Complex relations of genetic polymorphisms with nutritionally influenced biomarkers

Robert J Glynn

The article by Yang et al (1) nicely shows the methodologic challenges in the evaluation of whether a nutritionally influenced biomarker has a causal relation with an important clinical outcome of sufficient magnitude to warrant a public health intervention. We have long recognized the limitations of basing causal inferences entirely on traditional epidemiologic studies, including case-control and prospective cohort designs, because unmeasured characteristics of study participants can also influence both the targeted biomarker as well as the study outcome. Examples include plasma β carotene, vitamin E, and vitamin C, each inversely correlated with development of several major chronic diseases, yet each of uncertain causal significance. Indeed, numerous randomized placebo-controlled trials of vitamin supplements with proven ability to modify those specific biomarkers have largely found no important differences in rates of diseases between subjects randomly assigned to receive active compared with placebo supplements. The time and expense required to complete these trials, which typically have required large sample sizes for reliable estimation of treatment effects, have highlighted the challenges of complete and accurate confounder assessment in traditional epidemiologic approaches and reasonably led investigators to consideration of alternative study designs, including instrumental variable approaches, to evaluate causal relations.

The specific example of homocysteine and its association with cardiovascular disease further shows challenges in causal inference. Initial interest in homocysteine as a cardiovascular risk factor was sparked by observations that individuals with very high concentrations due to homocystinuria, a rare autosomal recessive condition, have markedly elevated rates of both arterial and venous thrombosis (2). A large number of observational studies of homocysteine concentration and risk of cardiovascular disease have generally found positive associations with risk of both stroke and ischemic heart disease. However, a meta-analysis of 30 observational studies found stronger associations in retrospective studies with blood collected after disease onset (3). Concerns that developing disease can influence concentrations of homocysteine indicated that prospective studies might give more reliable estimates. Initially quite strong associations in these pooled prospective data were substantially attenuated when age, sex, cigarette smoking, systolic blood pressure, and cholesterol concentration were controlled for. The final, adjusted estimate indicated that a 25% lower homocysteine concentration was associated with an 11% reduced risk of coronary heart disease (95% CI: 4%, 17%). Such an effect would still be quite meaningful if achievable through simple methods such as fortification of cereals with folic acid or use of supplements. However, additional concerns remain about unmeasured or mismeasured confounding variables. Indeed, the several documented determinants of homocysteine concentration include several that are sometimes unavailable in observational studies or that are measured with error including specific antihypertensive medications, number of cigarettes smoked, coffee and alcohol intake, socioeconomic status, and race-ethnic group (4).

A valuable alternative approach to observational analysis of the relation of a biomarker with a disease capitalizes on the availability of a genetic mutation that influences the biomarker. The term Mendelian randomization derives from the random allocation of alleles at the time of gamete formation. If a polymorphism at a specific locus has a strong relation with the biomarker of interest, and is unrelated to potential confounders of the biomarker-disease association of interest, then the relation of the polymorphism with the disease can be used to obtain an unbiased estimate of the biomarker-disease association. Note the critical assumption that the polymorphism affects the disease solely through the biomarker and not through additional pathways. Under these conditions, the polymorphism serves as an instrumental variable, and analytic strategies developed by econometricians are available to derive unbiased estimates of the biomarker-disease association (5). Katan (6) made the first application of this approach to medicine, with mutations of apolipoprotein E serving as an instrument to evaluate a possibly causal effect of low cholesterol concentrations on the incidence of cancer. Some clear strengths of the approach are that the polymorphism can generally be measured more reliably than the biomarker, and it does not change with time, which removes concerns about regression dilution bias and reverse causation. Indeed, the approach clearly focuses on lifelong determinants of a biomarker, as compared with the limitations of traditional observational studies, which...
evaluate biomarker concentrations within a few years of disease manifestation (and perhaps after subclinical disease has influenced biomarker concentrations), and randomized trials, which typically influence biomarker concentrations for only a few years.

