



UvA-DARE (Digital Academic Repository)

**Assuming independence of risk factor prevalences in simulation models like PREVENT.
When are the outcomes seriously biased?**

van de Mheen, P.J.; Gunning-Schepers, L.J.

Published in:
European Journal of Public Health

DOI:
[10.1093/eurpub/7.2.216](https://doi.org/10.1093/eurpub/7.2.216)

[Link to publication](#)

Citation for published version (APA):

van de Mheen, P. J., & Gunning-Schepers, L. J. (1997). Assuming independence of risk factor prevalences in simulation models like PREVENT. When are the outcomes seriously biased? *European Journal of Public Health*, 7, 216-220. <https://doi.org/10.1093/eurpub/7.2.216>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Assuming independence of risk factor prevalences in simulation models like PREVENT

When are the outcomes seriously biased?

PERLA J. VAN DE MHEEN, LOUISE J. GUNNING-SCHEPERS *

Little is known about the clustering of risk factors at a nation-wide level. As a result the prevalence of combinations of risk factors in models like PREVENT, designed to calculate the health benefits of a change in risk factor prevalences, is computed assuming an independent distribution. This assumption may not be valid. The aim of the present study was to quantify the maximum extent to which outcome measures of PREVENT may be biased, if the assumed independent distribution of risk factors is incorrect. We therefore calculated to what extent the life expectancy and the potential years of life gained were biased when independent risk factor prevalences were assumed, while they were in fact completely dependent. We used population data, mortality figures and risk factor prevalences from The Netherlands to obtain a realistic estimate of how serious the bias might be. Furthermore, sensitivity analyses were carried out to explore the extent of bias in the case of different risk factor prevalences. The results show that the assumed independence has little impact on the estimated life expectancy and the potential years of life gained, both in the case of the current risk factor prevalences and in the case of higher or lower prevalences. Given that the dependency between risk factors will probably be smaller in reality, we conclude that the assumption of independence may be used since it is not likely to cause substantial bias. This greatly reduces the data requirements necessary as input for simulation models such as PREVENT.

Key words: clustering, independent risk factors, simulation models, life expectancy, years of life gained

The PREVENT model was designed specifically for policy makers, to enable them to weigh policy alternatives quantitatively.^{1,2} This simulation model calculates the potential health benefits of primary prevention programmes that focus on reducing risk factor prevalences. The model is not used for analysis of empirical data, but rather to bring together information available from empirical studies for decision-making purposes at the population level. It uses the currently available information to quantify the future effects of changing risk factor prevalences in a population.

The methodology of the model is based on the potential impact fraction, a well-known epidemiological measure. PREVENT uses existing epidemiological knowledge about the relationship between risk factors and mortality and combines this with a dynamic population model to include demographic effects and interrelationships between causes of death. Another feature of the model is that a time dimension has been incorporated, to simulate a gradual reduction in excess risk after cessation of exposure.³ Furthermore, mortality risks have been linked through common risk factors, to include the fact that for instance a change in smoking behaviour in the population

will affect not only the coronary heart disease mortality rate, but also the mortality rates of lung cancer, stroke and chronic obstructive lung diseases. The population in PREVENT consists of several subgroups characterized by exposure to certain risk factors, e.g. part of the population may smoke and therefore increase its risk of several causes of death. Since individuals may be exposed to more than 1 risk factor, the prevalence of combinations of risk factors is also needed. Unfortunately, these are not available at a population level in The Netherlands. Therefore, an independent distribution of risk factor prevalences is assumed in PREVENT.

