

# In Search of Fewer Independent Risk Factors

Daniel J. Brotman, MD; Esteban Walker, PhD; Michael S. Lauer, MD; Ralph G. O'Brien, PhD

**M**ore than 1100 articles now appear annually investigating “independent risk factors” or “independent predictors” for various clinical outcomes. In medical research, independence is generally defined in a statistical sense: a variable is called an independent risk factor if it has a significant contribution to an outcome in a statistical model that includes established risk factors. As such, independence is based on a specific statistical model and depends on the set of established risk factors included in that model. Even when strong statistical evidence indicates that a variable is an independent risk factor for an outcome, this does not necessarily indicate that the risk factor causally contributes to the outcome. The opposite is also true: risk factors that have causal relationships with the outcome will not necessarily prove to be independent risk factors. These are basic statistical principles that are too often given short shrift in medical research. Herein, we discuss the clinical implications conferred by the above definition of *independence*, primarily using examples from recent cardiovascular literature. A glossary and schema are provided to help clinicians and researchers understand and discuss these matters effectively.

*Arch Intern Med.* 2005;165:138-145

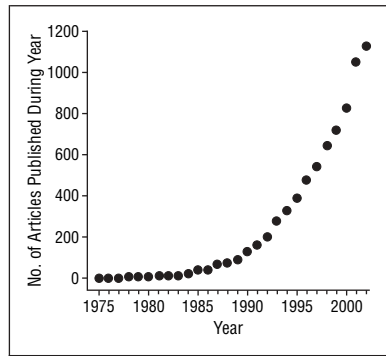
“The purpose of medicine is to prevent disease and prolong life.”<sup>1</sup> This can only be achieved when factors that cause disease are identified and prevented or treated. Well-controlled experimentation (manipulation of a particular factor while holding others constant) is the best way to determine whether a variable causally contributes to an outcome, but for practical and ethical reasons, it is usually impossible to hold biological variables constant in human research. Therefore, multivariable statistical models are used to assess possible causal factors for disease, permitting estimation of the unique effects of a particular variable on the outcome while (statistically) holding other variables constant. Such models help identify risk factors for disease, allow creation of clinical prediction tools, and of-

ten offer insights into pathophysiologic conditions. However, causation can never be definitively discerned from nonexperimental studies, regardless of the statistical tools used.

The accepted way to analyze relationships between multiple potential risk factors and an outcome is, if possible, to include all relevant risk factors in the statistical model to determine the adjusted effect of each risk factor on the outcome. If, after adjustment, a risk factor maintains a statistically significant association with the outcome, it is called an *independent risk factor* for the outcome. Based on this definition, independence (1) is a purely statistical concept, (2) depends on the unique set of variables included in the model, and (3) does not imply causality. Despite this, many clinicians believe that statistical independence strongly suggests causality or at least makes a risk factor more relevant. For example, a review article titled “Evidence That Triglycerides Are an Independent Coronary

**Author Affiliations:** Departments of General Internal Medicine (Dr Brotman), Biostatistics and Epidemiology (Drs O'Brien and Walker), Vascular Surgery (Dr Walker), and Cardiovascular Medicine (Dr Lauer), Cleveland Clinic Foundation, Cleveland, Ohio.

**Financial Disclosure:** None.



**Figure 1.** Increase in the number of articles per year that contain the terms *independent risk factor* or *independent predictor* in their title or abstract.

Heart Disease Risk Factor”<sup>2</sup> begins with the following statement:

In the past, the relation between hypertriglyceridemia and coronary heart disease (CHD) has been uncertain. However, a recent multivariate analysis of 8-year follow-up data from the large-scale Prospective Cardiovascular Munster study found hypertriglyceridemia to be an independent risk factor for major coronary events after controlling for low-density lipoprotein (LDL) and high-density lipoprotein (HDL) cholesterol.

The abstract concludes with the statement “To achieve the greatest possible reduction in CHD risk, antihyperlipidemic treatment strategies should also be aimed at reducing elevated triglycerides.”<sup>2</sup> The author also cites clinical trials demonstrating reduced CHD events in patients with hypertriglyceridemia treated with lipid-lowering medications. However, the title and abstract appear to convey the all-too-familiar assumption that knowing whether a risk factor is an independent risk factor is important in assessing its importance and whether to treat it.

The scope of this issue was revealed in a simple MEDLINE search of the number of articles per year that contained the terms *independent risk factor* or *independent predictor* in their title or abstract. This has increased steadily from 2 in 1975 (earliest occurrence) to 1129 in 2002 (**Figure 1**).

