



University
of Glasgow

Kooter, A.J., Kostense, P.J., Groenewold, J., Thijs, A., Sattar, N. and Smulders, Y.M. (2011) *Integrating information from novel risk factors with calculated risks: the critical impact of risk factor prevalence*. *Circulation*, 124 (6). pp. 741-745. ISSN 0009-7322

<http://eprints.gla.ac.uk/57604>

Deposited on: 18 November 2011

Circulation

JOURNAL OF THE AMERICAN HEART ASSOCIATION



Integrating Information From Novel Risk Factors With Calculated Risks : The Critical Impact of Risk Factor Prevalence

Albertus J. Kooter, Piet J. Kostense, Jan Groenewold, Abel Thijs, Naveed Sattar and Yvo M. Smulders

Circulation 2011, 124:741-745

doi: 10.1161/CIRCULATIONAHA.111.035725

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75214

Copyright © 2011 American Heart Association. All rights reserved. Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://circ.ahajournals.org/content/124/6/741>

Data Supplement (unedited) at:

<http://circ.ahajournals.org/content/suppl/2011/08/06/124.6.741.DC1.html>

Subscriptions: Information about subscribing to *Circulation* is online at

<http://circ.ahajournals.org/subscriptions/>

Permissions: Permissions & Rights Desk, Lippincott Williams & Wilkins, a division of Wolters Kluwer Health, 351 West Camden Street, Baltimore, MD 21202-2436. Phone: 410-528-4050. Fax: 410-528-8550. E-mail:

journalpermissions@lww.com

Reprints: Information about reprints can be found online at

<http://www.lww.com/reprints>

Integrating Information From Novel Risk Factors With Calculated Risks

The Critical Impact of Risk Factor Prevalence

Albertus J. Kooter, MD; Piet J. Kostense, PhD; Jan Groenewold, PhD; Abel Thijs, MD; Naveed Sattar, MD; Yvo M. Smulders, MD

Case vignette: a 60-year-old man visits his physician for assessment of his 10-year cardiovascular risk. On the basis of his systolic blood pressure, lipid profile, smoking status, and the fact that he is nondiabetic, the Framingham risk score estimates his risk to be 8%. The physician wonders if he could further specify the patient's risk by performing an additional test like coronary calcium score or microalbuminuria (MA). For matters of convenience and costs he decides to test MA, which turns out positive. Assuming that MA has an invariable and exact relative risk (RR), independent from the aforementioned classical risk factors, of 2.0, what would this man's estimated risk become?

The Problem

Prediction of absolute disease risk is an essential component of cost-effective disease prevention strategies. In cardiovascular disease (CVD) prevention, for example, antiplatelet and statin therapy is applied if absolute risk of CVD is considered sufficiently high. Various prediction models are available for the purpose of risk calculation. These models are derived from large population-based cohorts in which conventional CVD risk factors and prospective event registrations are available. Well known examples include the Framingham risk score and the risk model of the European SCORE consortium.^{1,2}

Obviously, with regard to individual risk estimation, risk models have inherent shortcomings in terms of precision and reliability. In an attempt to improve risk prediction, much focus has been on the potential benefit of adding information relating to novel risk factors. Various statistical methods have been developed to assess the ability of novel risk factors to improve risk stratification. These methods include assessment of discrimination and calibration of the conventional versus the updated risk model.^{3,4} The ultimate goal of adding novel risk factors is to improve a patient's health by correctly reclassifying him or her into high, intermediate, and low risk

categories for which the net reclassification improvement is one appropriate parameter.^{5,6}

Although models may, as judged from the net reclassification improvement, improve as a result of including a novel risk factor, such expanded models are hardly used in clinical practice. Moreover, literature addressing novel risk factors often does not provide these expanded risk models but, instead, provides an independent RR or standardized β of the novel risk factor.

Integrating a novel risk factor in a new model is very different from using a novel risk factor on top of an existing model. In the latter context, the model delivers a baseline risk and the independent RR from the novel risk factor must somehow be used to convert this baseline risk into a recalculated risk. Although several national and international guidelines encourage the use of novel risk factors, they do not describe how to obtain a new recalculated risk using this additional information. Intuitively, and sometimes explicitly, the RR of novel risk factor is directly translated into a multiplication factor.^{7,8} In other words, the risk of the patient in the case vignette would be multiplied by 2.0 to give a recalculated risk of 16%, assuming that the RR of 2 implies doubling of risk. We will explain that this reasoning is incorrect.

Imagine the RR and the multiplication factor to be identical. In the example of the case vignette, if the multiplication factor would be either 2.0 (MA present) or 1.0 (MA absent), the recalculated risk can only remain unchanged or adjusted upward but never downward. Hence, in the stratum (ie, the imaginary group of individuals with the same Framingham risk factor profile), the average risk would increase, solely by adding risk information. This cannot be correct because adding risk information simply cannot increase average risk. Hence, upward adjustments of risk due to presence of an additional risk factor in some individuals should be compensated by downward adjustments for absence of the same risk

From the VU University Medical Center, Department of Internal Medicine, Amsterdam, the Netherlands (A.J.K., P.J.K., A.T., Y.M.S.); Debye Institute for Nanomaterials Science, Utrecht, the Netherlands (J.G.); Division of Cardiovascular and Medical Sciences, BHF Cardiovascular Research Building, Faculty of Medicine, University of Glasgow, Scotland (N.S.).

