Naturally, there is a strong desire to substitute intellectual capital for labor. That is why investigators often try to base causal inference on statistical models.¹

In his 1999 historical review of the often contentious debate over the distinction between mathematical association and causation, statistician David Freedman observed that using observational (nonexperimental) data to make accurate causal inference requires hard work, coupled with an uncompromising willingness to explore multiple explanations for relationships among natural phenomena.¹ He cited the example of John Snow, a London physician who determined in 1855 that cholera was a waterborne infectious disease.

To reach his accurate conclusion, Snow performed a series of studies with a painstaking care that seems astonishing in light of today's pressure to publish findings as quickly as possible when they have commercial or political value. Snow observed the relationship between the timing of sailors' arrivals in London ports and contraction of the disease; described the case study of a man who contracted cholera shortly after occupying the apartment of an infected patient; mapped the locations of the victims, engaging in the meticulous work of tracing deaths to specific regions, apartment houses, and even to particular water pumps; and noted the absence of infection among employees of a local brewery who were permitted to drink the company product and “preferred ale to water.” Finally, he performed statistical analyses of mortality data for the years 1853-1854, comparing customers of the Lambeth water company, which in 1852 had moved its intake pipe upstream to a “relatively pure” water source, with customers of the Southwark and Vauxhall company, which continued to draw its water from the Thames River. History records Snow's striking findings: The cholera death rates per 10,000 houses were 315 for Southwark and Vauxhall customers, 37 for Lambeth customers, and 59 for customers in the rest of London. A compelling case for a waterborne infectious agent was made.¹

What does it take to draw accurate causal inference from observational data? As Freedman observed, the process is complex. “[A]n enormous investment of skill, intelligence and hard work seems to be a requirement. Many convergent lines of evidence must be developed. Natural variation needs to be identified and exploited. Data must be collected. Confounders need to be considered. Alternative explanations have to be exhaustively tested. Above all, the right question needs to be framed.”¹ To those who would argue that sophisticated statistical modeling obviates the need to consider confounding factors and alternative explanations when interpreting observational associations, Freedman's answer was simple: “The technology is relatively easy to use. … However, the appearance of methodological rigor can be deceptive.”¹ More contemporary guidelines for research reporting acknowledge the same understanding: when interpreting associations, multivariate modeling is often appropriate and valuable, but is still potentially vulnerable to the effects of confounding and bias and is therefore no substitute for a randomized design.²,³,⁴

Lessons Learned From Confusing Association With Causation In Chasing Biomarkers

Curtiss and Fairman observed in the July/August 2008 issue of JMCP that “evidence-based” interventions targeted to biomarkers frequently do not produce the end point “outcomes we love,” such as reductions in hospitalization rates or mortality.⁵ Notably, most of the instances cited were characterized by a single common pattern: the usurping of lower-quality evidence based on observational associations with higher-quality evidence garnered from experimental testing of hypothesized causal factors.

For example, the Action in Diabetes and Vascular Disease: Preterax and Diamicron Modified Release Controlled Evaluation (ADVANCE) trial was conducted because observational research had documented associations of blood glucose and HbA1c (A1c) levels with cardiovascular events. Researchers randomized 11,140 patients with type 2 diabetes to standard glucose control or intensive glucose control, targeted to achieve an A1c level of 6.5% or less.⁶ During a median 5 years of follow-up, the patients randomized to the intensive glucose control intervention did not have a significantly lower risk of major macrovascular events (hazard ratio [HR] = 0.94, 95% confidence interval [CI] = 0.84-1.06), cardiovascular mortality (HR = 0.88, 95% CI = 0.74-1.04), or all-cause mortality (HR = 0.93, 95% CI = 0.83-1.06), but did have an increased risk of severe hypoglycemia (2.7% intensive control vs. 1.5% standard control; HR = 1.86, 95% CI = 1.42-2.40).⁶