In the present article, we discuss the pitfalls of assuming that independent risk factors causally affect clinical outcomes and that nonindependent risk factors are not causal. We emphasize that statistical models are useful for *predicting* (estimat-

**Table 1. A Glossary of Terms for Risk Factor Research\***

Term	Definition
Risk factor	A variable with a significant statistical association with a clinical outcome.
Independent risk factor	A risk factor that retains its statistical association with the outcome when other established risk factors for the outcome are included in a statistical model. (Also referred to as an independent predictor.)
Nonindependent risk factor	A risk factor that loses its statistical association with the outcome when other established risk factors for the outcome are included in a statistical model.
Causal factor	A risk factor that has a causal relationship with a clinical outcome. There are 2 types: Direct causal factor: a causal factor that directly impacts the outcome (or the likelihood of the outcome); Indirect causal factor: a causal factor that impacts the outcome (or affects its likelihood of occurrence) by changing a direct causal factor. If the direct causal factor is prevented from changing, then changes in the outcome will not be produced. Causal factors are not defined statistically but are defined experimentally in that they are known to affect the outcome.
Noncausal risk factor	A variable that is associated with a causal factor but does not have a causal relationship with the outcome. However, by virtue of its association with a causal factor, a noncausal risk factor is indirectly associated with (and predictive for) the outcome. Others have referred to noncausal risk factors as “proxy” risk factors. <sup>3</sup>
Therapeutic target	A causal factor that can be modified with therapeutic intervention to improve or reduce the likelihood of the clinical outcome.
Therapeutic marker	A variable that predicts relative responsiveness to a particular intervention. Therapeutic markers need not be risk factors for the outcome and are often not modifiable. For example, responsiveness to a particular therapy may vary by genotype, sex, or race, even if none of these variables is a risk factor for the outcome.

\*Not all terms are exclusive; clinical variables may be defined by more than 1 of these terms.

ing the likelihood of) a particular disease or outcome and are sometimes also useful in *suggesting* possible causes, but they are never useful for *ascertaining* causation. For reference purposes, **Table 1** defines the terms we use. Types of risk factors are depicted graphically in **Figure 2**.

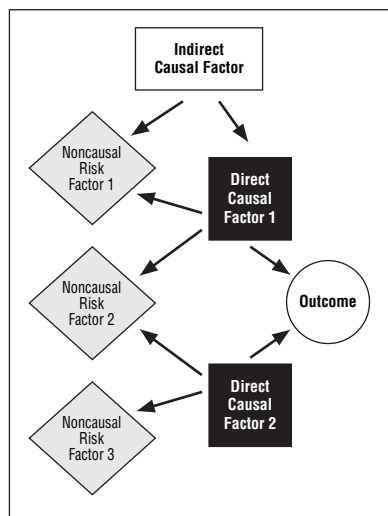
Recognizing that others have generated terminology to describe many of these concepts,<sup>3,4</sup> we have adopted some of their nomenclature. However, the definition of independent risk factor used by Kraemer and colleagues<sup>3</sup> requires that the defined risk factor be uncorrelated with other risk factors. Since most cardiovascular risk factors correlate with other risk factors,<sup>5-10</sup> this stringent criterion has not been adopted in the cardiovascular literature. Additionally, it is worth noting that the term *independent* (referring to a risk factor) as defined herein and as used in current literature is completely different from the statistical term *independence*, which implies an ab-

sence of association between variables (eg, the  $\chi^2$  test for independence).

#### INDEPENDENCE DEPENDS ON THE VARIABLES INCLUDED IN THE MODEL

Gray hair is undoubtedly a risk factor for myocardial infarction (MI): on average, a gray-haired person is more likely to suffer an MI than a person without gray hair. However, if 2 men are identical in age but only 1 has gray hair, it is unlikely that the gray-haired man is at increased risk for MI. Thus, hair color is probably not an independent risk factor for MI when age is considered. If, however, our patients' ages are unknown, then hair color may indeed be an independent risk factor for MI in a model that controls for blood pressure and dyslipidemia. In this case, hair color is a noncausal risk factor serving as a proxy for age.

This phenomenon can work in subtler ways: plasminogen activa-



**Figure 2.** Arrows denote causation in this idealized example to illustrate the distinctions between 3 types of risk factors. Bidirectional arrows, which may exist in real life, are not included, and this example does not include all possible relationships. In medical research, it may be impossible to discern whether a particular risk factor is a direct causal risk factor, a noncausal risk factor, or an indirect causal risk factor or if it falls into more than 1 category.

tor inhibitor-1 (PAI-1) activity is associated with impaired fibrinolysis, obesity, and cardiac disease.<sup>11</sup> Thogersen et al<sup>12</sup> labeled it an independent risk factor for first MI after controlling for established risk factors including diabetes and body mass index. However, others concluded that PAI-1 activity was not an independent risk factor for MI because the association was negated after controlling for insulin resistance.<sup>13</sup> When 2 risk factors are closely correlated (as are insulin resistance and PAI-1), 1 or both variables may lose independent predictive ability when included in a model simultaneously, even if either variable alone is a highly significant independent risk factor for the outcome. In this situation, the 2 correlated variables provide similar (overlapping) predictive information<sup>3</sup>; consequently, both variables may lose independence when modeled together.

Conversely, including additional variables in a model may lead to a finding of independence or even turn a variable that is not associated with the outcome at all into an independent risk factor. For example, 1 study found no unadjusted association between aspirin

use and all-cause mortality: death rates over 3.1 years were 4.5% for aspirin users and 4.5% for nonusers.<sup>14</sup> However, after adjusting for age, coronary history, and other variables, the researchers found that aspirin became an independent predictor for mortality owing to greater aspirin use among patients with cardiovascular risk factors.

