DISEASE MANAGEMENT
Volume 6, Number 2, 2003

An Assessment of the Total Population Approach for Evaluating Disease Management Program Effectiveness

ARIEL LINDEN, Dr.P.H., M.S.,[1] JOHN L. ADAMS, Ph.D.,[2] and NANCY ROBERTS, M.P.H.[3]

ABSTRACT

A key challenge currently facing the disease management industry is accurately demonstrating program effectiveness at controlling utilization of services and medical costs of populations with chronic disease. The most common method used in the disease management industry to date for determining financial outcomes is referred to as the "total population approach." This model is a pretest–posttest design, which is a relatively weak research and evaluation technique. This paper describes the "total population approach," details many of the biases and confounding factors that may influence outcomes using this method, and illustrates the potential consequences of these factors.

INTRODUCTION

A KEY CHALLENGE CURRENTLY facing the disease management (DM) industry is accurately demonstrating program effectiveness at controlling utilization of services and medical costs of populations with chronic disease.[1–4] (While a comprehensive evaluation of DM program effectiveness could include additional elements such as quality, continuity, and access to care measures, member and/or provider satisfaction, and health-related productivity metrics, this article will focus on the most common outcome of interest—medical costs.) While evolution in this area has eliminated several sources of measurement error, the most prevalent evaluation techniques are still limited by a variety of biases and confounding factors.[2,5]

The most common method used in the DM industry to date for determining financial outcomes is referred to as the "total population approach." This model is a pretest–posttest design, which is a relatively weak research and evaluation technique.[6–8] The most basic limitation of this design is that there is no control group with which comparisons of outcomes can be made. As a result, there may be several sources of bias and/or competing extraneous confounding factors that offer plausible alternative explanations for any change from baseline. Consequently, it is impossible to conclude that the difference is indeed due to the program's intervention.[6–8]

This paper will describe the "total population approach" used by the DM industry for evaluating program effectiveness, detail many of the biases and confounding factors that may influence outcomes, and illustrate the potential consequences of these factors.

[1]Director of Clinical Quality Improvement and [3]Regional Director, Managed Care Programs, Providence Health Plans, Portland, Oregon.
[2]RAND Corporation, Santa Monica, California.

THE TOTAL POPULATION APPROACH MODEL

Figure 1 illustrates a generalized model typically used for evaluating impact of DM programs on medical utilization and cost. The baseline measurement year usually denotes the 12-month period ending with the month prior to the program launch. Each subsequent measurement period equals the baseline period in duration. To improve measurement accuracy, analysis of each measurement year does not occur until after a complete claims run-out period (usually 3–4 months is sufficient to collect close to 100% of medical and pharmacy claims).

Table 1 provides two illustrations of components consistent with DM program evaluation models. There are two principal approaches to how costs are measured and compared across time periods. One uses the disease-specific member as the unit of analysis and aggregates their incurred costs for the period. The second approach uses specific utilization measures (procedure, test, or hospitalization) as the unit of analysis. In this scenario, the total number of occurrences and their related costs are aggregated to the specific utilization measure.

The types of medical costs included in the evaluation vary depending on the unit of analysis chosen. Evaluations using the disease-specific member as the measurement unit may include disease-specific medical costs (those medical costs directly related to the care of the chronic disease being managed) only or may be expanded to include all medical costs incurred by members identified with the disease. In either case, "medical cost" typically includes only those types of expenses captured in medical claims, and therefore does not include indirect health-related costs such as productivity. This model may also call for the exclusion of certain types of costs. For example, in evaluating a diabetes program all members who also have another specifically defined disease (such as cancer or AIDS) may be removed from the analysis. Likewise, certain types of costs (such as those related to trauma or infertility treatment) may be excluded from the analysis even when incurred by an otherwise eligible member.

The evaluation method may also call for medical costs to be aggregated separately by age and/or sex categories, or by lines of business (eg, commercial, Medicare, Medicaid). For many chronic diseases such subanalyses are prudent because of the possible variations in prevalence of the disease, differences in treatment methods, and costs in these categories.

