It Depends on What “Mean” Means: Averaging Versus Event Patterning in Analyses of Administrative Claims Data

Kathleen A. Fairman, MA, and Frederic R. Curtiss, PhD, RPh, CEBS

In 1963, the late Bobby Bragan, then manager of the Milwaukee Braves, poked fun at the statistics commonly used by baseball analysts, observing that “according to the percentage people” a person who was standing “with one foot in the oven and one foot in an ice bucket” was, on average, “perfectly comfortable.”¹ Nearly 5 decades later, the problem highlighted by Bragan’s colorful assessment of the limitations of summary statistics has grown exponentially, especially in health care. Bombarded daily with an estimated 34 gigabytes of information in nonwork activities alone,² American health care consumers are exposed to an overwhelming amount of statistical information about any health care condition that they choose to investigate. For example, a Google search on the term “H1N1 vaccine” returns more than 3 million hits in 0.28 seconds. The resulting potential for data overload should prompt us to pay close attention to how we package and summarize these data.

It is perhaps not surprising that a great deal of research in numerous medical specialties focuses on consumers’ perceptions of data about their health risks, often finding that consumers either underestimate their risk or fail to take action based on known risks.^{3–8} For example, in a case-control study of genetic counseling for BRCA1/2 testing in a sample of women with a family history of breast or ovarian cancer, Armstrong et al. (2005) found that 43 of 200 (21.5%) cases (i.e., patients who had undergone genetic counseling) and 21 of 181 (11.6%) controls perceived their lifetime risk of developing breast cancer as high. Risk was perceived as low for 121 of 200 (60.5%) cases versus 145 of 181 (80.1%) controls.³ Similarly, Moore et al. (2010) studied a sample of African-American women seeking care in urban health centers, finding that overweight and normal-weight women reported the same perceived risk of weight-related illnesses; and Holmes et al. (2007) studied auditory symptoms and hearing protection use in adults aged 18 to 27 years, finding that few used hearing protection consistently, although 6% reported hearing loss and 20% reported symptoms including ear pain and tinnitus after exposure to loud noises.^{6,7} Slightly rephrased, perhaps the issue at the heart of all of these studies is to what extent health care consumers believe that summary statistics, such as averages, actually apply to them.

The challenge for managed care, as both a user and purveyor of data and information, is to ensure that data-weary consumers are provided with accurate but comprehensible summaries

of which behaviors and treatments are most likely to benefit instead of harm. This task, already challenging because of the complexity of health care information, has been made even more difficult by the increasing use and promulgation of retrospective observational research designs. That is, when health care outcomes are analyzed retrospectively, the only limits to the methodological options available are the imaginations and budgets of the investigators.⁹ The resulting myriad of measures and outcomes in the published research literature can cause both conflicting findings and bewilderment about what study results really mean. Retrospective analyses of administrative claims data pose several particular challenges in summarizing risks and benefits. A few key examples are discussed in this editorial.

Are Average Laboratory Values Clinically Meaningful?

In this issue of *JMCP*, Menzin et al. describe the results of a retrospective observational analysis of the association between glycemic control and diabetes-related hospitalizations in a sample of 9,887 patients aged 30 years or older with type 1 or type 2 diabetes who were treated in the primary care clinic of a managed care organization.¹⁰ All study subjects had at least 2 claims with a diagnosis of diabetes mellitus (*International Classification of Diseases, Ninth Revision, Clinical Modification* codes 250.xx) and at least 2 hemoglobin A1c tests administered within any 12-month period from January 1, 2002, through December 31, 2006. A diabetes-related hospitalization was defined as a hospital stay in which any of 10 available diagnosis fields on the hospital claim contained any of 16 diagnoses, including not only known microvascular and macrovascular complications (e.g., retinopathy, nephropathy, and ischemic heart disease) but also possibly related conditions as defined by the study authors (e.g., septicemia, urinary tract infection, and electrolyte imbalance).