That risk factors cluster more than expected under the assumption of independence has been shown by several authors. Criqui et al.⁴ showed that clustering of cardiovascular risk factors (smoking, high blood pressure, high cholesterol levels and obesity) was strongest in subjects with the highest levels of these risk factors. This means that persons at greater risk of 1 risk factor for cardiovascular disease, also have a higher risk of more risk factors. The Bogalusa Heart Study shows an example of clustering of cardiovascular disease risk factors at a younger age (5–24 years of age).^{5,6} Obese school children had more clustering of other risk factors than could be expected, assuming an independent distribution of risk factors. In The Netherlands, Kok et al.⁷ have shown that smoking, obesity, physical inactivity and inadequate nutrition clustered more than expected under the assumption of independent risk factor prevalences. Clustering of risk factors

* P.J. van de Mheen¹, L.J. Gunning-Schepers¹
Academic Medical Center, Institute of Social Medicine, Amsterdam,
The Netherlands

Correspondence: Ms. Perla J. van de Mheen, PhD, Academic Medical Center,
Institute of Social Medicine, Meibergdreef 15, 1105 AZ Amsterdam,
The Netherlands, tel. +31 20 5664892, fax +31 20 6972316

may be an issue when one is interested in the magnitude of the risk associated with a combination of risk factors, in order to find the group of people that is at extremely high risk. However, is it an issue in public health when one is interested in the average health or mortality of a population?

The aim of the present study was to estimate the maximum extent of bias in the estimated life expectancy, 1 of the outcome measures of PREVENT, if an independent distribution of risk factor prevalences is wrongly assumed. Since overall life expectancy may be a relatively insensitive measure, we also estimated the maximum bias in the potential years of life gained. Dutch data were used to calculate to what extent these outcome measures were biased when independent risk factor prevalences would be assumed, while the risk factor prevalences were completely dependent in reality. Since the observed dependence between risk factors is probably smaller than this complete dependence, the bias in practice will probably be smaller than calculated in this paper. Furthermore, the influence of the assumption in the case of higher and lower risk factor prevalences was explored, since these may exist in certain groups of the population or in other countries and this higher or lower prevalence may result in a different extent of bias in the outcomes of PREVENT. If the assumption would not affect the outcome measures of PREVENT, this would greatly reduce data requirements for the input data and enable people to use the model with the data already available.

METHODS

Details of the basic methodology of PREVENT can be found in the *appendix*. Since PREVENT is based upon life table techniques, standard life table techniques⁸ were used to calculate the mortality experience of a cohort of men from 0 to 95 years of age, using Dutch mortality data.⁹ As in PREVENT, the cohort was assumed to consist of subgroups characterized by exposure to different combinations of the risk factors: smoking, hypertension and hypercholesterolaemia. For smoking, a distinction was made between current smokers, former smokers and never smokers. For the other risk factors, we distinguished only 2 exposure categories: with or without the risk factor. The prevalences of the risk factors were taken from a representative sample of the Dutch population.¹⁰ The population thus consisted of 40.8% of smokers, 9% of persons with hypertension and 19% of persons with hypercholesterolaemia. The proportion of former smokers, however, was not reported in that sample. We assumed 45% to be a realistic estimate, as we had calculated in an earlier study.¹¹ We furthermore assumed that exposure to risk factors occurred from the age of 20 years onwards. For smokers this age is reported in other studies.¹² The relative risks of death were used to quantify the higher risk of those exposed compared to those not exposed to the risk factor and these were taken from published prospective studies.¹³⁻¹⁷

At every age, the mortality experience for each subgroup was calculated assuming independent risk factor prevalences.

On the basis of the mortality rates for each subgroup, the mortality rate in the total population and the life expectancy were calculated. Then the total mortality rate was computed in the case of completely dependent risk factor prevalences, assuming that the mortality rate for each subgroup was the same as in the case of independent risk factor prevalences, so that the only difference was the prevalence of combinations of risk factors. Completely dependent risk factor prevalences for this purpose were defined as all persons with hypertension who also had hypercholesterolaemia and were (former) smokers. Given a difference in the prevalences of smoking, hypertension and hypercholesterolaemia, with the prevalence of hypertension being the lowest, this definition will give the maximum extent of dependence between the 3 risk factors.