Several hundred studies have examined the association between methylenetetrahydrofolate reductase \((MTHFR)\) C677T polymorphisms and the risk of various manifestations of cardiovascular disease (7, 8). Overall, homocysteine concentrations averaged \(~20\%\) higher with the \(TT\) than with the \(CC\) genotype in the populations studied. Some evidence of publication bias was noted, whereby the estimated relative odds of coronary heart disease in \(TT\) compared with \(CC\) genotypes was significantly weaker in unpublished studies than in published studies, and the overall conclusion was that the estimated OR is consistent with the overall null results of the 10 large trials of folate supplementation (which reduced homocysteine concentrations by \(~25\%)\) that had a summary rate ratio of 1.02 (95% CI: 0.96, 1.02) for major coronary events (7). Also, the pooled prior studies showed considerable evidence of raised average folate concentrations after the introduction of folate supplementation in the late 1990s in the United States, Canada, Australia, and New Zealand. The possible effect modification by time in the relation of \(MTHFR\) C677T polymorphisms with cardiovascular mortality in the study by Yang et al (1) is consistent with a differential response to this folate supplementation by polymorphism status. Possible changes in the relation of \(MTHFR\) C677T polymorphisms with cardiovascular death over time are also consistent with observed modification of the effect of these polymorphisms with risk of stroke by population dietary folate intake in meta-analyses (8). The absence of a consistent effect of \(MTHFR\) C677T polymorphisms on risks of cardiovascular diseases over time challenges the implementation of instrumental variable approaches to estimate the causal effect of folate intake in meta-analyses (8). The absence of a consistent effect of \(MTHFR\) C677T polymorphisms with cardiovascular disease (7, 8). Overall, homocysteine concentrations averaged \(~20\%\) higher with the \(TT\) than with the \(CC\) genotype in the populations studied. Some evidence of publication bias was noted, whereby the estimated relative odds of coronary heart disease in \(TT\) compared with \(CC\) genotypes was significantly weaker in unpublished studies than in published studies, and the overall conclusion was that the estimated OR is consistent with the overall null results of the 10 large trials of folate supplementation (which reduced homocysteine concentrations by \(~25\%)\) that had a summary rate ratio of 1.02 (95% CI: 0.96, 1.02) for major coronary events (7). Also, the pooled prior studies showed considerable evidence of raised average folate concentrations after the introduction of folate supplementation in the late 1990s in the United States, Canada, Australia, and New Zealand. The possible effect modification by time in the relation of \(MTHFR\) C677T polymorphisms with cardiovascular mortality in the study by Yang et al (1) is consistent with a differential response to this folate supplementation by polymorphism status. Possible changes in the relation of \(MTHFR\) C677T polymorphisms with cardiovascular death over time are also consistent with observed modification of the effect of these polymorphisms with risk of stroke by population dietary folate intake in meta-analyses (8). The absence of a consistent effect of \(MTHFR\) C677T polymorphisms on risks of cardiovascular diseases over time challenges the implementation of instrumental variable approaches to estimate the causal effect of homocysteine on risk.

Yang et al (1) also find strong associations of broad race-ethnic categories with the prevalence of \(TT\) genotype. Because these race-ethnic groups have quite different rates of cardiovascular mortality, this constitutes population stratification (9). Furthermore, because ethnicity is measured crudely in broad categories, and correlated variables such as socioeconomic status are also measured with error, residual confounding likely remains in the adjusted analyses.

The nuanced results of Yang et al provide a cautionary note about simple approaches to causal inference. Even in their setting with a prevalent minor allele and a fairly strong association with the biomarker of interest, inference is complicated because the relation of this polymorphism with disease risk may vary with time and by dietary patterns of populations. Overall, the course of this long-argued topic suggests support for the idea of Rothman and Greenland (10) that causal inference in epidemiology is best viewed as an exercise in estimating effects instead of a criterion-guided process for determining whether or not an effect is present.

The author has received grants to his institution (Brigham and Women’s Hospital) from AstraZeneca and Novartis for clinical trials development, monitoring, and analysis. He has signed a consulting agreement with Merck to give a grand rounds talk.

REFERENCES