Notably, the measure of glycemic control used in the study by Menzin et al. was average A1c value—that is, the sum of all A1c values obtained for the patient over a follow-up period from a minimum of 1 to a maximum of 5 years, divided by the number of tests. Because the clinic’s laboratory data file includes only tests ordered by clinic providers, most tests ordered by specialists were not included in the A1c average.

In this use of the *average* per test A1c as the predictor of interest, Menzin et al. departed from some previous work in this topic area, which related glycemic control to medical

utilization by *patterning* events in sequence. For example, a widely cited historical cohort study by Wagner et al. (2001) classified a sample of patients with diabetes (predominantly type 2, mean age 60 years) into cohorts based on whether they had achieved improvements in glycemic control from 1992 through 1993 and sustained those improvements in 1994; subsequent health care utilization was examined from 1994 through 1997.¹¹ Wagner et al. observed a significant association between better glycemic control and lower total health care costs in 1995 through 1997 (but not in 1994); when analyzed by subgroup, these savings were statistically significant only for those with baseline A1c levels of at least 10%. Rates of primary care use were lower for the patients with improved glycemic control in all follow-up years, but hospital utilization did not significantly differ by cohort in any follow-up year.¹¹ Similarly, Gilmer et al. (2005) found in a study of patients with type 1 or type 2 diabetes that higher baseline A1c, measured at the start of a follow-up period of up to 3 years, predicted higher *subsequent* total health care costs, although comorbid heart disease, depression, and hypertension were more important predictors.¹²

Despite the inconsistency of the method used by Menzin et al. with some previous work, the method of contemporaneous measurement of average A1c as a predictor and health care utilization as an outcome measure has been used in previously published work, including a previous study by Menzin et al.^{13,14} Moreover, methodological limitations aside, studies of this question have generally concluded that, primarily among patients with very high A1c levels (e.g., at least 10%), improvement in A1c is associated with health care cost reduction.¹⁰⁻¹⁴

In discussing the limitations of their work, Menzin et al. appropriately note that “calculating a mean of all observed A1c values to classify patients is not sensitive to changes over time in A1c levels” and suggest the use of alternative approaches in future research. Although accurate, this statement perhaps does not go far enough to describe the impact of using an average laboratory test value. In the study by Menzin et al., hospitalizations were measured from the date of the first A1c test until the earlier of plan disenrollment, death, or the study end date. It is possible that for an unknown number of study patients, the hospital stay(s) identified as outcomes preceded all but 1 of the A1c measures that were used to predict them, and the mean (SD) number of A1c tests for the full sample was 7.6 (4.3). Thus, this analytic approach violates a key factor in the so-called “Bradford Hill’s considerations” in determining the likelihood that an association is attributable to a causal relationship: “the factor must precede the outcome it is assumed to affect.”¹⁵

Averaging in Measures of Medication Possession Ratio and Cost Sharing

Studies of average laboratory values are not the only examples of problems in summary measures that are frequently used in

analyses of administrative claims data. One of the most common measures, medication possession ratio (MPR), is broadly defined as a ratio of medication availability (supply) to time spent in treatment or intended treatment. However, in practice MPR has been operationalized in numerous ways, sometimes to the detriment of the applicability of study findings to typical clinical practice.

For example, in some previous work, the numerator of the MPR has been defined as a simple sum of all days supply for all medications dispensed during the time period of interest. For patients taking more than 1 drug within a therapeutic class—including those switching from one drug or strength to another or receiving augmentation of therapy with additional drug(s)—the days supply values for all drugs were summed;¹⁶⁻¹⁸ reports of studies in which this method was used often indicate that MPR was truncated at 1.0 (100%) because the resulting summed days of therapy exceeded the number of calendar days studied.^{16,17} When assessing patients making product or strength switches, this calculation is particularly problematic because the drug supply remaining after a switch date is unlikely to be consumed. For example, if a patient is initially dispensed a 30-day supply of Drug A and switches to Drug B after 7 days of treatment, a 23-day supply of Drug A is unconsumed and therefore should not be counted towards MPR. To account for this problem, better designs use a more complicated calculation, the assessment of days covered (sometimes called an assessment of medication gaps), in which (a) calendar days with and without medication are identified by summing the fill date plus days supply for each initial fill and refill of the medication(s) of interest, and (b) days with and without medication are summed across the study period.^{19,20}