Similarly, the Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial was initiated because of high-quality prospective observational evidence suggesting that after “adjusting for other risk factors,” each 1% decrease in A1c was associated with a 21% decrease in the risk of diabetes-related mortality, a 14% decrease...
in the risk of all-cause mortality, and a 37% decrease in the risk of microvascular complications. ACCORD investigators randomized 10,251 patients to intensive therapy, targeted to achieve an A1c level below 6.0%, or standard therapy, targeted to achieve a level of 7.0%-7.9%. Contrary to observational evidence, treatment groups did not significantly differ on the primary study outcome, a composite of nonfatal myocardial infarction, nonfatal stroke, or death from cardiovascular causes (HR = 0.90, 95% CI = 0.78-1.04). However, the intensive therapy group experienced significantly higher rates of hypoglycemia requiring medical assistance (10.5% vs. 3.5%, P < 0.001) and weight gain greater than 10 kilograms (27.8% vs. 14.1%, P < 0.001). More importantly, all-cause mortality rates were higher in the intensive-treatment than in the standard-treatment group (HR = 1.22, 95% CI = 1.01-1.46), prompting early termination of the trial by the National Institutes of Health after a mean of 3.5 years of follow-up.

Lessons Learned From the Medicare Health Support Experiment: Higher-Quality Evidence Contradicts Observational Studies of Disease Management

Sometimes the negative effects of misinterpreting observational evidence are only economic. For example, the proponents of widespread adoption of disease management programs by Medicare largely ignored the effects of selection bias (i.e., the attraction of health-conscious consumers to disease management programs) and regression to the mean (i.e., the statistical phenomenon in which higher-cost patients incur lower costs over time without any intervention) when interpreting observational associations between participation in disease management programs and positive outcomes. Disease management proponents were understandably unhappy and outspoken when the Congressional Budget Office announced in October 2004 that the evidence was insufficient to conclude that applying disease management programs (i.e., for congestive heart failure, coronary artery disease, and diabetes) in Medicare would reduce overall health spending. Proponents advocated for randomized trials of disease management interventions by Medicare as the path to show definitively that these interventions would simultaneously improve beneficiary health and reduce medical expenditures. But the evidence proved otherwise. The results of the much-anticipated Medicare Health Support (MHS) experiment, launched with its first randomization in August 2005, produced disappointment when researchers concluded in June 2007 that the cost of the MHS vendor fees had greatly exceeded medical savings, resulting in termination of the project.

Lessons Learned in Treatment Guideline Development

More seriously, the danger of confusing association with causation in establishing treatment guidelines and policies is a lesson learned many times over from the sometimes untoward experiences of patients whose providers relied on published observational research or on treatment guidelines that failed to account for quality of evidence. Today’s reputedly gold-standard protocols, if based on poor-quality evidence, can change unexpectedly tomorrow. Examples of widely respected treatment recommendations that were based on observational research, but subsequently called into question by more rigorous evaluation, include the use of hormone replacement therapy to reduce cardiovascular risk in postmenopausal women, influenza vaccination of persons aged 65 years or older to reduce mortality, and clinical protocols used in the treatment of patients with chronic kidney disease. “Although numerous observers have emphasized that clinical practice guidelines should not be translated into clinical performance measures in the absence of high-grade evidence supporting a relationship between the intervention and outcome,” Himmelfarb observed in 2007, “this is precisely what has transpired in the management of [end-stage renal disease].”

Confusing Association With Causation in Reporting the “Impact” of Medication Adherence on Health Outcomes

New evidence of hazards inherent in relying on observational evidence comes from an analysis conducted by Dormuth et al., who studied the relationship between adherence to statin therapy and positive health outcomes. Such studies of associations between medication adherence and various outcomes—including mortality, hospitalization, and health care expenditures—have become familiar fare to managed care decision makers. Typically consisting of retrospective analyses of administrative claims data, almost always with statistical adjustments for measured confounders, these studies routinely find that higher rates of medication adherence are associated with better outcomes and lower disease-related or all-cause health care costs.