The online-only Data Supplement is available with this article at <http://circ.ahajournals.org/lookup/suppl/doi:10.1161/CIRCULATIONAHA.111.035725/-/DC1>.

Correspondence to Albertus J. Kooter, MD, Department of Internal Medicine, VU Medical Centre, De Boelelaan 1117, P.O. Box 7057, 1007 MB, the Netherlands. E-mail jkooter@vumc.nl

(*Circulation*. 2011;124:741-745.)

© 2011 American Heart Association, Inc.

Circulation is available at <http://circ.ahajournals.org>

DOI: 10.1161/CIRCULATIONAHA.111.035725

factor in other individuals, eventually generating a stable risk for each risk stratum. Literature on high-sensitivity C-reactive protein (hsCRP), for example, explains that hsCRP has the ability to reclassify upward as well as downward.⁹ Although this may seem obvious, many guidelines describe novel (emerging) risk factors as principally being helpful to place patients from intermediate- into high-risk categories.^{10–12}

It follows that the RR of the novel risk factor and the multiplication factor for baseline risk cannot be considered synonymous. So how can we translate the RR into a multiplication factor. In the case vignette, if the RR is indeed 2.0, how does the patients risk change in case of presence or absence of MA?

Below, we will focus on the principles and mathematics when the novel risk factor is a dichotomous (like the one presented in the case vignette), ordinal, or continuous risk factor. The conclusion is sobering: Even perfect knowledge of the exact and truly independent RR of a novel risk factor is insufficient without a reliable estimate of its prevalence. We explain this in the context of CVD risk stratification, but the principles outlined apply to risk prediction in every clinical context, in risk models as well as predictions based on clinical experience.

The Solution for a Dichotomous Risk Factor

In the example given, the additional risk information consisted of a dichotomous variable with a RR of 2.0.¹³ An intuitive response might be that the presence of MA would increase the risk by a factor 2.0, so from 8% to 16%. We will explain why this is incorrect.

Firstly, consider that the risk of 8% for this individual is in fact an estimate for a mix of persons that, together, make up the stratum of people with the same profile of conventional (Framingham) risk factors. Some in this stratum will have had MA and some will not. Therefore, the 8% should be regarded as a weighted mean of the risks for those with and those without MA.

Imagine now, for the sake of argument, that over 99% of persons in the original 8%-risk stratum would have had MA. Then, the risk of the patient in the example would not increase by any significant margin by having MA (multiplication factor ≈ 1.0). After all, the person was already expected to have MA, so having MA does not change risk appreciably. However, in the exceptional case of a patient with this risk profile not having MA, his risk would be substantially lower than 8%. As the defined RR is 2.0, his risk should thus be $\approx 4\%$ (multiplication factor ≈ 0.5). Since he is a rare individual in this stratum, the baseline average risk remains unaffected.

Conversely, if $<1\%$ of individuals in the stratum would have had MA, not having MA would not change the patient's risk by any considerable margin (multiplication factor ≈ 1.0). Having MA, on the other hand, although a rare event, increases his risk by almost the full multiplication factor of 2.0, thus reaching $\approx 16\%$. Again, the baseline risk will be 8% because this individual with MA is a rare one.

It follows that the RR of 2.0 translates into varying multiplication factors and, as demonstrated above, the crucial determinant of the variation of the multiplication factor is its

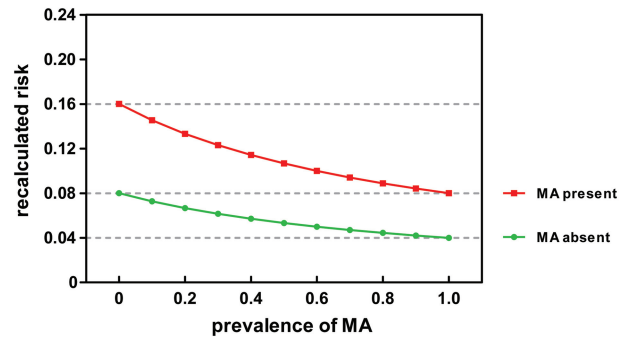


Figure 1. Association between the prevalence of MA and recalculated risk. The recalculated risk is presented for a patient as in the case vignette: a baseline risk of 0.08 (8%), and a novel risk factor (ie, MA) with a consistent RR of 2.0. MA indicates microalbuminuria.

prevalence. Depending on the prevalence of MA in the stratum in which a patient fits, presence of MA translates into a multiplication factor of between 1.0 (if prevalence of MA is near 100%) and 2.0 (if prevalence of MA approaches 0%). Conversely, the absence of MA translates into a multiplication factor of between 0.5 (if prevalence is near 100%) and 1.0 (if prevalence is near 0%).

In less extreme examples of prevalence, the multiplication factors vary within the margins indicated, and what is required now to convert RRs into multiplication factors for baseline risk is calculation of the weighted mean RR.