Finally, costs may be normalized in a variety of ways to ensure that the effect of change in the population size is accounted for. The two most widely used methods are: to spread the

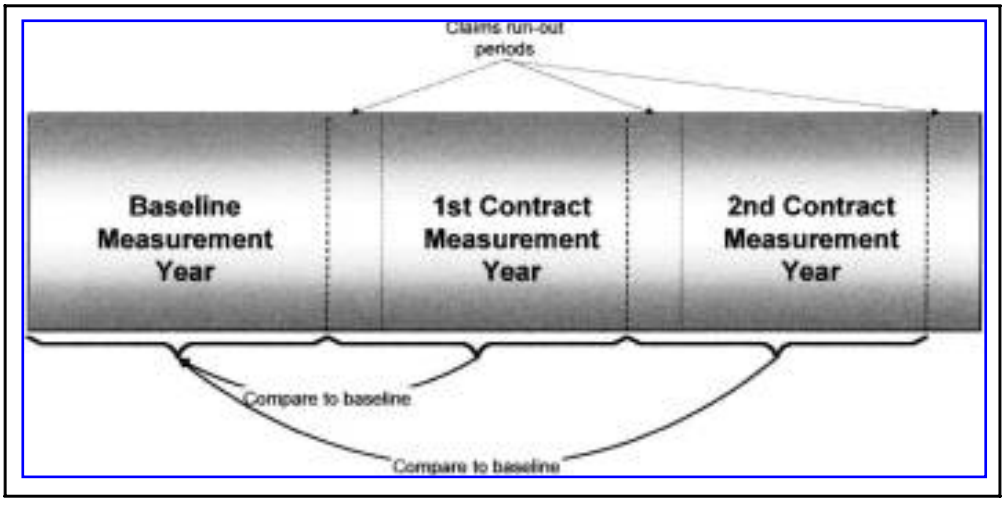

FIG. 1. Conceptual model of the "total population approach" to DM program evaluation. Each contract measurement year is compared with the baseline, after a claims run-out is completed (usually within 3–4 months after the year ends).

Table 1. Illustration of Two Common DM Evaluation Models

Disease of interest	*Unit of analysis*	*Type of costs included*	*Type of costs excluded*	*Subcategories*	*Population denominator*	*Outcome variable*
Diabetes	Members with diabetes	Total cost	Diabetic members with AIDS Claims for cost related to trauma and infertility treatment	Separate analyses by line of business: • Medicare • Commercial • Medicaid	Total membership	PMPY
COPD	Utilization variables Specific COPD-related events: • Diagnostic tests (eg, pulmonary function test) • Procedures (eg, bronchoscopy) • In-patient admissions • Emergency room visits	Specific COPD-related events only Only for such events when accompanied by a diagnosis of COPD	All costs not associated with the specifically defined COPD events All costs for the specifically defined COPD events without an accompanying COPD diagnosis	Separate analyses by age/sex categories	Disease-specific membership only	PDMPY

PMPY, per-member-per-year; PDMPY, per-diseased-member-per-year.

costs over the entire population, or to limit the costs to the disease-specific members. Thus, the "standardized unit" for the final outcome variable may be either per-member-per-year (PMPY), or per-diseased-member-per-year (PDMPY).

THREATS TO VALIDITY OF THE TOTAL POPULATION APPROACH

Irrespective of which component parts are used to develop the outcome variable, or which of these outcome variables is ultimately chosen, the total population approach does not control for the effects of many potential sources of bias or confounding factors.[2,6–8] Bias is any systematic error in collecting or interpreting data, while confounding factors are any additional variables that may influence the results outside of the variables under observation.

Table 2 provides a matrix of confounding factors and biases that may influence the results attained using the total population model. The first column indicates the source of the bias or confounding, the second column specifies the type of bias or confounding (vis-à-vis the nomenclature normally used in the research literature), the third column explains how these factors may be manifested, and the fourth column shows how these variables may affect the results.