#### OTHER DETERMINANTS OF INDEPENDENCE

In addition to which predictor variables are used, other factors may affect the independence of a potential risk factor. These include type of outcome variable (continuous, time-to-event, ordinal, or categorical), type of model (parametric vs semiparametric or linear vs nonlinear), transformations of outcomes and predictors, how candidate variables are treated (forced inclusion vs stepwise), and whether variables are considered individually or combined into an index (eg, the Framingham Risk Score). Furthermore, different conclusions regarding a given variable's independence in predicting a given outcome may depend on the number of patients with the risk factor, the number of outcome events, and the type of patients studied.

Regardless of the methods used to define independence, these statistical tools all address the same basic question: If 2 hypothetical patients are identical on all predictor variables included in the model except for a difference in predictor X, would they then have different expected outcomes? If the answer is yes, then X is an independent risk factor for the outcome. We will not discuss further the statistical methods used to define independence but will instead focus on the limitations of independence itself—limitations that apply even when a well-designed and adequately powered study is analyzed using appropriate models that are carefully validated.

#### TO WHAT EXTENT IS CONTROLLING FOR ALL ESTABLISHED RISK FACTORS REALLY POSSIBLE?

In most epidemiologic studies of cardiovascular disease, an effort is made

to control for as many risk factors as possible, but controlling for all risk factors is impossible. For example, Kenchaiah et al<sup>15</sup> recently found obesity to be an independent risk factor for congestive heart failure (CHF) after controlling for age, smoking, alcohol consumption, total cholesterol, valvular heart disease, hypertension, diabetes, left ventricular hypertrophy, and prior MI. They concluded that 11% to 14% of CHF cases were attributable to obesity alone. However, it is potentially misleading to conclude that obesity is an independent risk factor for CHF outside the context of this particular study. Other cardiac risk factors that correlate with obesity were unmeasured, such as leptin,<sup>16</sup> PAI-1,<sup>17</sup> fasting insulin,<sup>18</sup> sleep-disordered breathing,<sup>19</sup> C-reactive protein (CRP),<sup>20</sup> interleukin 6,<sup>21</sup> heart rate variability,<sup>22</sup> urinary catecholamines,<sup>22</sup> triglycerides, HDL cholesterol,<sup>23</sup> and oxidative stress.<sup>24</sup> It is impossible to know if obesity would have predicted CHF had the authors controlled for all of these variables.<sup>25</sup> A future study might examine other variables and conclude that obesity is indeed a risk factor for CHF but not an independent one.

As more physiologic variables are linked to cardiovascular disease, it has become impossible to measure them all (**Table 2**). No study will ever properly model all cardiovascular risk factors to assert that a given variable is truly an independent risk factor for cardiovascular disease. Any claim that variable X is an independent risk factor for a given cardiovascular outcome (except within the context of a particular study) ignores the likelihood of residual confounding—ie, that valuable predictors also associated with X have been excluded, poorly measured, or incorrectly modeled. In some cases, residual confounding may completely undermine the conclusions (eg, observational studies of hormone replacement and use of antioxidant vitamins).<sup>26</sup>

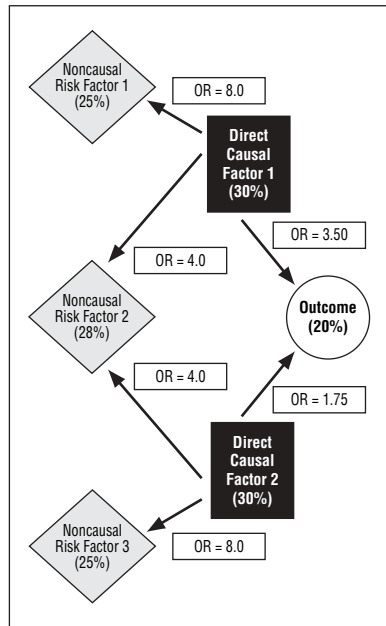
#### NONCAUSAL RISK FACTORS CAN BE INDEPENDENT RISK FACTORS

**Figure 3** shows an idealized structure in which 2 uncorrelated causal

**Table 2. Examples of Risk Factors That Have Achieved “Independence” in Statistical Models for Various Cardiovascular Outcomes After Controlling for Other Cardiovascular Risk Factors\***