For example, if the prevalence of MA in the original 8%-risk stratum of the example is 60%, the risks for those with and without MA would be 10% and 5%, respectively (Formula 1, further explained in online-only Data Supplement). This can be understood as follows: The difference in risk between the groups with and without MA always corresponds to the RR of 2.0 for MA which, after all, was considered to be precise and reliable across all risk strata. The original 8% thus must be a weighted mean of those with and those without MA and the risks differing by a factor of 2.0. From these components, risks for those with and without MA can be calculated as 10% and 5%, respectively, and indeed, $0.6 \times 10\% + 0.4 \times 5\% = 8\%$.

Formula 1

$$[r] = p \cdot RR + (1 - p)$$

$$MF_{(+)} = RR/[r]$$

$$R_{(+)} = R_{bl} \cdot MF_{(+)}$$

$$R_{(-)} = R_{(+)} / RR$$

where $[r]$ = weighted mean RR; p = prevalence of novel risk factor; RR = relative risk of the novel risk factor; $MF_{(+)}$ = multiplication factor for R_{bl} to obtain $R_{(+)}$; $R_{(+)}$ = recalculated risk in presence of novel risk factor; R_{bl} = baseline risk; and $R_{(-)}$ = recalculated risk in absence of novel risk factor.

A graphic representation of how the recalculated risk (after applying the novel risk factor) depends on both RR and prevalence of MA is shown in Figure 1. A generic Figure (for

any dichotomous novel risk factor) is presented in the online-only Data Supplement.

The Solution for Ordinal Risk Factors

Some novel risk factors are separated in >2 categories, for example hsCRP.¹⁴ Imagine that the physician in the example case would have opted for measurement of hsCRP instead of MA to recalculate risk. High-sensitivity C-reactive protein is commonly divided into 3 categories: hsCRP <1 mg/L, 1 to 3 mg/L and >3 mg/L. Suppose now that, in the risk stratum to which the patient is allocated on the basis of the conventional risk-factor profile, the (hypothetical) prevalences of these 3 categories are 0.1, 0.4, and 0.5, with corresponding RRs of 1.0, 1.25, and 2.0.

As is the case for dichotomous risk factors, the first step is to define a weighted mean RR for the entire stratum, which is determined by the RRs of the categories as well as their prevalences. The weighted mean RR for the example presented = 0.1 × 1 + 0.4 × 1.25 + 0.5 × 2 = 1.6.

We now compare the RR of the patient to this weighted mean RR. Our patient’s hsCRP is 2.2 mg/L, corresponding to a RR of 1.25. His baseline risk was 8%. The recalculated risk will be (1.25/1.6) × 8% = 6.3%. Thus, although having an added risk factor with a RR higher than 1.0 (ie, 1.25), his risk is adjusted downward because most patients in his stratum have the higher RR of 2.0.

The mathematical relationship between recalculated risk and prevalence of an ordinal novel risk factor for any level of RR and prevalence (P) is shown in Formula 2 (further explained in the online-only Data Supplement).

Formula 2

$$[r] = p_1 \cdot RR_1 + p_2 \cdot RR_2 \dots + p_n \cdot RR_n$$

$$MF_x = RR_x / [r]$$

$$R_x = R_{bi} * MF_x$$

where [r] = weighted mean RR; p₁ = prevalence of lowest level of risk; p_n = prevalence of highest level of risk; RR₁ = relative risk of lowest level of risk (=1); RR_n = relative risk of highest level of risk; x = xth level of risk; R_{bi} = baseline risk; MF_x = multiplication factor for patient with RR_x; RR_x = relative risk of xth level of risk; and R_x = recalculated risk for patient with RR_x.

Figure 2 shows the recalculation of risk by means of the weighted mean RR for 3 different levels of hsCRP with their accompanying RR. A generic Figure (for any ordinal novel risk factor) is presented in the online-only Data Supplement.

The Solution for Continuous Risk Factors

Suppose the physician in the case vignette would have ordered a homocysteine (hcy) measurement. Homocysteine is commonly considered as a continuous risk factor for CVD.¹⁵ Compared with dichotomous and ordinal risk factors, recalculation is much more complicated for a continuous risk factor. Usually, the relationship between the continuous risk factor and risk is multiplicative (ie, risk increases by a factor). This multiplicative association precludes simple recalculation

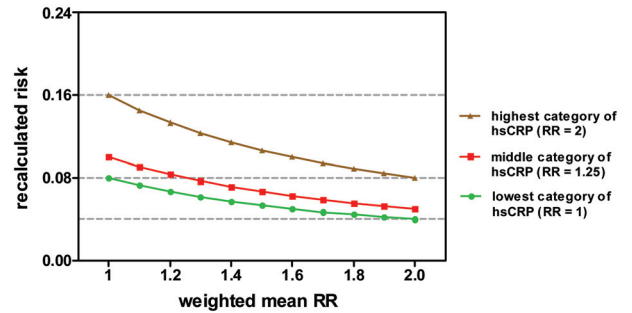


Figure 2. Association between weighted mean RR and recalculated risk for 3 levels (ordinals) of hsCRP. RR indicates relative risk; weighted mean RR indicates a weighted RR of the whole stratum, weighted by prevalence and RR of each of the 3 levels (ordinals) of hsCRP; hsCRP, high-sensitive C-reactive protein. Baseline risk = 0.08 (8%).

of baseline risk. A more thorough description of this problem is presented in the online-only Data Supplement.