Since the overall life expectancy may be a relatively insensitive outcome measure, we also investigated the influence of the assumption of independence on the potential years of life gained. An intervention was simulated reducing the overall prevalence of smoking by 50%, thereby increasing the prevalence of former smokers and assuming that the dependency of risk factors was not affected by the intervention (i.e. completely dependent risk factors were also completely dependent after the intervention). Dutch population¹⁸ and mortality data were used to assess whether the effect of an intervention in terms of the potential years of life gained would be greatly overestimated or underestimated by wrongly assuming an independent distribution of risk factor prevalences.

Sensitivity analyses

It was tested whether the results were sensitive to the relative risks chosen for joint exposure of risk factors. We used multiplicative versus additive relative risks, which can be seen as the 2 extremes reported in epidemiological research. Multiplicative risks are often assumed. However, Silberberg¹⁹ found, using coronary heart disease death rates from the population screened for the MRFIT study, that the relationship between cholesterol, smoking and blood pressure was closer to additive than to multiplicative. We therefore initially used multiplicative relative risks, with additive relative risks as an alternative. Furthermore, it was assessed whether the results were sensitive to the magnitude of the risk factor prevalences. The effect of the assumption was evaluated in the case of 25% higher risk factor prevalences and in the case of 50% lower risk factor prevalences. Moreover, 2 analyses were carried out to test whether the results were sensitive to a smaller difference between the overall risk factor prevalences. In the first analysis the effect of the assumption was estimated in the case of a 100% higher prevalence of hypertension and a 100% higher prevalence of hypercholesterolaemia (i.e. 40.8% of smokers, 18% of the population with hypertension and 38% of the population with hypercholesterolaemia). In the second analysis the effect of the assumption was evaluated in the case of a 50% lower prevalence of smoking and no change in the prevalences of hypertension and hypercholesterolaemia.

RESULTS

Table 1 shows that the estimated life expectancy at different ages is only slightly biased by wrongly assuming independent risk factor prevalences, both when multiplicative and additive relative risks are used for joint exposure to risk factors. The possible bias introduced by wrongly assuming an independent distribution of risk factors is strongest in the case of additive relative risks for joint exposure and at older ages but is still only 0.1 and 1.5% respectively. Part of this minimal bias may be due to the fact that the overall life expectancy is a relatively insensitive outcome measure. Table 2 shows the effect of the assumption on the potential years of life gained. When independent risk factors are wrongly assumed, the effect of the intervention is overestimated in the case of multiplicative relative risks for joint exposure to risk factors, and underestimated in the case of additive relative risks. However, as a percentage of the total years lived by the average Dutch population in that year, the bias is smaller than 1%. Expressed as a percentage of the total effect of the intervention, the bias introduced by this assumption is also around 1% (data not shown). The bias in the estimated potential years of life gained was slightly stronger when an intervention was simulated that reduced the proportion of individuals exposed to 3 risk factors by 50%, but the bias remained below 1% of the total number of years lived (data not shown).

Sensitivity analyses

Another reason for the minimal error due to wrongly assuming an independent distribution of risk factors, may be the rather small risk factor prevalences. There may be subgroups within the population with higher risk factor prevalences, in which the assumption of independent risk factor prevalences may lead to serious bias in the estimated life expectancy or the potential years of life gained. Tables 3 and 4 show that the possible bias due to

the assumption in estimated life expectancy is approximately the same and thus very small. However, in the case of a smaller difference between the overall prevalence of the risk factors, the possible bias in the estimated life expectancy increases (tables 5 and 6). The possible bias is still small, ranging from 0.1% in life expectancy at birth, to 2.4% in life expectancy at 85 years of age in the case of multiplicative relative risks for joint exposure and 100% higher prevalences of hypertension and hypercholesterolaemia (table 5). The possible bias in the estimated potential years of life gained also increases, but remains below 1% (data not shown).