The member

DM program interventions typically target the health plan's members identified with or at significant risk for the disease, and attempt to improve their selfmanagement of the disease process using health education, nurse monitoring and coaching, and a variety of other techniques. Therefore, successful interventions are highly dependent on which members choose to enroll. People who are motivated to overcome their disease or physical limitations are more likely to enroll in a DM program than those who are not motivated. Similarly, sicker members may be scared enough by the progression of their disease to prompt them to enroll. Since selfselection poses such a major bias in influencing results, many DM programs are beginning to rely on an enrollment process called the engagement model, in which members are presumptively enrolled and must "opt-out" of the program.[9] Even if 100% of the diseased population were to enroll in the program, certain members may be more apt to disenroll than others. Of course, members may disenroll from the health plan, which would cause their immediate ineligibility in the DM program.

Additionally, chronic illness is progressive, and over time, members will get sicker and eventually die. Therefore, it can be assumed that a diminishing health status will impact costs and certain quality outcome measures.

The aforementioned factors may influence DM program effectiveness through their impact on program enrollment. Health plan benefit design is an additional confounding factor that may influence DM program outcomes. For example, the types of benefits covered, the presence and amount of copay, co-insurance, and deductibles, and the cost-sharing percentages in "tiered" pharmacy plans, all play a role in when and how members access services. Changes in any of these benefit design elements may influence utilization rates of services, unrelated to the impact of DM program interventions.

The health plan or DM program

For most DM programs, demonstrating a reduction in medical costs from baseline to subsequent measurement periods is the ultimate measure of program effectiveness. The cost to treat a disease-specific population from year to year consists of two major components: (1) the type and amount of medical services used (utilization) and (2) the rates paid for those specific services (unit cost). DM program interventions principally affect the utilization of services. With very few exceptions, DM programs have little influence on unit cost changes from year to year. Failure to adequately control for unit cost changes poses a major threat to the accuracy of measuring DM program effectiveness. Unit cost changes may manifest in several ways: (1) general fee schedule increases (eg, provider conversion factor increased by 4%), (2) changes in reimbursement methodology (eg, alterations in the Resource Based Relative Value Scale or a shift from capitation to

Table 2. Threats to Validity in the Total Population Approach

Source	*Type*	*How evidenced?*	*Possible result?*[a]
Member	Selection bias	Motivated members more likely to enroll and achieve desired behavior	+
		Sicker members may enroll due to the "fear factor"	−
		Voluntary enrollment model may enroll motivated members	+
		Engagement model will force members to enroll, including the unmotivated	−
	Loss to attrition	Members may disenroll from the program or health plan, or die	?
	Maturation	Progressive disease, patients will get sicker	−
	Benefit design	Member cost sharing may influence use of health services	?
Health plan/program	Unit cost increases	Costs will appear higher if unadjusted for changes in pricing of services	−
		Reimbursement method and coding changes may alter unit cost	?
	Regression to the mean	High cost members in the 1st year will cost less in the 2nd year	+
		Low cost members in the 1st year will cost more in the 2nd year	−
	Case-mix	Turnover of health plan membership may change population health status	?
	Treatment interference	Members may be exposed to more than one competing intervention	+
	Access to services	Utilization may be impacted by access to and availability of providers	?
	New technology	New technologies may increase some costs and/or reduce others	?
Physician/provider	Hawthorne effect	Changing practice patterns due to being observed	+
	Reimbursement method	Practice patterns may be influenced by risk and reimbursement models	?
Data	Sensitivity/specificity	Imprecise disease identification algorithms may miss suitable members	?
	Missing information	Algorithms may require specific data sources that may be unavailable	−
	Seasonality	Outcomes may be effected if unadjusted for seasonal influences	?
Measurement	Reliability	Can the same results be produced repeatedly?	?
	Validity	Does the outcome measure make sense?	?
General	Secular trends	External factors may mask the impact of the DM program intervention	?