Risk Factors	
Demographic factors	Psychological factors
Age	Depression
Sex	Inadequate relaxation
Race	Lack of social support
Family history	Lack of a sense of personal control
Myocardial infarction or premature atherosclerosis	Impaired sleep
Sudden death	Negative emotions
Anthropometric measures	Lifestyle factors
Body mass index	Tobacco abuse
Waist-to-hip ratio	Dietary factors
Anatomic cardiovascular factors	Physical exercise habits
Carotid intima media thickness	Alcohol consumption
Left ventricular dysfunction	Inflammatory factors
Known coronary artery atherosclerosis	Macrophage colony stimulating factor
Cerebrovascular disease	C-reactive protein
Peripheral vascular disease	Interleukin 6
Markers of fluid/electrolyte homeostasis	Lipoprotein-associated phospholipase A <sub>2</sub>
Angiotensin II	Tumor necrosis factor $\alpha$
B-type natriuretic peptide	Presence of connective tissue disease
Atrial natriuretic peptide	Pregnancy-associated plasma protein
Genetic polymorphisms affecting the renin-angiotensin-aldosterone system	Genetic polymorphisms of matrix metalloproteinase
Electrocardiographic (ECG) findings	Myeloperoxidase
Atrial fibrillation	Placental growth factor
Left ventricular hypertrophy	Physiologic parameters
Bundle branch blocks	Systolic blood pressure
Frequent premature ventricular contractions	Diastolic blood pressure
Nonsustained ventricular tachycardia	Arterial pulse pressure
QT-interval dispersion	Diurnal blood pressure variability
Other ECG abnormalities	Aortic pulse-wave velocity
Hemostatic variables	Peak workload with exercise
Fibrinogen	Ventilatory response to exercise
Tissue-type plasminogen activator	Arterial elasticity
Thrombomodulin	Metabolic conditions and markers
von Willebrand factor	Plasma glucose
Plasminogen activator inhibitor-1 (PAI-1)	Insulin resistance and hyperinsulinemia
Factor VII	Diabetes mellitus
Factor VIII	Triglycerides
D-dimer	Nonesterified (free) fatty acid concentrations
Plasmin-antiplasmin complex	High-density lipoprotein cholesterol
Platelet volume	Low-density lipoprotein (LDL) and LDL particle size
Platelet aggregation	Fasting cholesterol
Genetic polymorphisms affecting thrombomodulin, platelet membrane glycoproteins, and PAI-1 activity	Lipoprotein(a)
Aspirin resistance	Leptin
CD40 ligand	Adiponectin
Autonomic nervous system variables	Genetic polymorphisms affecting lipid metabolism
Heart rate variability	Vascular endothelium-related factors
Levels of catecholamines at rest or with exercise	Endothelin
Heart rate recovery after exercise	Endothelium-mediated vasodilatation
Chronotropic response to exercise	Genetic polymorphisms in nitric oxide metabolism
Markers of myocardial strain or damage	Miscellaneous
Cardiac enzymes	Lewis blood group phenotype
Renal parameters	$\gamma$ -Glutamyltransferase
Serum creatinine	Serum calcium
Dialysis-dependent renal failure	Serum potassium
Urinary albumin excretion or proteinuria	Albumin
Pharmacotherapy	Soluble adhesion molecules
Treatment with digoxin	Homocysteine and genetic polymorphisms of homocysteine metabolism
Treatment with antiarrhythmic drugs	Uric acid
Markers of oxidative stress	Serum ferritin
Genetic polymorphisms in glutathione metabolism	Choline
Low activity of glutathione peroxidase 1	Headaches
Advanced oxidation products	Enterolactone
Nitrotyrosine	Sleep-disordered breathing
	Infection with <i>Mycoplasma pneumoniae</i>

\*Citations for this table are available from the corresponding author.



**Figure 3.** Hypothetical model with 2 uncorrelated direct causal factors, 3 noncausal risk factors, and an outcome. All variables are binary. Odds ratios (ORs) define relationships between the variables. Percentages define prevalence rates for the risk factors and the incidence rate for the outcome. For simplicity, bidirectional arrows and interaction terms are not included in this idealized example. Table 3 lists power probabilities for various statistical models applied to this scenario.

factors affect an outcome. The same causal factors also affect 3 noncausal risk factors. **Table 3** lists the statistical powers of various models for testing whether each predictor is an independent risk factor for the outcome.<sup>27,28</sup> If both causal factors are modeled (model B) and there is no measurement error, then noncausal risk factors have no independent predictive ability. But excluding a causal factor gives the corresponding noncausal risk factors independent predictive ability, as in models C and D.

#### EFFECTS OF MEASUREMENT ERROR AND BIOLOGICAL VARIABILITY ON INDEPENDENCE

Suppose now that both causal factors are measured imperfectly in that 10% of patients have their values reversed (Figure 3 and Table 3). Likewise, suppose the noncausal values are reversed 3% of the time. The noncausal risk factors have independent predictive ability in every model, even when all causal factors

are included, as in model B. Since no biological variable is ever measured perfectly, statistical independence of noncausal risk factors may sometimes be an artifact of imperfect measurement of causal factors.

Consider the metabolic syndrome as a causal factor for CHD. This syndrome is characterized by a clustering of risk factors for cardiovascular disease, including central obesity, insulin resistance, hypertriglyceridemia,<sup>29</sup> elevated free fatty acid levels,<sup>29</sup> impaired heart rate variability,<sup>30</sup> increased norepinephrine release,<sup>31</sup> hypertension, inflammation,<sup>32,33</sup> endothelial dysfunction,<sup>34</sup> microalbuminuria,<sup>35</sup> hyperleptinemia,<sup>36</sup> left ventricular hypertrophy,<sup>37</sup> carotid intima media thickness,<sup>38</sup> low insulinlike growth factor 1 levels,<sup>39</sup> overactivity of the renin-angiotensin-aldosterone system,<sup>40</sup> hyperuricemia,<sup>41</sup> and altered hemostasis.<sup>29</sup> Perhaps each of these features of the metabolic syndrome is a causal factor for CHD, but it is likely that some, such as hyperuricemia,<sup>42</sup> merely indicate the presence of the metabolic syndrome (and are noncausal risk factors). In other words, if the metabolic syndrome itself is a causal risk factor for CHD, but we cannot measure it perfectly (there is no gold standard test for it), then each variable associated with it—whether causal or not—may be an independent cardiovascular risk factor, as in model C (Table 3).