Choice of denominator also affects the validity of the MPR calculation. Although it is common to measure the denominator as time from the therapy start date to therapy end date, often defined as the fill date plus days supply for the patient’s final claim,^{16,18} doing so can produce findings that are misleading or confusing.¹⁹ For example, a patient who fills only the initial prescription and has no subsequent refills has a calculated MPR of 100%. MPRs calculated using this approach should be interpreted cautiously if they are used at all. Better—that is, more clinically relevant—designs measure the ratio of days supply to a fixed calendar time frame that represents an intention to treat,¹⁷ such as the first 12 months following the start date of therapy.¹⁹

Another measure that is sometimes misinterpreted, average out-of-pocket cost per prescription or per month of therapy, appears commonly in the literature on patient cost sharing.^{20,21} Often used when specific benefit design information is unavailable to researchers, average out-of-pocket cost is a problematic measure because it is not truly exogenous (independent) but is itself a function of drug selection. Thus, the researcher who uses mean copayment as a predictive measure could be

measuring *either* the effect of the patient's out-of-pocket cost *or* the effect of the medication choice. Systematic bias can occur if, for example, a health plan's pharmacy and therapeutics committee has selected preferred brand drugs based on tolerability or efficacy, with more tolerable or efficacious drugs assigned preferred status (and therefore lower copayment). Similarly, patient out-of-pocket cost is generally lower for generic than brand medications. In a class such as antidiabetic medications, in which generic metformin is both more tolerable and safer than newer brand oral antidiabetic drugs,²²⁻²⁴ a systematic relationship between lower cost-sharing amount and a more favorable drug safety profile is introduced. In both of these situations, an analysis that uses mean copayment as an independent variable is biased to find an association between lower cost-sharing amount and greater persistence with therapy. In better designs, researchers have access to and control for specific benefit design features, including copayment amounts for each medication and other relevant factors, such as formularies, step therapy policies, and prior authorization requirements.

The Tradeoff Between Expediency and Accuracy: How to Choose?

There is no such thing as a perfect research study. All studies have limitations, usually made necessary by an inability to account for every possible source of bias or confounding—although, of course, some studies have more limitations than others. We have argued previously that the most important aspects of any research report are clarity and transparency, which allow readers and decision makers to determine the degree of potential bias and the usefulness of study conclusions for their organizations.²⁵ Still, a primary endpoint outcome measure that is clinically invalid informs no one. How should a researcher choose a measure, and how should a managed care decision maker interpret it?

Perhaps a good rule of thumb is that if one cannot imagine a clinician using an outcomes measure in the way that it is being used in a research study, it is probably not an appropriate measure. For example, try to picture a clinician saying something like this to a patient with type 2 diabetes: “Well, Mrs. Smith, thanks to treatment with metformin, your A1c has declined from 8.0% at your first test, to 7.5% in your second test, and your latest A1c reading was 7.0%. Still, the average of these 3 test values is 7.5%, which puts you at elevated risk of complications. It's time to add a new medication.” Similarly, one cannot picture a clinician saying to a patient: “During the past 3 months, I have started you on 4 different medications and you have stopped taking all 4 after only 1 week. But, when I sum up the days supply of all of your dispensed drug tablets, including those that you consumed and those that are still sitting in your medicine cabinet, and I divide that sum by 90 treatment days, I get a ratio of 133%. So, good job. I'll see you at your checkup next year.”