These associations between adherence and improved health outcomes are often cited in assessments of various treatments that increase pharmacy benefit spending, such as newer products with less frequent dosing or fixed-dose single-pill combinations containing drugs available individually as generic products. The objective of this type of assessment is, of course, to suggest that if the payer is willing to adopt the new product, the increased prescription drug expense potentially could be partly or fully offset by reductions in total medical cost. Similarly, a recent “prescription for national healthcare reform” cited associations of adherence with lower total health care cost in making the case that $177 billion could be saved annually in the United States by improving adherence to prescribed medication and reducing medication-taking errors. Studies using observational designs and measuring associations are also cited by proponents of value-based insurance design (VBID) as evidence that copayment reductions, although increasing payer expenditures for prescription drugs, would improve medication adherence, which would presumably lead to health improvements and potentially to medical cost offsets.

These conclusions rely on a key assumption—that associations between medication adherence and positive health care outcomes
represent cause-and-effect relationships that can be replicated through interventions targeted to the assumed causal factor, medication adherence. This key assumption is logically analogous to the hypotheses tested and refuted in the ADVANCE and ACCORD studies that interventions targeted to biomarkers, which are associated with clinical end points, would improve those end points in a cause-and-effect relationship.

### New Evidence of the Hazards of Confusing Association With Causation: Adherence to Statins Reduces Workplace and Motor Vehicle Accidents

Dormuth et al.’s study of medication adherence in 141,086 patients who were prescribed a statin for primary prevention provides new reason to be skeptical about the assumption that associations will consistently translate into effective interventions. Instead of studying the expected physiological outcomes of statin therapy, such as reduced rates of myocardial infarction or lower cardiovascular expenditures, Dormuth et al. assessed outcomes that are physiologically highly unlikely to be related to statin use. After statistically adjusting for confounders in a Cox regression analysis, using a method similar in approach to those used in other adherence studies, Dormuth et al. found that statin-adherent patients were less likely than nonadherent patients to have motor vehicle accidents (HR = 0.75, 95% CI = 0.72-0.79) or workplace accidents (HR = 0.77, 95% CI = 0.74-0.81) or to develop diseases unlikely to be related to the biological effects of a statin (HR = 0.87, 95% CI = 0.86-0.89). Statin-adherent patients were also more likely than nonadherent patients to use preventive screening services (HR = 1.17, 95% CI = 1.15-1.20).

Dormuth et al.’s results provide new support for the hypothesis of the “healthy adherer effect,” a phenomenon “whereby adherence to drug therapy may be a surrogate marker for overall healthy behaviour,” documented by Simpson et al. in a 2006 meta-analysis of 21 studies (46,847 subjects), of which 8 (19,633 subjects) were placebo controlled. Simpson et al. assessed the relationship between mortality and adherence for a variety of medications and disease states, such as beta blockers following myocardial infarction, antiretroviral therapy for human immunodeficiency viral infection, digoxin for heart failure, and statins for hypercholesterolemia. The studies had assessed adherence using a variety of measures that included patient self report, electronic drug monitoring, refill data, clinician estimates, and tablet counts. As expected, Simpson et al. found that good adherence (as compared with poor adherence) to “harmful drug therapy” nearly tripled the odds of mortality (OR = 2.90, 95% CI = 1.04-8.11). Odds of mortality were cut by approximately one-half for patients who were prescribed a statin for primary prevention and had good adherence to “beneficial drug therapy” (OR = 0.55, 95% CI = 0.49-0.62). However, Simpson et al.’s findings strongly suggested that these effects were not entirely attributable to the biological effect of the medication, since good adherence to placebo was associated with approximately the same mortality reduction (OR = 0.56, 95% CI = 0.43-0.74). Similar results were obtained in previous studies of propranolol, amiodarone, and candesartan, all of which found that adherence to either placebo or active drug therapy was associated with reduced risk of mortality.

### Quasi-Experimental Studies of Cost Sharing and Adherence: Better Guidance for Pharmacy Benefit Design

Providing another example of the hazards of confusing association with causation, well-designed quasi-experimental studies conducted in commercially insured populations have consistently shown, contrary to the findings of research conducted with less rigorous cross-sectional designs, that typical copayment increases (approximately $5 to $13) in commercially insured populations produce cost savings, especially for payers cost net of patient copay, with little effect on utilization overall, and modest effect on total prescription drug expenditure and no effect on use of other medical services.