Measurement error and biological variability are not identical for all variables. Consequently, a noncausal risk factor that can be measured precisely may become independent, while a causal factor may lose independence because of greater measurement error. For example, glycosylated hemoglobin is probably not a causal factor for CHD, whereas hyperglycemia probably is. Nevertheless, a single glycosylated hemoglobin value may offer greater predictive ability than a random glucose measurement because a glucose level fluctuates widely and that of glycosylated hemoglobin does not. If both glycosylated hemoglobin and random glucose levels are included in a model to predict CHD, it would not be surprising for glycosylated hemoglobin to be an independent risk factor, while random glucose was no

longer predictive after adjustment for glycosylated hemoglobin. If there were a good way to measure glucose values continuously, then using actual glucose values might be ideal for assessing glycemia. However, cost and feasibility should be considered when creating any clinically useful model, and noncausal risk factors, such as glycosylated hemoglobin level, may be more practical to measure than the corresponding causal factors.

#### INDIRECT CAUSAL FACTORS MAY BE NONINDEPENDENT RISK FACTORS

An indirect causal factor may lose independence when all direct causal factors for the outcome are included in the model (Figure 2). For example, obesity leads to many physiologic changes that are probably causal factors for cardiovascular disease, including glucose intolerance, autonomic imbalance, dyslipidemia, hypertension, inflammation, oxidative stress, and abnormal coagulation. Suppose that Kenchaiah et al<sup>15</sup> had controlled for all biochemical and physiologic correlates of obesity, and obesity was no longer an independent risk factor for CHF in their model. Would this imply that obesity does not cause CHF and is not a therapeutic target? Of course not. Weight loss will improve each direct causal factor and is less costly than polypharmacy with antithrombotics, adrenergic blockers, hypoglycemic agents, and lipid-lowering medications. Regardless of whether obesity is an independent risk factor for CHF in a given model—whether 14% or 0% of CHF cases are attributable to obesity alone—obesity is still a causal factor for CHF and a good therapeutic target.

#### THERAPEUTIC MARKERS NEED NOT BE RISK FACTORS

Therapeutic markers help predict the relative efficacies of competing therapies for a given patient. For example, hydroxymethyl glutaryl coenzyme A reductase inhibitors (statins) are most useful in patients with hyperlipidemia or elevated CRP levels.<sup>43</sup> Angiotensin-converting enzyme inhibitors are particularly use-

**Table 3. Calculated Adjusted Odds Ratios (ORs) and Powers Relating Risk Factors to Outcome (All Binary) Based on the Hypothetical Scenario Shown in Figure 3**

Predictive Factor	Measurement Error Absent*		Measurement Error Present†		Comments
	OR	Power‡	OR	Power‡	
<b>Model A</b>					
Causal 1	3.50	>0.99	2.62	>0.99	In the ideal situation, both causal factors (and the outcome) are measured perfectly and the model contains no unnecessary predictors. Therefore, the modeled associations (here, the ORs) are on target, and the power is maximized. Measurement error of the causal factors weakens the associations and reduces power.
Causal 2	1.75	0.98	1.52	0.86	
<b>Model B</b>					
Causal 1	3.50	>0.99	2.36	>0.99	When all causal factors are measured perfectly and included in the model, the noncausal risk factors are null predictors. The modeled associations are on target, but power is reduced. With imperfect measurement, however, the noncausal risk factors become independent risk factors.
Causal 2	1.75	0.93	1.42	0.67	
Noncausal 1	1.00	Null§	1.25	0.33	
Noncausal 2	1.00	Null	1.21	0.26	
Noncausal 3	1.00	Null	1.10	0.09	
<b>Model C</b>					
Noncausal 1	1.69	0.96	1.63	0.93	No causal factors are included. All noncausal variables become independent risk factors for the outcome, even though none is causal.
Noncausal 2	1.53	0.86	1.50	0.82	
Noncausal 3	1.21	0.26	1.20	0.24	
<b>Model D</b>					
Causal 2	1.56	0.80	1.37	0.58	Only 1 causal factor is included, so variables associated with the unmeasured causal factor are independent risk factors, even though they are not causal.
Noncausal 1	1.72	0.96	1.64	0.94	
Noncausal 2	1.38	0.60	1.42	0.69	
Noncausal 3	1.00	Null	1.09	0.08	

\*Idealized models in which the causal factors are measured perfectly.

†Ten percent of causal factor values and 3% of noncausal values are measured opposite their true values, reflecting 1 type of real-life clinical variation and/or imprecision.

‡Power calculations are based on likelihood ratio tests of standard logit models for a sample size of 1500 and  $\alpha = .05$  (2-tailed).<sup>27,28</sup>

§Null predictors have no statistical association with the outcome (OR = 1; null power =  $\alpha$  [.05 in this example]) and can only be identified as independent risk factors via type I error.