As in many things in life, the more difficult path is often the better path in analyses of administrative claims data. “Quick and easy” approaches—like measuring independent and dependent variables contemporaneously, calculating average laboratory test values over an entire study period, and assessing MPRs without accounting for realistic clinical patterns such as drug switching and augmentation—may be a means to rapid calculation and dissemination of study findings, but they are also potentially misleading. In another editorial in this issue, we argue that an excess volume of poorly targeted communications to prescribers about drug safety may threaten the degree to which prescribers respond to information about true drug-related risks to their patients.²⁶ Perhaps the same is true of information that is provided to health care consumers and decision makers without sufficient consideration of whether a study design is clinically meaningful. Researchers who use observational study designs should take time to think through the logical connections between realistic clinical scenarios and study methodology to ensure that research results will have real-world applicability. To do otherwise does not serve the needs of clinicians or their patients.

Authors

KATHLEEN A. FAIRMAN, MA, is Associate Editor and Senior Methodology Reviewer and FREDERIC R. CURTISS, PhD, RPh, CEBS, is Editor-in-Chief of the *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

The authors report no conflicts of interest related to the subjects or products discussed in this article.

REFERENCES

1. Quotations about statistics. Quote garden. Available at: <http://www.quote-garden.com/statistics.html>. Accessed April 23, 2010.
2. Bohn RE, Short JE. How much information? 2009 report on American Consumers. December 2009. Available at: http://hmi.ucsd.edu/pdf/HMI_2009_ConsumerReport_Dec9_2009.pdf. Accessed April 21, 2010.
3. Armstrong K, Micco E, Carney A, Stopfer J, Putt M. Racial differences in the use of BRCA1/2 testing among women with a family history of breast or ovarian cancer. *JAMA*. 2005;293(14):1729-36. Available at: <http://jama.ama-assn.org/cgi/reprint/293/14/1729>. Accessed April 21, 2010.
4. Allen JK, Purcell A, Szanton S, Dennison CR. Perceptions of cardiac risk among a low-income urban diabetic population. *J Health Care Poor Underserved*. 2010;21(1):362-70.