[a]A plus indicates that the remeasurement period may show a better result than the baseline, a minus indicates a worse result than baseline, and a question mark indicates that the result may be difficult to determine.

fee-for-service for primary care), and (3) coding changes (eg, adding a newly approved expensive procedure to an existing code, thereby changing the average cost of that code). Despite the fact that DM programs rarely directly impact unit cost, adjustment for year-to-year unit cost increases is critical to accurately assessing DM program impact on medical cost. Unit cost adjustment methods can vary from quite simple to extraordinarily complex. Examples include local market medical inflation index, using premium inflation as a proxy for unit cost, comparing the actual year-to-year cost changes of a specific limited set of procedures or admission types, or detailed analysis of changes in health plan provider contracts.

At the aggregate health plan or DM program level, regression to the mean poses a serious threat to the validity of the results. Also referred to as statistical regression, this concept suggests that, without the effect of the intervention, members with high costs and utilization in the baseline year will tend to cost less and use fewer services in the following year (a move toward the mean). Conversely, members using few services in the baseline year will use more services and accrue higher costs in the subsequent year. Some DM companies contend that, because of the nature of progressive chronic disease, increased costs can, and should be, expected year over year. The results shown in Figure 2 should dispel that notion. As illustrated in this single health plan example (using a continuously enrolled cohort over the course of 2 years during which no chronic disease interventions were in place), regardless of type of chronic illness, the regression to the mean occurs. From 7% to 11% of members in the highest cost quintile in year 1 will move to the lowest cost quintile in year 2. Conversely, 11–17% of members in the lowest cost quintile in year 1 will move to the highest cost quintile in year 2. While these data show that there is movement toward the mean, the actual total costs may vary from year to year, and from quintile to quintile. Therefore, it would be incorrect to conclude that total average costs remain stable across measurement periods.

From a programmatic perspective, these data suggest that a DM program should incorporate a predictive modeling tool for identifying which members will be at highest risk for future costs, and provide interventions specifically tailored toward their needs.[10,11] Many DM programs use stratification methods, which target today's highest-cost members for nursing intervention, while those members who fall into the low-cost/low-risk pool typically receive nothing more than periodic educational mailings.

Measures of a DM program's success may be additionally influenced by the health plan's overall case-mix. As mentioned earlier, high disenrollment rates from the health plan may impact who remains in the DM program. Disenrollment may be effected by a decrease in member benefit coverage, a decline in satisfaction with the plan, or the effects of some cost-containment strategies.[12] Conversely, a more "favorable" case-mix may result from a health plan offering more competitive premium rates for healthier employer groups. Some factors affecting case-mix, such as age, sex, number of chronic conditions, employer group mix (by industry type classification codes), and health plan line of business mix (ie, Medicare, Medicaid, commercial), are easily measured and monitored. Year-to-year changes in the measures, outside an agreed-upon corridor, might trigger additional adjustments in the evaluation methodology.

Another serious impediment to ensuring the validity of the outcome measure is the effect of different interventions on the population occurring simultaneously. There may be initiatives that the health plan has implemented that coincide with the DM program (such as an inpatient concurrent review program or claims audit program), making it difficult to ferret out the individual contribution of each initiative to the change in outcomes from baseline. Another common problem muddying the waters is the interaction between various comorbid conditions and the targeted disease. For example, a subset of patients with congestive heart failure (CHF) may also have diabetes, or chronic obstructive pulmonary disease (COPD), or coronary artery disease (CAD). As a result interventions targeting those conditions will also have an impact on the DM program-specific disease.