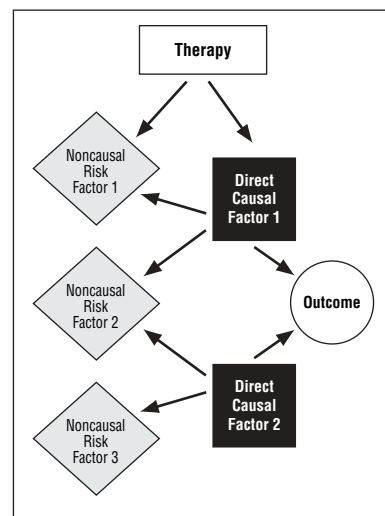
ful in patients with diabetes or microalbuminuria. Therapeutic markers play the role of independent risk factors when the outcome is a favorable response to treatment. Specifically, 2 patients hypothetically identical in every way except for a therapeutic marker will have different expected outcomes when treated identically. However, in the absence of treatment, that same marker might have no prognostic value. If a clinical trial demonstrates that the efficacy of a treatment is related to CRP or demographic variables such as sex<sup>44,45</sup> or race,<sup>46</sup> it does not matter whether CRP, sex, or race is an independent risk factor for the outcome in the absence of treatment.

#### RANDOMIZED CONTROLLED TRIALS CANNOT ALWAYS DISTINGUISH CAUSAL FROM NONCAUSAL RISK FACTORS

Randomization promises to distribute confounding variables evenly

among intervention groups, ensuring that any treatment effect stems from the intervention rather than from known or unknown confounders. When experimentally modifying a risk factor improves an outcome, it is likely that this risk factor is at least indirectly causal for the outcome. However, this may not hold true in small randomized trials because the confounding variables may not be well balanced. Even in the largest trials, when a treatment has multiple (pleiotropic) effects, only some of which are causal (**Figure 4**), it may be challenging to discern whether a risk factor modified by therapy is causal for the outcome. This concept has been overlooked by some authors who have defined causal risk factors based solely on whether manipulation of the risk factor changes the outcome.<sup>4</sup>

For example, since a variety of antihypertensive medications with different mechanisms of action all im-



**Figure 4.** Hypothetical model in which a therapy affects a noncausal risk factor and a causal factor. When multiple clinical variables are favorably affected by an efficacious therapy, it may be impossible to discern which of these variables are causal for the outcome.

prove cardiovascular outcomes, it is likely that hypertension is a causal factor for cardiovascular disease. In

contrast, observing that statins reduce LDL cholesterol levels while reducing cardiovascular mortality<sup>47</sup> does not prove that LDL causes CHD, since statins also affect other cardiovascular risk factors. In the sample model pictured in Figure 4, LDL cholesterol could be non-causal risk factor 1 modified by therapy or direct causal factor 1 also modified by therapy. Since statins affect inflammation,<sup>48</sup> endothelial function,<sup>49</sup> oxidative stress,<sup>50</sup> and coagulation,<sup>48</sup> we cannot conclude that LDL cholesterol is atherogenic based on statin studies alone. This requires a convincing mechanistic explanation and an array of consistent evidence supporting it. Of course, the weight of pathologic and physiologic evidence indicates that LDL cholesterol is important in the formation of atherosclerotic plaques and is indeed a causal factor for atherosclerosis.

## CONCLUSIONS

Spurred by the ease with which investigators can fit multivariable models for almost any outcome, the search for independent risk factors has become commonplace in biomedical research. Such research helps to formulate risk stratification tools and may suggest possible disease mechanisms. Yet causality cannot generally be assessed from observational data sets, regardless of the statistical tools used, including testing for independence. And even in controlled clinical trials, causality of the targeted risk factor cannot be definitively ascertained when the intervention affects other factors.

A variable may appear, correctly, to be an independent risk factor for an outcome in one study, but may appear, equally correctly, not to be so in another study because of different populations, copredictors, or statistical techniques. Independent risk factors in a given study may be clinically or pathophysiologically less relevant than other risk factors that fail to achieve statistical independence. Finally, independence can result from imprecision in biological measurements, risk-factor clustering, and/or the inability to measure all possible risk factors. In summary, terms such as *independ-*

*ent risk factor* and *independent predictor* have meaning only in the context of a particular statistical model; yet these terms are often used casually and in such a way that many physicians and researchers misconstrue them to imply causality.

Accepted for Publication: September 30, 2004.

Correspondence: Daniel J. Brotman, MD, Department of General Internal Medicine/S70, Cleveland Clinic Foundation, 9500 Euclid Ave, Cleveland, OH 44195 (brotmad@ccf.org).