Instead, we propose a pragmatic approach, which is to convert the continuous risk factor to risk categories. One could dichotomize the risk factor in low or high value or trichotomize, as in in the example of hsCRP, or make more categories, such as in the example of the coronary calcium score.^{14,16} The recalculation of risk then follows the same pattern as outlined above for dichotomous and ordinal risk factors.

Discussion

We have shown that using additional risk information on top of information in a traditional risk model requires much more than a precise estimate of the independent RR of the novel risk factor. In fact, knowledge of the RR is insufficient without knowing the prevalence of the novel risk factor.

The importance of the distribution of novel risk factors has been reported previously.¹⁶ We have followed up on this by offering a generic and quantitative solution for translating RRs into multiplication factors for baseline risk.

Recently, the American Heart Association listed recommendations for reporting of novel risk markers.¹⁷ These recommendations involved reporting the (independent) RR of the new marker and its statistical significance, accuracy, and discrimination properties. If the novel risk factor is to be used within a new model, these recommendations suffice. However, if the novel risk factor is to be used on top of an existing model, which the American Heart Association recommendations consider appropriate in some situations, knowledge of its prevalence becomes pivotal. However, reporting the prevalence of novel risk markers was not included in the American Heart Association recommendations. We have shown that the impact of prevalence is large, and without knowledge of prevalence recalculation of risk is impossible.

Recalculating using prevalence data involves 2 assumptions that merit discussion. First, we need to know the prevalence of the novel risk factor in the original cohort because this is the cohort we refer to if we estimate baseline risk. It may be possible that the novel risk factor, like MA, has been measured in the original cohort and these data are available. It is more likely, as is true for more laborious and

expensive markers (eg, coronary calcium score), that the novel risk factor was not measured in the original cohort. If this is the case, we could project the prevalence of the novel risk factor in a different cohort onto the original cohort. We would then assume the original cohort (eg, Framingham) to be comparable in all respects to the alternative cohort (eg, the subjects for whom the RR of the novel risk factor has been obtained). Any difference in confounder adjustments or differences in other characteristics could make this projection somewhat inaccurate.

Second, even if the overall prevalence of the novel risk factor in the original cohort is known, it is not precise enough because prevalence presumably will differ across different strata within the cohort. Commonly, prevalence increases when risk is higher because risk factors are often correlated, and with increasing risks the prevalence of the novel risk factor is also likely to increase. Therefore, it is of utmost importance to know the prevalence in each risk stratum (ie, in patients with risk profiles that are comparable to those of patients for whom risk will be re-estimated). A diabetic patient, for example, who smokes and has high blood pressure is more likely to have MA compared with one having none of these risk factors. For our patient, in the stratum with a risk of 8%, knowledge of the prevalence of MA in the whole Framingham cohort would not be sufficiently precise. In fact, we would have to know its prevalence within this stratum.

If distribution of the novel risk factor is not known, one could estimate it by means of regression analysis using classical risk factors as independent predictors and novel risk factor as outcome. This has been done in some instances, like for hsCRP and coronary calcium score, but is certainly not commonly practiced, causing significant lack of clarity on risk-factor distributions in the literature.^{16,18}

Notwithstanding these shortcomings, and even in the absence of information on prevalence, understanding the impact of prevalence is important. Indeed, assuming the novel risk factor to be correlated with other risk factors, one should realize that a higher risk accompanies a higher prevalence of the novel risk factor. The higher the prevalence of the novel risk factor, the less the likelihood that presence of the novel risk factor substantially reclassifies upward. Consequently, absence of the novel risk factor reclassifies downward more substantially.

Conversely, in subjects at low risk (and thus an anticipated low prevalence of the novel risk factor) presence of the novel risk factor reclassifies more substantially than its absence. This is close to intuition, which tells us that the expected is less likely to be informative than the unexpected.

Correct recalculation of risk therefore depends on several assumptions that are difficult to meet in practice. Still, guidelines on cardiovascular risk explicitly encourage the use of specific novel risk factors to improve risk stratification but do not explain how to use this extra information. Usually, novel risk factors are used to place patients in higher categories of risk. The National Cholesterol Education Program–Adult Treatment Panel III (NCEP-ATPIII) guideline, for example, states that for some patients “emerging risk factors might be integrated into ATP III risk assessment ... to

elevate persons ... to the category of CHD risk equivalent.”¹⁰ The Seventh Report of the Joint National Committee on the Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC VII) guideline states that “albuminuria is associated with increased cardiovascular risk,” and the US Preventive Services Task Force considers a novel risk factor as clinically useful and “when assessed in intermediate-risk persons, it should reclassify a substantial portion of them as high-risk.”^{11,12} However, as we have clearly illustrated, upward reclassification is just 1 side of the coin. If added risk information is introduced, upward and downward reclassification in the entire stratum must balance out.