Elasticity (price sensitivity) is remarkably low for prescription drugs, estimated at less than 0.2 in many analyses and just 0.1 in a 2007 panel data analysis of a large commercially insured sample (n = 17,798). Following copayment increases of up to $13, existing users of chronic medications, including antihypertensives and statins, do not discontinue therapy at higher rates compared with those experiencing no copayment change. Results for larger copayment increases are mixed. One study found higher discontinuation rates among users of proton pump inhibitors, antihypertensives, and statins following a change from a $7 to $30 copayment, but another study found price-inelastic response to copayment changes of $15-$25 among those with 2 or more claims for chronic medication in the 3 months prior to the cost-sharing increase.

Greater sensitivity to cost sharing was exhibited in quasi-experimental studies of noncommercially insured, more vulnerable populations, such as elderly Medicaid enrollees with at least 8 claims per year for at least 3 chronic medication classes, Medicaid enrollees with schizophrenia, and commercially insured groups that are subject to extreme and atypical copayment changes (e.g., a $23 increase from a single-tier plan at $7 to a $30 nonpreferred brand copayment). Unfortunately, studies of these atypical groups are often cited by VBID proponents as evidence of the harmful effects of cost sharing in commercially insured populations, despite clear lack of external validity (generalizability) in making the comparison.

The few studies of cost-sharing decreases performed to date in commercially insured populations using quasi-experimental designs similarly suggest little price sensitivity. Karter et al. found that, despite an association between lower cost-sharing level and greater use of glucose testing strips among patients with diabetes, providing free testing strips shifted costs from patients to the payer without significant improvement in adherence to glucose testing protocols. Similarly, Sedjo and Cox, who used a difference-in-difference
design to compare matched cohorts of patients using brand simvastatin versus other brand statin medications before and after simvastatin’s patent expiration in June 2006, found only “modest” differences in medication possession ratio (MPR).54 Among brand simvastatin users (n = 13,319), who had experienced a copayment decrease upon patent expiration (from brand to generic copayment, reductions of up to approximately $20 depending on benefit design), adjusted mean MPR increased by an absolute (percentage point) 0.32%. Among users of other brands (n = 26,569), adjusted mean MPR decreased by 2.02%. The resulting difference of 2.54%, although statistically significant because of the extremely large sample size, represented a clinically unimportant 9.3 additional days of statin therapy per year. Elasticity was estimated at only 0.02, essentially no price sensitivity, for copayment reductions of more than $15.54.

Chernew et al. produced a similar finding in a study of copayment decreases from $5/$25/$45 to $0/$12.50/$22.50 for generic drugs, preferred brand drugs, and nonpreferred brand drugs, respectively; the absolute (percentage point) MPR change for statins was 3.39%, representing a clinically unimportant 12.4 additional days of statin therapy annually.55,56 Results for antihypertensives (angiotensin-converting enzyme inhibitors and angiotensin II receptor blockers), beta blockers, and diabetes drugs were similar at 9.5 to 14.7 days of therapy per year. Although noting a remarkable lack of transparency in the Chernew et al. study report, Fairman and Curtis applied national data to its results and estimated that to achieve these tiny gains in adherence the annual per member per year (PMPY) intervention costs across the entire insured population would be large, $11.53, $9.10, and $18.60, respectively, for antihypertensives, diabetes drugs, and statins.56

Will Current VBID Research Provide Evidence That Payers Can Really Use?

A study currently in process, the MHealthy: Focus on Diabetes trial, is an observational difference-in-difference (interrupted time series) evaluation, comparing employees and dependents of the University of Michigan (n = 2,507), whose 3-tier copayments for selected chronic medications changed on July 1, 2006, from $7/$14/$24 to $0/$12.50/$22.50 for generic drugs, preferred brand drugs, and nonpreferred brand drugs, respectively; the absolute (percentage point) MPR change for statins was 3.39%, representing a clinically unimportant 12.4 additional days of statin therapy annually.55,56 Results for antihypertensives (angiotensin-converting enzyme inhibitors and angiotensin II receptor blockers), beta blockers, and diabetes drugs were similar at 9.5 to 14.7 days of therapy per year. Although noting a remarkable lack of transparency in the Chernew et al. study report, Fairman and Curtis applied national data to its results and estimated that to achieve these tiny gains in adherence the annual per member per year (PMPY) intervention costs across the entire insured population would be large, $11.53, $9.10, and $18.60, respectively, for antihypertensives, diabetes drugs, and statins.56