## REFERENCES

1. Mayo W. The aims and ideals of the American Medical Association [Address given at the 66th Annual Meeting of the National Education Association, Minneapolis, July 1-6, 1928]. *Proc Natl Educ Assoc.* 1928;66:158-163.
2. Cullen P. Evidence that triglycerides are an independent coronary heart disease risk factor. *Am J Cardiol.* 2000;86:943-949.
3. Kraemer HC, Stice E, Kazdin A, Offord D, Kupfer D. How do risk factors work together? mediators, moderators, and independent, overlapping, and proxy risk factors. *Am J Psychiatry.* 2001; 158:848-856.
4. Kraemer HC, Kazdin AE, Offord DR, Kessler RC, Jensen PS, Kupfer DJ. Coming to terms with the terms of risk. *Arch Gen Psychiatry.* 1997;54: 337-343.
5. Osterud B, Elvevoll EO, Brox J, et al. Haemostatic parameters related to lipids and adhesion molecules. *Blood Coagul Fibrinolysis.* 1999; 10:465-470.
6. Jensen-Urstad K, Johansson J, Jensen-Urstad M. Vascular function correlates with risk factors for cardiovascular disease in a healthy population of 35-year-old subjects. *J Intern Med.* 1997;241: 507-513.
7. Tamakoshi K, Yatsuya H, Kondo T, et al. The metabolic syndrome is associated with elevated circulating C-reactive protein in healthy reference range, a systemic low-grade inflammatory state. *Int J Obes Relat Metab Disord.* 2003;27:443-449.
8. Felmeden DC, Spencer CG, Chung NA, et al. Relation of thrombogenesis in systemic hypertension to angiogenesis and endothelial damage/dysfunction (a substudy of the Anglo-Scandinavian Cardiac Outcomes Trial [ASCOT]). *Am J Cardiol.* 2003;92:400-405.
9. Mertens I, Van Gaal LF. Obesity, haemostasis and the fibrinolytic system. *Obes Rev.* 2002;3:85-101.
10. Pedrinelli R, Dell'Omo G, Penno G, et al. Microalbuminuria and pulse pressure in hypertensive and atherosclerotic men. *Hypertension.* 2000; 35:48-54.
11. Juhan-Vague I, Morange P, Alessi MC. Fibrinolytic function and coronary risk. *Curr Cardiol Rep.* 1999;1:119-124.
12. Thogersen AM, Jansson JH, Boman K, et al. High plasminogen activator inhibitor and tissue plasminogen activator levels in plasma precede a first acute myocardial infarction in both men and women: evidence for the fibrinolytic system as an independent primary risk factor. *Circulation.* 1998; 98:2241-2247.
13. Juhan-Vague I, Pyke SD, Alessi MC, Jespersen J, Haverkate F, Thompson SG. Fibrinolytic factors and the risk of myocardial infarction or sudden death in patients with angina pectoris: ECAT Study Group: European Concerted Action on Thrombosis and Disabilities. *Circulation.* 1996;94:2057-2063.
14. Gum PA, Thamilarasan M, Watanabe J, Blackstone EH, Lauer MS. Aspirin use and all-cause mortality among patients being evaluated for known or suspected coronary artery disease: a propensity analysis. *JAMA.* 2001;286:1187-1194.
15. Kenchaiah S, Evans JC, Levy D, et al. Obesity and the risk of heart failure. *N Engl J Med.* 2002; 347:305-313.
16. Kennedy A, Gettys TW, Watson P, et al. The metabolic significance of leptin in humans: gender-based differences in relationship to adiposity, insulin sensitivity, and energy expenditure. *J Clin Endocrinol Metab.* 1997;82:1293-1300.
17. Sudi KM, Gallistl S, Trobinger M, et al. The influence of weight loss on fibrinolytic and metabolic parameters in obese children and adolescents. *J Pediatr Endocrinol Metab.* 2001;14:85-94.
18. Smith SR, Lovejoy JC, Greenway F, et al. Contributions of total body fat, abdominal subcutaneous adipose tissue compartments, and visceral adipose tissue to the metabolic complications of obesity. *Metabolism.* 2001;50:425-435.
19. Newman AB, Spiekerman CF, Enright P, et al. Daytime sleepiness predicts mortality and cardiovascular disease in older adults: the Cardiovascular Health Study Research Group. *J Am Geriatr Soc.* 2000;48:115-123.
20. Lemieux I, Pascot A, Prud'homme D, et al. Elevated C-reactive protein: another component of the atherothrombotic profile of abdominal obesity. *Arterioscler Thromb Vasc Biol.* 2001;21:961-967.
21. Vozarova B, Weyer C, Hanson K, Tataranni PA, Bogardus C, Pratley RE. Circulating interleukin-6 in relation to adiposity, insulin action, and insulin secretion. *Obes Res.* 2001;9:414-417.
22. Emdin M, Gastaldelli A, Muscelli E, et al. Hyperinsulinemia and autonomic nervous system dysfunction in obesity: effects of weight loss. *Circulation.* 2001;103:513-519.
23. Daniels SR, Morrison JA, Sprecher DL, Khoury P, Kimball TR. Association of body fat distribution and cardiovascular risk factors in children and adolescents. *Circulation.* 1999;99:541-545.
24. Davi G, Guagnano MT, Ciabattini G, et al. Platelet activation in obese women: role of inflammation and oxidant stress. *JAMA.* 2002;288:2008-2014.
25. Massie BM. Obesity and heart failure: risk factor or mechanism? *N Engl J Med.* 2002;347:358-359.
26. Lawlor DA, Davey Smith G, Kundu D, Bruckdorfer KR, Ebrahim S. Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? *Lancet.* 2004;363:1724-1727.
27. Agresti A. *Categorical Data Analysis.* 2nd ed. New York, NY: John Wiley; 2002.
28. O'Brien R. Using the SAS system to perform power analysis for log-linear models. In: *Proceedings of the 11th Annual SAS Users Group Conference.* Cary, NC: SAS Institute; 1986:778-784.
29. Lormeau B, Arousseau MH, Valensi P, Paries J, Attali JR. Hyperinsulinemia and hypofibrinolysis: effects of short-term optimized glycemic con-