5. Licht AS, Murphy JM, Hyland AJ, Fix BV, Hawk LW, Mahoney MC. Is use of the human papillomavirus vaccine among female college students related to human papillomavirus knowledge and risk perception. *Sex Transm Infect.* 2010;86(1):74-78.
6. Moore SE, Harris C, Wimberly Y. Perception of weight and threat to health. *J Natl Med Assoc.* 2010;102(2):119-24.
7. Holmes AE, Widén SE, Erlandsson S, Carver CL, White LL. Perceived hearing status toward noise in young adults. *Am J Audiol.* 2007;16(2):S182-S189.
8. Seale H, Heywood AE, McLaws ML, et al. Why do I need it? I am not at risk! Public perceptions towards the pandemic (H1N1) 2009 vaccine. *BMC Infect Dis.* 2010;10(1):99. [Epub ahead of print]
9. Fairman KA. Differentiating effective data mining from fishing, trapping, and cruelty to numbers. *J Manag Care Pharm.* 2007;13(6):517-27. Available at: <http://www.amcp.org/data/jmcp/pages%20517-27.pdf>.
10. Menzin J, Korn JR, Cohen J, et al. Relationship between glycemic control and diabetes-related hospital costs in patients with type 1 or type 2 diabetes mellitus. *J Manag Care Pharm.* 2010;16(4):264-75. Available at: <http://www.amcp.org/data/jmcp/264-275.pdf>.
11. Wagner EH, Sandhu N, Newton KM, McCulloch DK, Ramsey SD, Grothaus LC. Effect of improved glycemic control on health care costs and utilization. *JAMA.* 2001;285(2):182-89. Available at: <http://jama.ama-assn.org/cgi/reprint/285/2/182>. Accessed April 22, 2010.
12. Gilmer TP, O'Connor PJ, Rush WA et al. Predictors of health care costs in adults with diabetes. *Diabetes Care.* 2005;28(1):59-64. Available at: <http://care.diabetesjournals.org/content/28/1/59.full.pdf+html>. Accessed April 22, 2010.
13. Menzin J, Langley-Hawthorne C, Friedman M, Boulanger L, Cavanaugh R. Potential short-term economic benefits of improved glycemic control: a managed care perspective. *Diabetes Care.* 2001;24(1):51-55. Available at: <http://care.diabetesjournals.org/content/24/1/51.full.pdf+html>. Accessed April 22, 2010.
14. Oglesby AK, Secnik K, Barron J, Al-Zakwani I, Lage MJ. The association between diabetes-related medical costs and glycemic control: a retrospective analysis. *Cost Eff Resour Alloc.* 2006;4:1. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1369002/pdf/1478-7547-4-1.pdf>. Accessed April 22, 2010.
15. Höfler M. The Bradford Hill considerations on causality: a counterfactual perspective. *Emerg Themes Epidemiol.* 2005;2:11. Available at: <http://www.ete-online.com/content/pdf/1742-7622-2-11.pdf>. Accessed April 22, 2010.
16. Robertson TA, Cooke CE, Wang J, Shaya FT, Lee HY. Effect of medication burden on persistent use of lipid-lowering drugs among patients with hypertension. *Am J Manag Care.* 2008;14(11):710-16. Available at: http://www.ajmc.com/media/pdf/AJMC_NovRobertson_710to716.pdf. Accessed April 23, 2010.
17. Yoon J, Ettner SL. Cost-sharing and adherence to antihypertensives for low and high adherers. *Am J Manag Care.* 2009;15(11):833-40. Available at: http://www.ajmc.com/media/pdf/AJMC_09Nov_Yoon_833to840.pdf. Accessed April 23, 2010.
18. Taira DA, Wong KS, Frech-Tamas F, Chung RS. Copayment level and compliance with antihypertensive medication: analysis and policy implications for managed care. *Am J Manag Care.* 2006;12:678-83. Available at: http://www.ajmc.com/media/pdf/AJMC_06novTaira_678to683.pdf. Accessed April 23, 2010.
19. Fairman KA, Motheral BR. Evaluating medication adherence: which measure is right for your program? *J Manag Care Pharm.* 2000;6(6):499-504. Available at: http://www.amcp.org/data/jmcp/ce_v6_499-506.pdf.
20. Ellis JJ, Erickson SR, Stevenson JG, Bernstein SJ, Stiles RA, Fendrick AM. *J Gen Intern Med.* 2004;19:638-45. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1492382/pdf/jgi_30516.pdf. Accessed April 23, 2010.
21. Hunt J, Rozenfeld Y, Shenolikar R. Effect of patient medication cost share on adherence and glycemic control. *Manag Care.* 2009;18(7):47-53.
22. Bolen S, Feldman L, Vassy J, et al. Systematic review: comparative effectiveness and safety of oral medications for type 2 diabetes mellitus. *Ann Intern Med.* 2007;147(6):386-99. Available at: <http://www.annals.org/content/147/6/386.full.pdf+html>. Accessed April 23, 2010.
23. Selvin E, Bolen S, Yeh H, et al. Cardiovascular outcomes in trials of oral diabetes medications: a systematic review. *Arch Intern Med.* 2008;168(19):2070-80. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2765722/pdf/nihms82476.pdf>. Accessed April 23, 2010.
24. Eurich DT, McAlister FA, Blackburn DF, et al. Benefits and harms of antidiabetic agents in patients with diabetes and heart failure: systematic review. *BMJ.* 2007;335(7618):497. [Epub 2007 Aug 30].
25. Fairman KA, Curtiss FR. Rethinking the “whodunnit” approach to assessing the quality of health care research: a call to focus on the evidence in evidence-based practice. *J Manag Care Pharm.* 2008;14(7):661-74. Available at: http://www.amcp.org/data/jmcp/661-674_FairmanCurtiss-Final.pdf.
26. Curtiss FR, Fairman KA. Quality improvement opportunities in prescriber alert programs. *J Manag Care Pharm.* 2010;16(4):292-96. Available at: <http://www.amcp.org/data/jmcp/292-296.pdf>.