Table 1 Effect of assuming independent risk factor prevalences on estimated life expectancy

Life expectancy	Multiplicative relative risks		Additive relative risks	
	Independent risk factors	Dependent risk factors	Independent risk factors	Dependent risk factors
At birth	73.85	73.84	73.85	73.91
At 65 years of age	14.06	14.14	14.06	14.13
At 85 years of age	4.55	4.62	4.55	4.61

Table 2 Overestimation of the potential years of life gained due to wrongly assuming an independent distribution of risk factor prevalences (intervention: 50% reduction of smoking prevalence)

	Multiplicative relative risks		Additive relative risks	
	Number	% of total years lived	Number	% of total years lived
All ages	952	0.01	-847	-0.01
<65 years	1651	0.03	-25	-0.00
≥65 years	-699	-0.10	-822	-0.11

Table 3 Effect of assumption in the case of 25% higher risk factor prevalences

Life expectancy	Multiplicative relative risks		Additive relative risks	
	Independent risk factors	Dependent risk factors	Independent risk factors	Dependent risk factors
At birth	73.85	73.84	73.85	73.90
At 65 years of age	14.06	14.13	14.06	14.11
At 85 years of age	4.55	4.61	4.55	4.60

Table 4 Effect of assumption in the case of 50% lower risk factor prevalences

Life expectancy	Multiplicative relative risks		Additive relative risks	
	Independent risk factors	Dependent risk factors	Independent risk factors	Dependent risk factors
At birth	73.85	73.84	73.85	73.90
At 65 years of age	14.06	14.12	14.06	14.11
At 85 years of age	4.55	4.59	4.55	4.58

Table 5 Effect of assumption in the case of 100% higher prevalences of hypertension and hypercholesterolaemia

Life expectancy	Multiplicative relative risks		Additive relative risks	
	Independent risk factors	Dependent risk factors	Independent risk factors	Dependent risk factors
At birth	73.85	73.84	73.85	73.94
At 65 years of age	14.06	14.18	14.06	14.16
At 85 years of age	4.55	4.66	4.55	4.64

Table 6 Effect of assumption in the case of 50% lower prevalences of smoking

Life expectancy	Multiplicative relative risks		Additive relative risks	
	Independent risk factors	Dependent risk factors	Independent risk factors	Dependent risk factors
At birth	73.85	73.84	73.85	73.94
At 65 years of age	14.06	14.17	14.06	14.16
At 85 years of age	4.55	4.63	4.55	4.62

DISCUSSION

The present study indicates that an assumed independent distribution of risk factor prevalences in simulation models like PREVENT is not likely to have a substantial influence on the estimated life expectancy at birth or the potential years of life gained, given the current level of 3 traditional risk factors. Should risk factors be added with much higher or lower prevalences, the bias in the outcome measures may increase if independence is wrongly assumed. Furthermore, in the case of risk factors with prevalences of equal magnitude, the impact of the assumption may be stronger. However, the sensitivity analyses in this study indicate that even in these cases the bias is very small. Moreover, the dependency of risk factors is probably smaller in reality, suggesting that the possible bias of the outcome measures will be smaller than calculated in this paper. The lack of impact of this assumption is caused by the fact that assuming complete dependency of risk factors leads to a simultaneous increase in the prevalence of people not exposed to any risk factor. In this way, the higher mortality due to a higher prevalence of people exposed to 3 risk factors is counterbalanced by the fact that the prevalence of people not exposed to any risk factor is also higher.

In this paper we only simulated the mortality experience and the influence of differential mortality. In general, differential mortality will result in smaller proportions of people exposed with increasing age and the proportion of people not exposed will increase. For risk factors such as hypertension and hypercholesterolaemia, however, the prevalence is thought to increase with age.²⁰ This would only affect our results if this increase with age were to differ between individuals exposed and not exposed to other risk factors. To our knowledge, it is not known whether this increase with age is different for exposure groups. Furthermore, Löwik et al.²¹ found in an elderly population that the risk factors smoking, hypertension, hypercholesterolaemia and obesity did not cluster more than expected under the assumption of independence. Given that a stronger clustering than expected under the assumption of independence is found at younger ages,⁷ the findings of Löwik et al.²¹ might be the result of differential mortality.