- trol with continuous insulin infusion in type II diabetic patients. *Metabolism*. 1997;46:1074-1079.
30. Colhoun HM, Francis DP, Rubens MB, Underwood SR, Fuller JH. The association of heart-rate variability with cardiovascular risk factors and coronary artery calcification: a study in type 1 diabetic patients and the general population. *Diabetes Care*. 2001;24:1108-1114.
  31. Landsberg L. Pathophysiology of obesity-related hypertension: role of insulin and the sympathetic nervous system. *J Cardiovasc Pharmacol*. 1994; 23:S1-S8.
  32. Piedrola G, Novo E, Escobar F, Garcia-Robles R. White blood cell count and insulin resistance in patients with coronary artery disease. *Ann Endocrinol (Paris)*. 2001;62:7-10.
  33. Festa A, D'Agostino R Jr, Howard G, Mykkanen L, Tracy RP, Haffner SM. Chronic subclinical inflammation as part of the insulin resistance syndrome: the Insulin Resistance Atherosclerosis Study (IRAS). *Circulation*. 2000;102:42-47.
  34. Naidoo DP. The link between microalbuminuria, endothelial dysfunction and cardiovascular disease in diabetes. *Cardiovasc J S Afr*. 2002;13: 194-199.
  35. Festa A, D'Agostino R, Howard G, Mykkanen L, Tracy RP, Haffner SM. Inflammation and microalbuminuria in nondiabetic and type 2 diabetic subjects: The Insulin Resistance Atherosclerosis Study. *Kidney Int*. 2000;58:1703-1710.
  36. Leyva F, Godsland IF, Ghatei M, et al. Hyperleptinemia as a component of a metabolic syndrome of cardiovascular risk. *Arterioscler Thromb Vasc Biol*. 1998;18:928-933.
  37. Davis CL, Kapuku G, Snieder H, Kumar M, Treiber FA. Insulin resistance syndrome and left ventricular mass in healthy young people. *Am J Med Sci*. 2002;324:72-75.
  38. Shinozaki K, Hattori Y, Suzuki M, et al. Insulin resistance as an independent risk factor for carotid artery wall intima media thickening in vasospastic angina. *Arterioscler Thromb Vasc Biol*. 1997; 17:3302-3310.
  39. Marin P, Kvist H, Lindstedt G, Sjostrom L, Bjorn-torp P. Low concentrations of insulin-like growth factor-I in abdominal obesity. *Int J Obes Relat Metab Disord*. 1993;17:83-89.
  40. Weber MA, Smith DH, Neutel JM. Differing roles of body mass and the renin-angiotensin system in mediating the hypertension syndrome. *Am J Nephrol*. 2000;20:169-174.
  41. Vuorinen-Markkola H, Yki-Jarvinen H. Hyperuricemia and insulin resistance. *J Clin Endocrinol Metab*. 1994;78:25-29.
  42. Anker SD, Doehner W, Rauchhaus M, et al. Uric acid and survival in chronic heart failure: validation and application in metabolic, functional, and hemodynamic staging. *Circulation*. 2003;107: 1991-1997.
  43. Albert MA, Staggars J, Chew P, Ridker PM. The pravastatin inflammation CRP evaluation (PRINCE): rationale and design. *Am Heart J*. 2001; 141:893-898.
  44. Woods SE, Noble G, Smith JM, Hasselfeld K. The influence of gender in patients undergoing coronary artery bypass graft surgery: an eight-year prospective hospitalized cohort study. *J Am Coll Surg*. 2003;196:428-434.
  45. Watanabe CT, Maynard C, Ritchie JL. Comparison of short-term outcomes following coronary artery stenting in men versus women. *Am J Cardiol*. 2001;88:848-852.
  46. Gray RJ, Nessim S, Khan SS, Denton T, Matloff JM. Adverse 5-year outcome after coronary artery bypass surgery in blacks. *Arch Intern Med*. 1996;156:769-773.
  47. Kjekshus J, Pedersen TR. Reducing the risk of coronary events: evidence from the Scandinavian Simvastatin Survival Study (4S). *Am J Cardiol*. 1995;76:64C-68C.
  48. Bickel C, Rupprecht HJ, Blankenberg S, et al. Influence of HMG-CoA reductase inhibitors on markers of coagulation, systemic inflammation and soluble cell adhesion. *Int J Cardiol*. 2002;82: 25-31.
  49. O'Driscoll G, Green D, Taylor RR. Simvastatin, an HMG-coenzyme A reductase inhibitor, improves endothelial function within 1 month. *Circulation*. 1997;95:1126-1131.
  50. Delbosc S, Morena M, Djouad F, Ledoucen C, Descomps B, Cristol JP. Statins, 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors, are able to reduce superoxide anion production by NADPH oxidase in THP-1-derived monocytes. *J Cardiovasc Pharmacol*. 2002;40:611-617.

### Correction

**Omission of Financial Disclosure.** In the Original Investigation by De Serres et al titled "Recurrence Risk of Oculorespiratory Syndrome After Influenza Vaccination: Randomized Controlled Trial of Previously Affected Persons," published in the November 8 issue of the ARCHIVES (2004;164:2266-2272), the Financial Disclosure was inadvertently omitted on page 2266. The statement of Financial Disclosure should have appeared beneath the Author Affiliations and read as follows: "Financial Disclosure: Drs De Serres, Skowronski, and Duval have received research grants and Dr Skowronski has received honoraria from Aventis Pasteur and Shire Biologics. Dr Guay has received grants from Wyeth-Ayerst Canada, Biochem Pharma, and Merck Frosst Canada." The journal regrets the error.