - RR_i) = 1 - \frac{1}{N}\sum_{i=1}^{N} RR_i.$$

In the special case that the relative risks RR_i of all individuals are the same, say equal to RR, the formula can be simplified to $PP = 1 - RR$, which is the formula that has already been used in the literature[8,9].

Appendix B – Estimating the proportion of cases preventable by a strategy

Assume that the logarithm of relative risk is a linear function of intake for intakes below a threshold T. Then, the relative risk associated with an increase of 160 g day^{-1} is constant as long as the increased intake does not exceed the threshold value. Let RR_0 be an estimate of this constant relative risk. Further, let x_i be the usual intake of the ith individual of a representative sample. Denoting the target value function of the prevention strategy by τ, the relative risk of the ith individual can be estimated by:

$$\widehat{RR_i} = RR_0^{\Delta_i/160},$$

with

$$\Delta_i = [\min(\tau(x_i), T) - \min(x_i, T)].$$

Substituting these expressions into the general formula of Appendix A gives an estimate of the proportion of cases preventable by the strategy.