Therefore, although clustering of risk factors may be an issue with regard to the (reduction of) risk associated with that clustering, it is not likely to be an issue in terms of the average health or mortality in a population. Our results indicate that even in the case of risk factors being completely dependent, the bias in the outcome measures of models such as PREVENT is very small. Since the

dependency between risk factor prevalences is probably smaller in reality, the bias in outcome measures will also be smaller. Only in the case of extremely high risk factor prevalences of approximately the same magnitude with very strong de-

pendency, will the outcome measures possibly be biased. However, given the current risk factor prevalences, this is not very likely to occur simultaneously. This study suggests that we may then assume independence of risk factors, which will greatly reduce the data requirements needed for models such as PREVENT and enable people to use the model with the data already available.

REFERENCES

- Gunning-Schepers LJ, Barendregt JJ, van der Maas PJ. Population interventions reassessed. *Lancet* 1989;i:479-81.
- Gunning-Schepers LJ. The health benefits of prevention: a simulation approach. *Hlth Policy* 1989;12:1-256.
- Gunning-Schepers LJ, Barendregt JJ. Timeless epidemiology or history cannot be ignored. *J Clin Epidemiol* 1992;45:365-72.
- Criqui MH, Barrett-Connor E, Holdbrook MJ, Austin M, Turner JD. Clustering of cardiovascular disease risk factors. *Prevent Med* 1980;9:525-33.
- Smoak CG, Burke GL, Webber LS, Harsha DW, Srinivasan SR, Berenson GS. Relation of obesity to clustering of cardiovascular disease risk factors in children and young adults: the Bogalusa heart study. *Am J Epidemiol* 1987;125:364-72.
- Webber LS, Voors AW, Srinivasan SR, Frerichs RR, Berenson GS. Occurrence in children of multiple risk factors for coronary artery disease: the Bogalusa heart study. *Prevent Med* 1979;8:407-18.
- Kok FJ, Matroos AW, van den Ban AW, Hautvast JGAJ. Characteristics of individuals with multiple behavioral risk factors for coronary heart disease: The Netherlands. *Am J Public Hlth* 1992;72:986-91.
- Namboodiri K, Suchindran CM. Life table techniques and their applications. Orlando: Academic Press, 1987.
- Central Bureau of Statistics. Life tables, 1990 and 1986-1990. *Monthly Popul Stat* 1991;12:74-5.
- Kromhout D, Obermann-de Boer GL, Blokstra A, Verschuren WMM. Monitoring cardiovascular diseases 1990 (in Dutch). Bilthoven: Rijks Instituut voor Volksgezondheid en Milieuhygiëne (RIVM), 1991.
- van de Mheen PJ, Gunning-Schepers LJ. Reported prevalences of former smokers in survey data: the importance of differential mortality and misclassification. *Am J Epidemiol* 1994;140:52-7.
- US Department of Health and Human Services. Reducing the health consequences of smoking: 25 years of progress. Rockville, MD: US Department of Health and Human Services, Office on Smoking and Health, 1989.
- US Department of Health and Human Services. The health benefits of smoking cessation. Rockville, MD: US Department of Health and Human Services, Public Health Service, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 1990.
- Doll R, Peto R. Mortality in relation to smoking: 20 years' observation on male British doctors. *BMJ* 1976;2:1525-36.
- Kannel WB. The role of blood pressure in cardiovascular morbidity and mortality. *Prog Cardiovasc Dis* 1974;17:5-24.
- Yano K, McGee D, Reed DM. The impact of elevated blood pressure upon 10-year mortality among Japanese men in Hawaii: the Honolulu Heart Program. *J Chronic Dis* 1983;36:569-79.
- Cowan LD, O'Connell DL, Criqui MH, Barrett-Connor E, Bush TL, Wallace RB. Cancer mortality and lipid and lipoprotein levels: the lipid research clinics program mortality follow-up

study. *Am J Epidemiol* 1990;131:468-82.

18 Central Bureau of Statistics. Age composition of the Netherlands, January 1st 1990. *Monthly Popul Stat* 1990;4:24-5.

19 Silberberg JS. Estimating the benefits of cholesterol lowering: are risk factors for coronary heart disease multiplicative? *J Clin Epidemiol* 1990;43:875-9.