Changes to a health plan's benefit design or

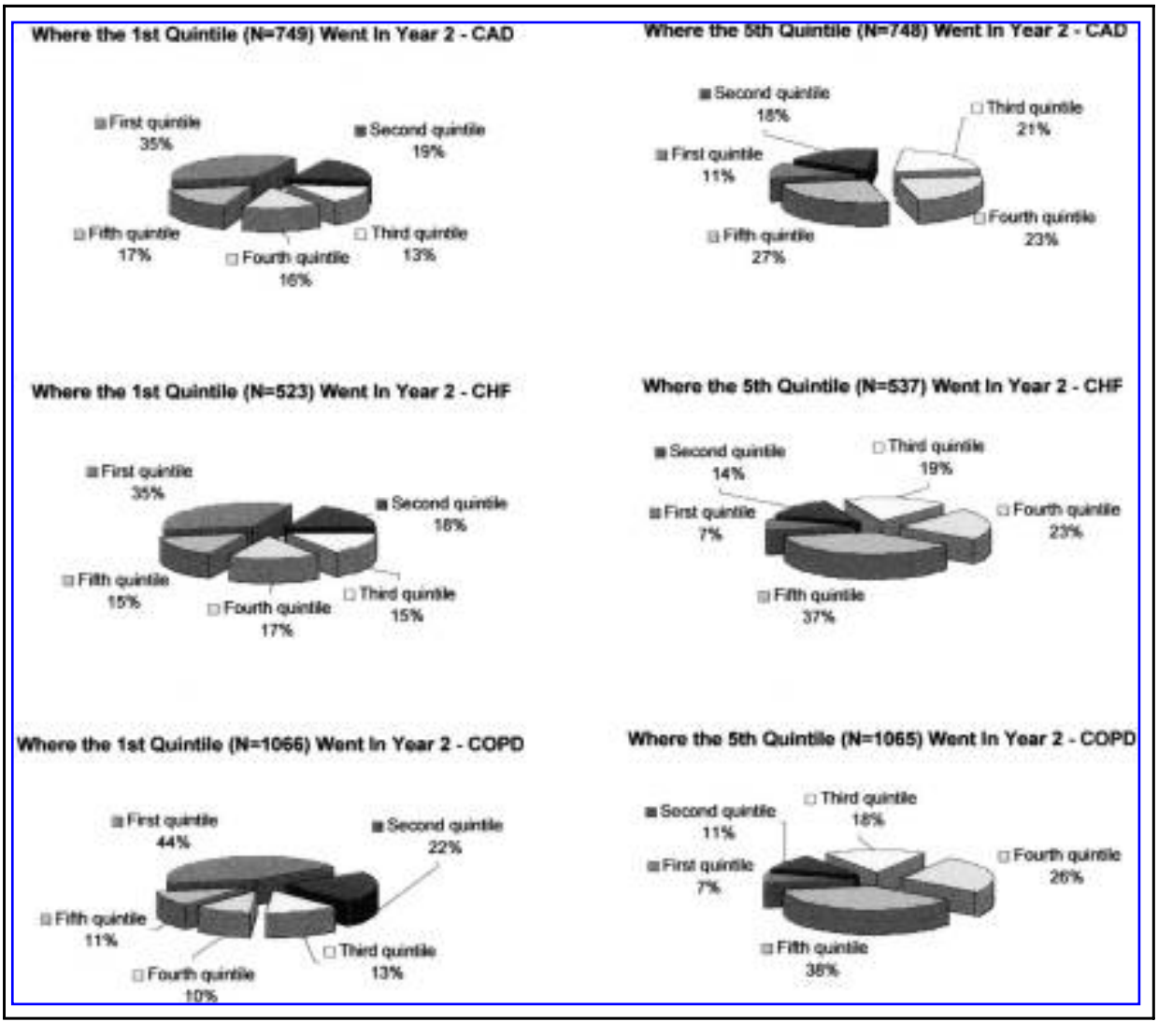

FIG. 2. Actual data illustrating the regression to the mean phenomenon in CAD, CHF, and COPD. Quintiles are ranked from 1 to 5, with 1 being the lowest cost group and 5 being the highest. All patients were continuously enrolled during the 2-year period.

provider network may impact availability of and access to medical services. For example, a narrowing of a plan's specialist panel available to members may reduce use of these services. Likewise, moving from a primary care provider (PCP) "gatekeeper" system to a direct access model may increase the utilization of specialty services.