In summary, guidelines stimulate the use of novel risk factors above the well-known risk factors that make up existing risk models. How to use the novel risk factor is less well explained, with disproportionate emphasis on upward reclassification. We show that if a novel risk factor is used to recalculate risk, detailed knowledge of the prevalence of this risk factor, specific to the patients profile, is essential.

Disclosures

None.

References

1. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97:1837–1847.
2. Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetière P, Jousilahti P, Keil U, Njølstad I, Oganov RG, Thomsen T, Tunstall-Pedoe H, Tverdal A, Wedel H, Whincup P, Wilhelmsen L, Graham IM; SCORE project group. Estimation of ten year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J*. 2003;24:987–1003.
3. Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *JAMA*. 2007;297:611–619.
4. Diamond GA. What price perfection? Calibration and discrimination of clinical prediction models. *J Clin Epidemiol*. 1992;45:85–89.
5. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27:157–172.
6. Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. *Ann Intern Med*. 2009;150:795–802.
7. McGeechan K, Macaskill P, Irwig L, Liew G, Wong TY. Assessing new biomarkers and predictive models for use in clinical practice. *Arch Intern Med*. 2008;168:2304–2310.
8. Wallis EJ, Ramsay LE, Jackson PR. Cardiovascular and coronary risk estimation in hypertension management. *Heart*. 2002;88:306–312.
9. Ridker PM, Paynter NP, Rifai N, Gaziano JM, Cook NR. C-reactive protein and parental history improve global cardiovascular risk prediction: the Reynolds Risk Score for men. *Circulation*. 2008;118:2243–2251.
10. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III) final report. *Circulation*. 2002;106:3143–3421.
11. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL Jr, Jones DW, Materson BJ, Oparil S, Wright JT Jr, Roccella EJ; National Heart, Lung, and Blood Institute Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure; National High Blood Pressure Education Program Coordinating Committee. The seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA*. 2003;289:2560–2572.
12. Helfand M, Buckley DI, Freeman M, Fu R, Rogers K, Fleming C, Humphrey LL. Emerging risk factors for coronary heart disease: a

- summary of systematic reviews conducted for the US Preventive Services Task Force. *Ann Intern Med.* 2009;151:496–507.
13. Valmadrid CT, Klein R, Moss SE, Klein BEK. The risk of cardiovascular disease mortality associated with microalbuminuria and gross proteinuria in persons with older-onset diabetes mellitus. *Arch Intern Med.* 2000; 160:1093–1100.
 14. Ridker PM, Wilson PWF, Grundy SM. Should C-reactive protein be added to metabolic syndrome and to assessment of global cardiovascular risk? *Circulation.* 2004;109:2818–2825.
 15. Humphrey LL, Fu R, Rogers K, Freeman M, Helfand M. Homocysteine level and coronary heart disease incidence: a systematic review and meta-analysis. *Mayo Clin Proc.* 2008;83:1203–1212.
 16. Pletcher MJ, Tice JA, Pignone M, McCulloch C, Callister TQ, Browner WS. What does my patient's coronary artery calcium score mean? Combining information from the coronary artery calcium score with information from conventional risk factors to estimate coronary heart disease risk. *BMC Med.* 2004;24:31–42.
 17. Hlatky MA, Greenland P, Arnett DK, Ballantyne CM, Criqui MH, Elkind MS, Go AS, Harrell FE Jr, Hong Y, Howard BV, Howard VJ, Hsue PY, Kramer CM, McConnell JP, Normand SL, O'Donnell CJ, Smith SC Jr, Wilson PW; American Heart Association Expert Panel on Subclinical Atherosclerotic Diseases and Emerging Risk Factors and the Stroke Council. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation.* 2009;119:2408–2416.
 18. Wong ND, Pio J, Valencia R, Thakal G. Distribution of C-reactive protein and its relation to risk factors and coronary heart disease risk estimation in the National Health and Nutrition Examination Survey (NHANES) III. *Prev Cardiol.* 2001;4:109–114.

KEY WORDS: risk factors ■ epidemiology ■ cardiovascular diseases

SUPPLEMENTAL MATERIAL

Formula 1 provides the mathematical relationships between recalculated risk and prevalence of a dichotomous novel risk factor for any level of relative risk (RR) and prevalence (P).

Formula 1

$$[r] = p \cdot RR + (1-p)$$

$$MF_{(+)} = RR/[r]$$

$$R_{(+)} = R_{bl} \cdot MF_{(+)}$$

$$R_{(-)} = R_{(+)} / RR$$

[r] = weighted mean relative risk; p= prevalence of novel risk factor; RR=relative risk of the novel risk factor; $MF_{(+)}$ = multiplication factor for R_{bl} to obtain $R_{(+)}$; $R_{(+)}$ = recalculated risk in presence of novel risk factor; R_{bl} = baseline risk; $R_{(-)}$ = recalculated risk in absence of novel risk factor

Derivation of formula 1 for a dichotomous risk factor

Imagine a stratum with a baseline risk on a certain outcome. The subjects in this risk stratum can be divided into those with and those without the novel risk factor. The relative risk (RR) of the novel risk factor applies to those subjects who have the novel risk factor. In absence of the novel risk factor, $RR=1$.