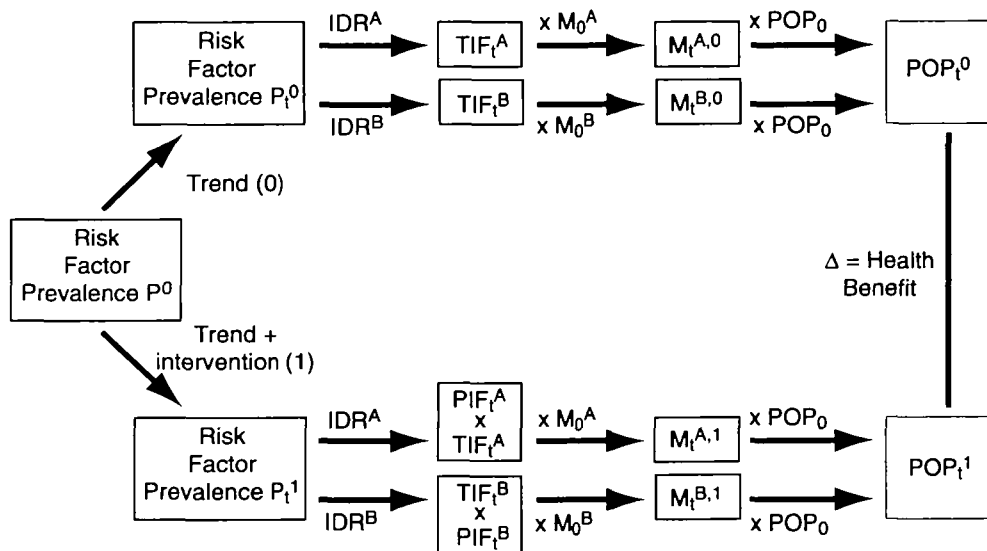
20 Whelton PK. Blood pressure in adults and the elderly. In: Bulpitt CJ, editor. *Epidemiology of hypertension*. Amsterdam,

New York, Oxford: Elsevier, 1985:51-69.

21 Lówik MRH, Borsboom ABG, Bun CJE, Kok FJ, Schouten EG, Odink J. Clustering of cardiovascular risk indicators in the elderly (in Dutch). *Hart Bull* 1991;22:95-9.

Received 22 August 1995, accepted 1 November 1995

Appendix Basic methodology of PREVENT model ^a



$$PIDR_t^{r,z,s,A} = \sum_{n=1}^{cn} \sum_{i=0}^{ID} P_{t-LAT^z}^{r,z,s,A,n,i} IDR^{r,z,s,A,n,i}$$

$$TIF_t^{r,z,s,A} = \frac{PIDR_0^{r,0,z,s,A} - PIDR_t^{r,0,z,s,A}}{PIDR_0^{r,0,z,s,A}}$$

$$PIF_t^{r,z,s,A} = \frac{PIDR_t^{r,0,z,s,A} - PIDR_t^{r,1,z,s,A}}{PIDR_t^{r,0,z,s,A}}$$

$$M_t^{0,s,A} = M^{s,A} - \sum_{z=1}^{\pi} TIF_t^{z,s,A} M^{z,s,A}$$

$$M_t^{1,s,A} = M^{s,A} - \sum_{z=1}^{\pi} [1 - (1 - TIF_t^{z,s,A})(1 - PIF_t^{z,s,A})] M^{z,s,A}$$

Where:

P: proportion
 cn: total number of exposure categories
 r: index for risk factor
 i: index for ex-exposure level
 A: index for age
 z: index for disease

IDR: incidence density ratio
 n: index for exposure category
 ID: total number of ex-exposure levels
 j=0,1: index for reference (0) or intervention population (1)
 s: index for sex
 t: index for time

$M^{s,A}$: constant overall mortality quotient

$M_t^{j,s,A}$: adjusted overall mortality quotient