The MHealthy study analysis plan was published in Implementation Science nearly 3 years after the start of the 30-month intervention. Currently, the analysis plan raises more questions than answers about the adequacy and transparency of this nonrandomized evaluation. First, key characteristics of the study groups, including the prescription drug cost-sharing amounts for the comparison group, the industry sector(s) for the comparison group, and the formular(ies) for both groups, were undisclosed. Also undisclosed were key features of the medical benefit, such as monthly premium; prior authorization requirements; and cost sharing for emergency room, inpatient, physician office visit, and preventive services.

This lack of disclosure is important because the MHealthy authors’ assessment of their design—“any change in the control group values may reflect naturally occurring changes over time…while any change in the [University of Michigan] intervention group will reflect both the same naturally occurring trends, as well as the impact of the value-based co-payment reductions”—may be inaccurate if the study groups are not comparable.58 When study groups in an interrupted time series design are not comparable at baseline, it is difficult to know what is “driving” differences in trend, especially if the groups respond differently to changes introduced simultaneously with the intervention.59 For example, university employees are not necessarily representative of employees in other industries in their responses to information, a particular problem in the MHealthy study because both the intervention and comparison groups received an educational letter “detailing the importance of medication adherence in diabetes” as part of the project’s implementation.57 Generally, the “natural” trends in health care utilization and expenditures for groups with different occupations, education, and income may diverge over time independent of any interventions.

Additionally, the MHealthy study plan indicates that the intervention and comparison groups are served by different pharmacy benefit management companies (PBMs),57 raising serious questions about whether services often routinely provided to clients by PBMs, such as newsletters, availability of mail order pharmacy, or other policies and services that are designed to influence prescription drug use, differ in ways that will affect study results. The use of different PBMs also amplifies concerns about the absence of formulary information in the analysis plan. Formulary differences in tier placement of drugs are critically important, and formulary differences can also include tier placement for newly approved brand drugs (tier 2, tier 3, or not covered), influencing trend change separate from the effect of the absolute copayment amounts. An additional confounder is the dissolution of the study’s managed care organization in a merger that took place in December 2007, 12 months before the study end date.59 Whether the intervention and comparison group employers made different systematic choices for their employees in response to this major change is both undisclosed in the MHealthy analysis plan and a potentially crucial determinant of between-group differences in trend.
Second, the generic dispensing ratio, a critically important measure of whether the copayment reduction influences members to use higher-cost brand medications in lieu of less expensive and therapeutically equivalent generic medications, is not listed as an outcome measure in the MHealthy study analysis plan. In previous quasi-experimental research, groups experiencing copayment increases were more likely to increase use of formulary brands and generic medications than were those experiencing no copayment change. Whether copayment decreases prompt the reverse response, cost-ineffective use, should have been an outcome measure in this study. The investigators do mention, as the sixth limitation described in the analysis plan report, an assessment of the “extent of tier-shifting” from lower- to higher-tier drugs as “an empirical issue that we will explore,” but without describing any specific measures.

Third, the financial disclosures in the study implementation report do not mention that the Center for Value-Based Insurance Design at the University of Michigan, which employs 4 of the MHealthy investigators as faculty, is supported by 7 pharmaceutical manufacturers. Notably, the MHealthy financial disclosures also do not match those provided previously by several of the study authors in an April 2008 letter to the editor of JMCP on the topic of VBiD. Thorough and accurate disclosure of financial relationships and other potential conflicts of interest would help readers and decision makers interpret the VBiD and MHealthy study findings.