Baseline risk of the stratum is the sum of risks of both groups, each contributing with their specific weight. This weight is determined by the product of the prevalence (p) of this risk factor and its accompanying RR. In the presence of the novel risk factor, this weight = $p \cdot RR$. In absence of this risk factor, it will be $(1-p) \cdot 1 = 1-p$.

The sum of both weights is $p \cdot RR + (1-p)$ and can be considered as a weighted mean of relative risks. This 'weighted mean RR' = [r]

The next step is to compare the RR of the novel risk factor to this 'weighted mean RR'. This will give us a multiplication factor (in presence of novel risk factor) for baseline risk.

$$MF_{(+)} = RR/[r]$$

Finally, to obtain recalculated risk, we must multiply baseline risk with this multiplication factor, so:

$$\text{recalculated risk for patient with the novel risk factor} = R_{bl} * MF_{(+)} = R_{(+)}$$

In absence of the novel risk factor, recalculated risk $R_{(-)}$ will be $R_{(+)}/RR$ since the ratio between patients with and without the novel risk factor is the RR.

For example, if baseline risk is 8%, the prevalence of the risk factor is 0.6, and the risk factor has an independent relative risk of 2, then

$$[r] = 0.6 * 2 + (1 - 0.6) = 1.6$$

$$MF_{(+)} = 2 / 1.6 = 1.25$$

$$R_{(+)} = 0.08 * 1.25 = 0.1 = 10\%$$

$$R_{(-)} = 0.1 / 2 = 0.05 = 5\%$$

A graphic representation of how the multiplication factor depends on both relative risk and prevalence of a dichotomous novel risk factor is shown in **figure 1a**.

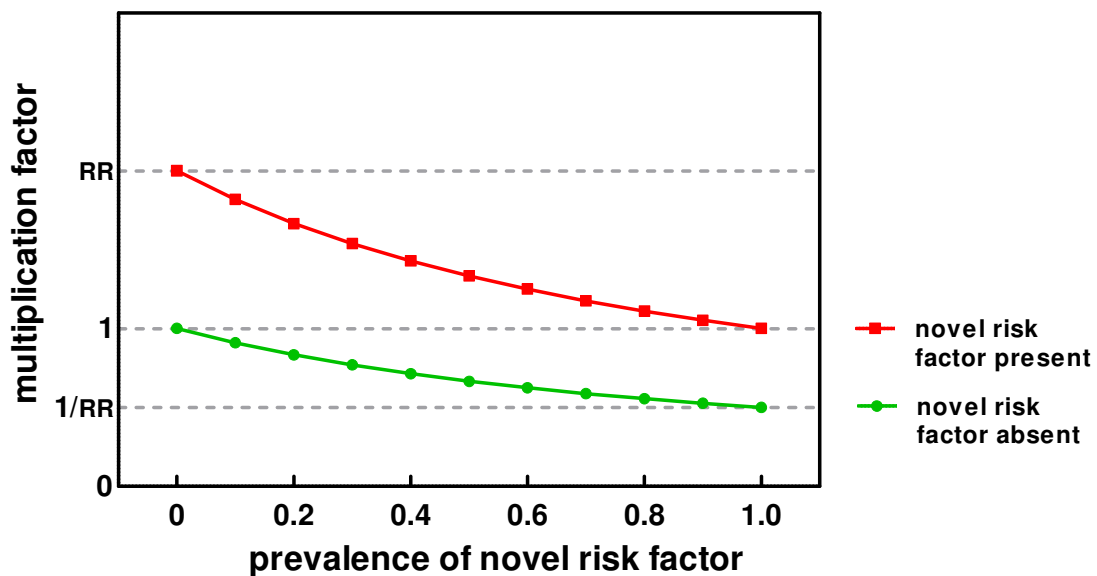


Figure 1a

Association between the prevalence of a novel risk factor and the multiplication factor.

RR indicates relative risk of the novel risk factor.

Formula 2 provides the mathematical relationships between recalculated risk and prevalence of an ordinal novel risk factor for any level of relative risk (RR) and prevalence (P).

Formula 2

$$[r] = p_1 * RR_1 + p_2 * RR_2 \dots\dots\dots + p_n * RR_n$$

$$MF_x = RR_x / [r]$$

$$R_x = R_{bl} * MF_x$$

[r] = weighted mean relative risk; p₁ = prevalence of lowest level of risk; p_n = prevalence of highest level of risk; RR₁ = relative risk of lowest level of risk (= 1); RR_n = relative risk of highest level of risk; x = xth level of risk; MF_x = multiplication factor for patient with RR_x; RR_x = relative risk of xth level of risk; R_x = recalculated risk for patient with RR_x; R_{bl} = baseline risk

Derivation of formula 2

Imagine a stratum with a baseline risk on a certain outcome. Within this stratum the novel risk factor is split into more than 2 possible levels, each with an accompanying RR. Baseline risk of the stratum is the sum of risks of all groups, each contributing with their specific weight. This weight is determined by the product of the prevalence (p) of each risk category and its accompanying RR. For the lowest level of risk, the accompanying RR is 1.0 by definition. The highest level of risk is noted as 'n'. The sum of all weights can be considered as a weighted mean of relative risks [r] , so:

$$\text{Weighted mean RR} = p_1 * RR_1 + p_2 * RR_2 \dots\dots\dots + p_n * RR_n = [r]$$

The next step is to compare the RR of a specific level (RR_x) of the novel risk factor to this 'weighted mean RR'. This will give us a multiplication factor for baseline risk.

$$MF_{(x)} = RR_x / [r]$$

Finally, to obtain recalculated risk, we have to multiply the baseline risk with this multiplication factor,
 so: recalculated risk for patient with level (ordinal) $x = R_{bi} * MF_x$

A graphic representation of how the multiplication factor depends on both relative risk and prevalence of different levels of an ordinal risk factor is shown in **figure 2a**.