One factor that is nearly impossible to control for is the impact of new technologies on medical costs and outcomes at the health plan or DM program level. At the early stages of deployment, a new testing device or procedure may cost several hundreds, if not thousands, of dollars more than the current standard of care. Moreover, the innovation may lead to higher hospitalization rates or longer recovery periods, even though health status may be significantly improved over the current methods. That said, a myriad of possibilities exist for the effect of introducing new technologies in health care.

Physicians and providers

Physicians and other providers may be a source of confounding that could impact measurement variables. One well-known phenomenon, referred to as the Hawthorne effect, sug-

gests that people will alter their behavior while knowingly being observed. It is plausible that providers may change their practice patterns in advance of the introduction of a DM program. For example, the mere discussion of introducing a CHF DM program may be enough to remind some physicians to review their angiotensin I-converting enzyme (ACE) inhibitor prescription practice. A DM program administrator may argue that this is a natural byproduct of the intervention and that these effects should be considered integral to the program effect. However, from an impact assessment perspective, this could be considered a confounding variable that would dilute the ability to measure efficacy of the patient-centered program.

Physicians may also alter their practice patterns because of changes in risk models or reimbursement methodology. For example, reducing the financial risk borne by the PCP for use of specialty services may increase referrals to specialists. A shift in PCP reimbursement method from capitation to fee-for-service may drive up the rate of office visits and other services previously under the capitation rate.

Data

Issues pertaining to the availability, accuracy, and use of data may impact outcomes measurement as well as influence how effectively patients are identified for inclusion in the program. Typically, DM programs use a claims-based algorithm to initially identify patients suitable for the intervention. Through subsequent physician and/or member contact, some of those identified through claims are deemed not to have the disease of interest (ie, "false-positives"). Some DM programs may use suitable members in the baseline measurement period, yet use enrolled members in subsequent measurement periods. If the algorithm leads to a high false-positive rate, then costs may be overstated in the baseline. This would occur because the baseline measure assumes that those members identified would have the disease, and would therefore be suitable for the program. During the program years, false-positives would be weeded out and thus would not be included in the subsequent measurements.

These algorithms usually require several data sources for maximizing the potential for accurate disease identification. If certain data sources are unavailable, then the accuracy of the identification process may be compromised. For example, if the health plan does not have a pharmacy benefit for its Medicare members, and the algorithm needs ACE inhibitor data to accurately identify CHF patients, the end result will be missed cases suitable for the intervention.

Seasonality or cyclical trends may influence the outcome measure if left unaccounted for. For example, if the DM program begins in March of year 1, and the first measurement period will be shortened to accommodate a calendar year cycle, then the highest-cost period of the year, December and January,[13] will be missed from the measurement. While most DM programs operate on a 12-month cycle, there may be cyclical trends that occur less often because of external factors.

Measurement error

One of the biggest concerns facing health plans when contracting with DM vendors is the reconciliation process. Because of the sometimes large discrepancies found in the results when data analyses are performed by both the vendor and the plan, this period has been coined "the reconciliation blues."[14] This is a textbook example for demonstrating the importance of having reliable outcome measures. In other words, if the iterative process for arriving at the results is clearly defined and followed, then the outcome measure should be identical whether the vendor analyzes the data repeatedly, or whether the health plan and the vendor run the analysis separately and compare results in the end.

Poorly chosen outcome measures may invalidate the entire program evaluation. In the simplest of terms, validity means that the outcome measure is a true portrayal of what it is supposed to represent. Sometimes the measure appears to be valid, but the underlying process to achieve the results invalidates the

outcome. For example, assume that the chosen outcome measure is PMPY costs, and that total healthcare costs are included in the calculation (including disease-specific and non–disease-specific costs). It is entirely possible that disease-specific costs increased (as a result of increased hospitalizations and emergency department visits), while costs decreased enough in the non-disease category to outstrip those disease-specific increases. As a result, the total PMPY costs for the measurement period would be lower than in the baseline. Clearly, a successful DM program would be expected to reduce utilization in these categories, not increase them, yet the inclusion of non-disease costs into the equation could bias the resulting outcome measure.