Illuminating Results From “Plausibility Calculators” for Medication Adherence Interventions

For the decision maker who seeks objective information in a world that is often long on claims of success and short on high-quality evidence, “plausibility calculators” are highly valuable tools, designed to help decision makers model the potential medical cost savings that could result from adherence promotion efforts. Following a concept originally advanced by the Disease Management Purchasing Consortium, plausibility calculators rely upon algorithms derived from published randomized controlled trials of the relationship between use of medications to treat chronic disease, such as statins and antidiabetic drugs, and adverse outcomes such as disease-related hospitalizations and emergency room visits.

Plausibility calculators for disease management and VBiD are available online free of charge. Key assumptions, such as copayment reduction amounts for VBiD programs and engagement rates (the percentage of patients who will be contacted) for disease management programs, are entered by users. Users can either enter hospitalization and emergency room utilization rates for their specific population or apply the rates pre-coded into the calculators based on published and nonpublished evidence. Instead of producing a single point estimate, the calculators produce ranges for various levels of key factors, such as adherence rate change, representing “what if” scenarios.

The results of plausibility calculators typically impose a sobering reduction on the sometimes overly enthusiastic estimates of medical cost offsets that are made by proponents of investments in disease management or reduction of prescription drug copayments. Generally, results suggest that to produce overall cost savings, interventions intended to promote adherence should (a) target only patients in whom previous high-quality research has demonstrated high risk of high-cost adverse events, and (b) provide copayment reductions solely or primarily for generic medications.

Association and Causation: Recommendations for Managed Care

What does evidence about association versus causation mean for managed care decision makers today? First, it strongly suggests that the purported outcomes of interventions to reduce or offset medical expenditures by increasing medication adherence should be viewed with healthy skepticism from a “caveat emptor” perspective if those interventions are supported primarily by observational data. To understand the potential harm in interpreting predictors as if they represented causal factors, an example presented by statistician Jane Miller is helpful: White hair and mortality rates may be highly correlated, “but that does not make white hair a cause of high mortality.” Just as managed care decision makers should not invest in hair coloration products to reduce mortality or in statin adherence promotion to reduce automobile accidents, they should not adopt medication adherence interventions that are based on low-quality observational data. Decision makers should judiciously target interventions to improve medication adherence in the high-risk patients who are most likely to benefit, using the most cost-effective medication to achieve the therapeutic goal and keeping in mind that the findings of “healthy adherer” studies suggest a limitation on the expected outcomes of interventions targeted to medication adherence.

Second, we should be reminded—again—of the importance of establishing a base of high-quality evidence for making decisions that affect the cost and quality of health care. Efforts to set systemic policy based primarily on observational evidence, such as those currently being made by proponents of the application of VBiD to health care reform and Medicare Part D should be addressed in a manner similar to what Centers for Medicare & Medicaid Services applied in the MHS—rigorous experimental testing and provision for early termination in the event of program failure. Confident projections should be replaced by evidence that is based on a randomized study design and reported transparently with complete financial disclosure by the authors. To do otherwise is to risk investing precious resources in interventions that, when put to the test of real-world use, do not work.
Still Looking for Health Outcomes in All the Wrong Places?
Misinterpreted Observational Evidence, Medication Adherence Promotion, and Value-Based Insurance Design

Authors

KATHLEEN A. FAIRMAN, MA, is Associate Editor and Senior Methodology Reviewer, and FREDERIC R. CURTISS, PhD, RPh, CEBs, is Editor-in-Chief, Journal of Managed Care Pharmacy.

AUTHOR CORRESPONDENCE: Kathleen A. Fairman, MA, Kathleen Fairman LTD, P.O. Box 31278, Phoenix, AZ 85046. Tel.: 602.867.1343; E-mail: kfairman@amcp.org

DISCLOSURES

Fairman discloses that during the past year, she was compensated by CareScientific for performing a database analysis, CareScientific provides, free of charge, the plausibility calculators mentioned in this article. Curtiss reports no conflicts of interest related to the subjects or products discussed in this article and notes that JMCP author attestation guidelines require disclosure of all relationships with individuals or companies that have a commercial or other interest in the subject of a manuscript, regardless of whether authors or editors perceive these relationships to represent real or perceived conflicts of interest.

REFERENCES