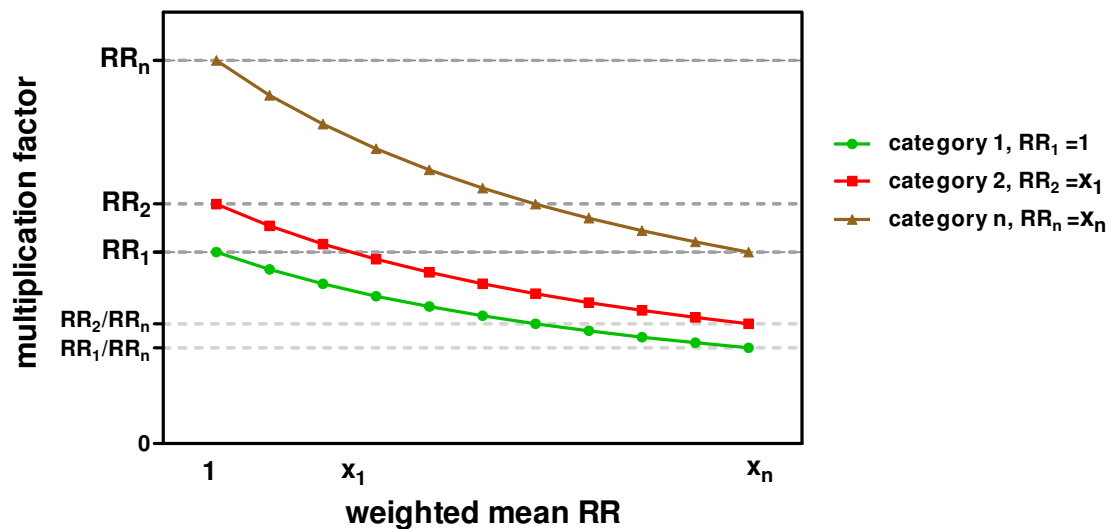


Figure 2a.

Association between weighted mean relative risk and multiplication factor for n levels of a risk factor. RR indicates relative risk of the novel risk factor.

The mathematical problems with continuous risk factors

Recalculation with a continuous risk factor is relatively easy if the relationship between exposure and risk is considered additive i.e. if the relation between the level of the risk factor and risk is represented by a straight line. The baseline risk then corresponds to the mean value of the risk factor, for example homocysteine (hcy), and the linear change in risk is simply the regression coefficient (β). If the hcy-level of a patient is n units different from the mean hcy, the recalculated risk will be: baseline risk + $n\beta$ (**formula 3**).

Formula 3 provides the mathematical relationships between recalculated risk and mean value of a continuous risk factor for any value of the risk factor in a linear model.

Formula 3

$$\text{Risk} = (A - \bar{A}) * \beta + R_{bl}$$

A = value of the continuous risk factor in the patient; \bar{A} = mean value of the risk factor in the stratum of the patient; β = regression coefficients); R_{bl} = baseline risk

Derivation of formula 3

If risk changes linearly with every change of the continuous novel risk factor on the x-axis of 1 unit, then recalculation is carried out by measuring the distance on the x-axis between the level of the risk factor in the patient A and the mean level of this risk factor \bar{A} .

This difference, multiplied with the regression coefficient (β) shows the absolute difference in risk compared to baseline risk. If baseline risk is added, recalculated risk can be obtained.

A graphic representation of the relationship between a continuous risk factor (hcy) and risk in an additive model is shown in **figure 3**.