General

Long-term trends occurring in the community in which the intervention takes place is referred to as a secular trend, or drift. Secular trends can make the intervention appear successful when in fact rates of the particular illness are decreasing in the community as a whole. For example, a smoking cessation program implemented by an asthma DM program over multiple contract years may show a positive impact on smoking rates in the disease-specific population. If the general community's smoking prevalence decreased at the same rate, the secular trend should be credited for the decrease, not the DM program intervention.

The impact of multiple DM programs in place in the same community may overstate a single program's impact. Physicians may change their practice behavior because of the introduction of a DM program in another health plan to which they belong. The last health plan in town to introduce a program may already have the influence of the prior programs present in its baseline data. Likewise, having multiple plans, sharing overlapping provider networks, or using the same DM program may have different effects than that same DM program alone in another community.

Availability of medical services in the community may also impact DM program outcome measures. For example, an undersupply of primary care physicians in the community may drive up emergency department visits. Overflowing emergency departments and hospitals at near 100% bed capacity may cause higher use of home care services.

CONCLUSIONS

The purpose of this paper was to present a detailed review of the pretest–posttest design, which is referred to in the DM industry as the "total population approach." The shortcomings of this design were thoroughly illustrated in order to provide readers with the tools necessary to identify potential problem areas during the DM program evaluation.

Given the limitations of this evaluation design, the pervasiveness of its use, and the importance of the programs it is used to measure, opportunity currently exists to design methods more appropriately suited to the DM program construct. The natural setting in which these programs function lends itself to quasi-experimental designs such as time-series analyses, etc. Hopefully these options will be explored in the near future.

REFERENCES

1. Baldwin A 3rd. Financial and risk considerations for successful disease management programs. Manag Care 1999;8(11):52–65.
2. Diamond F. DM's motivation factor can skew study results. Manag Care 1999;8(6):45–50.
3. Goldstein R. The disease management approach to cost containment. Nurs Case Manage 1998;3(3):99–103.
4. Kosma C. Faith or evidence: what is the purpose of disease state management? Manag Care Interface 1999;12(6):68, 73.
5. Carroll J. Health plans demand proof that DM saves them money. Manag Care 2000;9(11):25–30.
6. Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1966.
7. Cook TD, Campbell DT. Quasi-experimentation: design and analysis issues for field settings. Chicago: Rand McNally College Publishing Company, 1979.
8. Shadish SR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin, 2002.
9. Schu B. Zeroing in on DM. Manag Healthcare News 2000;16(9):14–16.
10. Cousins MS, Shickle LM, Bander JA. An introduction

to predictive modeling for disease management risk stratification. Dis Manage 2002;5:157–167.
11. Linden A, Schweitzer SO. Applying survival analysis to health risk assessment data to predict time to first hospitalization. In: AHSRHP 18th Annual Meeting. Washington, DC: Academy for Health Services Research and Health Policy, 2001:26.
12. Linden A, Schweitzer SO. Medicare HMO ambulatory service denials: determinants and consequences. In: AHSRHP 17th Annual Meeting. Washington, DC: Academy for Health Services Research and Health Policy, 2000:17.
13. Linden A, Schweitzer SO. Using time series ARIMA modeling for forecasting bed-days in a Medicare HMO. In: AHSRHP 18th Annual Meeting. Washington, DC: Academy for Health Services Research and Health Policy, 2001:25.
14. Cellini GL. Disease management—the reconciliation blues. Available at: http://www.healthleaders.com/news/feature1.php?contentid=39481&CE_Session=e53af34ff8342f37dc0fa4cd29b04690. Accessed December 3, 2002.

Address reprint requests to:
Ariel Linden, Dr.P.H., M.S.
Director, Clinical Quality Improvement
Providence Health Plans
3601 SW Murray Boulevard, Suite 10
Beaverton, OR 97005

E-mail: lindena@providence.org