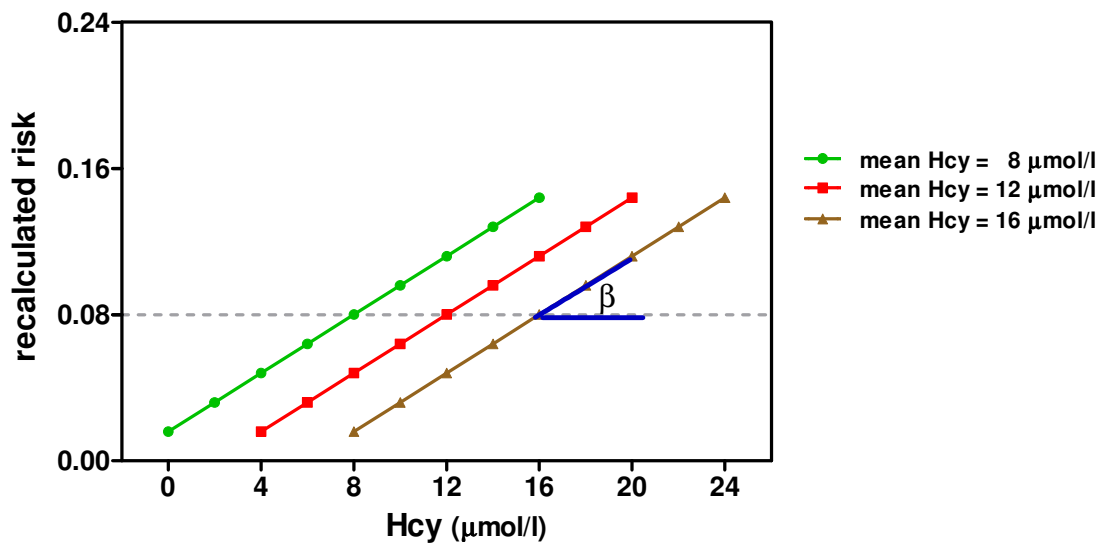


Figure 3 Association between the level of homocysteine of the patient on the x-axis and recalculated risk, for 3 different levels of mean homocysteine in an additive model. Hcy indicates homocysteine level of the patient; 'mean hcy' indicates mean hcy for the stratum with a baseline risk of 0.08; β indicates the regression coefficient ($=0.008$, i.e. if hcy increases 1 $\mu\text{mol/l}$, absolute risk changes 0.008)

However, most continuous risk factors are presented in a multiplicative model, i.e. risk increases by a factor for a given change of the risk factor. The strength of the relation is usually expressed as a relative risk (RR) or odds ratio (OR). This type of relationship is not linear and can graphically be represented as a curve. In a non-linear model, the assumption that the mean value on the y-axis fits the mean value on the x-axis is incorrect. We explain this in **figure 4**.

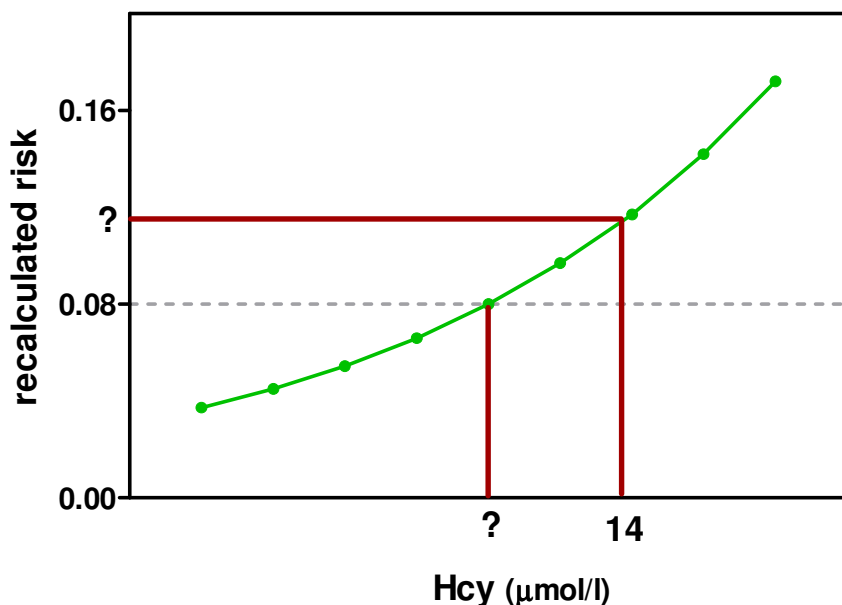


Figure 4 Association between the level of homocysteine and recalculated risk in an multiplicative model for a stratum with a mean hcy of 14 $\mu\text{mol/l}$ and a baseline risk of 0.08 (8%). Hcy indicates homocysteine.

Imagine the level of the continuous risk factor on the x-axis and risk on the y-axis. The baseline risk for our patient is 8%. Going in a straight line from this point to the right, one crosses the curve at a point with an accompanying value on the x-axis. Which point at the x-axis is it? In other words, if, for example homocysteine (hcy) is on the x-axis, which value of hcy fits the baseline risk of 8%?

It is certainly not the mean value of this parameter, it must be lower. Let us imagine hcy to have a normal distribution, so there are as many values higher as there are lower values than mean-hcy. The higher values increase risk more than lower values will bring it down because of the multiplicative nature of the relationship. So the mean value of hcy correlates with a risk on the y-axis which is higher than 8%. In conclusion, the mean value on the x-axis (hcy), does not correspond to the mean value on the y-axis (=baseline risk)

In fact, something very odd happens. Since the baseline risk of 8% corresponds to a hcy-value lower than the mean stratum-specific value of hcy, it follows that having the mean level of hcy is associated with a risk higher than 8%! So, in a multiplicative model, having the mean level of a (normally distributed) determinant increases risk above baseline risk.

Theoretically, a mathematical formula could be constructed to calculate which hcy-value accompanies the baseline risk of 8%. However, things get increasingly difficult if the novel risk factor is not perfectly normally distributed, which will often be the case. In that case one would also need to account for skewness in the formula.

For reasons mentioned above, we propose a pragmatic approach, which is to convert the continuous risk factor to risk categories which is discussed in the main text.