PART 1

AN INTRODUCTION TO CARE MANAGEMENT INTERVENTIONS AND THEIR IMPLICATIONS FOR ACTUARIES

Paper 2: Actuarial Issues in Care Management Evaluations

By Henry G. Dove, Ph.D.[1] and Ian Duncan, FSA, FIA, FCIA, MAAA[2]

August 10, 2004

INTRODUCTION

This paper addresses specific details of measurement principles and practice that the actuary should consider when planning to conduct or review a study of care management interventions. Three major topics are covered: **Measurement Principles**, which discusses basic principles that should be considered in any evaluation; **Study Design Issues**, which explores issues that arise when assessing or planning a study; and **Risk Factors**, which covers factors that influence the inherent risk in a population being managed and which therefore influence the measured outcomes.

In the evolution of managed care, actuaries have tended to function within traditional roles (product development, pricing, rate filings, reserving and underwriting) while care management functions have been provided by professionals with a clinical background. Often, the two professions have operated in separate functional areas, coming together only at the most senior level of the health plan. One consequence of this separation of clinical and financial functions has been the establishment, in many health plans, of a separate informatics and evaluation function within the care management area, staffed by non-actuarial health professionals.

More recently, however, as health care costs continue to escalate despite many and varied clinical intervention programs, the senior financial managements of health plans have begun to look to the actuarial profession for counsel. Because the health care actuarial profession has traditionally been involved in rigorous financial calculations, actuaries understand health insurance and health claims data. Although much of the debate in care management evaluation concerns methodologies, methodologies are just a part of a larger set of issues concerning the validation of financial outcomes.

Assuming that actuaries become more active in care management issues, we believe they will be involved in three important areas: the economics of care management programs, risk adjustment and predictive modeling, and financial outcomes evaluation.

---

[1] Yale University, Division of Health Policy and Administration

[2] Lotter Actuarial Partners, New York, NY

The Economics of Care Management Programs

One factor common to the seven care management programs described in the first paper is that they all involve, to different degrees, highly qualified and costly clinical resources. While considerable attention has been paid to evaluating outcomes and savings from these programs, as discussed in detail in Paper 3, fewer questions appear to be asked about the relationship between inputs and outputs, or the appropriateness of the level and volume of clinical resources and programs to the outcome. Rather than analyzing the economics of a particular care management opportunity, a health plan is more likely to determine its level of case management intervention by using industry norms or benchmarks from other plans, potentially replicating over- or under-resourcing mistakes made elsewhere in the industry. Paper 4 addresses this issue in more depth.

Risk-Adjustment and Predictive Modeling

Risk-adjustment and predictive modeling are processes for: comparing different populations, where to devote clinical resources, how to evaluate programs, and how to profile and reimburse providers. Risk-adjustment and predictive modeling have been addressed elsewhere in the actuarial profession.[3, 4] Predictive modeling is used standardly in Disease Management (DM) to identify candidates for intervention programs. DM companies differ in the balance that they strike between "risk" and "impactibility." In this context, "risk" implies that a group of members are highly likely to experience high cost; "impactibility" introduces the idea of suitability for DM (for example, those members who exhibit signs that they are ready to change behavior, or who have a condition that, while less risky, is more amenable to telephonic management). We are also beginning to see interest in using risk adjustment or similar techniques (for example, propensity scoring) in the process of assessment of outcomes. (We will return to this topic in more detail in Paper 5.)

Financial Outcomes Evaluation

Program evaluations have generally tended to validate the savings of programs, despite continued escalating health plan costs. There are many issues with the methodologies chosen for these studies, which we will cover in a future paper. As important as the choice of a methodology, however, are the adjustments made to achieve comparability between the reference and the intervention population. (In this series of papers, the terms "reference population" and "comparison population" or "comparison group" are used interchangeably.) Many of the issues

---

[3] See, for example: Cumming, R.B., Knutson, D., Cameron, B.A. and Derrick, B. "A Comparative Analysis of Claims-Based Methods of Health Risk Assessment for Commercial Populations." Society of Actuaries, 2002. Dove, H., Duncan, I.G. and Robb, A.S. "A Prediction Model for Targeting Low-Cost, High-Risk Members of Managed Care Organizations." American Journal of Managed Care, 2003, 9 (5): 381-389.

[4] Duncan, I.G. and Robb, A.S. "Population Risk Management: Reducing Costs and Managing Risk in Health Insurance." in Jain, L.H., and A.S. Shapiro, eds. : "Intelligent and other Computational Techniques in the Insurance Industry - Theory and Applications." World Scientific, December, 2003.

faced by researchers evaluating equivalence are the same issues faced by actuaries in pricing and underwriting different populations. Actuaries, using their background and training, can help to bridge the gap between program outcomes and the overall trend in health plan costs.

While there has been general acceptance of intervention programs clinically, the same is not true of financial results of interventions. The most significant ongoing issue for any form of intervention program is its ability to justify itself financially. A recent meta-analysis survey of *clinical* outcomes of disease management programs showed that these clinical outcomes were generally favorable.[5] A similar survey of financial outcomes found mixed results.[6] Since it is an axiom of the managed care industry that "higher quality" leads to lower cost,[7] the apparent inconsistency in these two studies should be of concern to all who work within the care management industry, and requires further analysis. Future papers in this series will address these issues in more detail.

Evaluating intervention programs has proved to be difficult because what is being measured is often something that did *not* occur. The objective way to measure the non-occurrence of a particular event is through a randomized control test. However, it is generally believed by health plans that conducting randomized trials is impractical (or even illegal). It is not considered to be feasible to design a study that withholds medical management services from an otherwise eligible health plan member, solely for the purpose of collecting information on equivalent patients who are not affected by intervention programs. Thus most studies that are conducted for business purposes use some form of non-randomized control methodology, or no control at all.

[5] Weingarten, S., Henning, J.M., Badamgarav, E., Knight, K., Hasselblad, V., Gano, A. and Ofman, J. "Interventions used in disease management programmes for patients with chronic illness—which ones work? Meta-analysis of published reports" British Medical Journal, 2002, 325.

[6] Krause, D.S. "Review of the Literature: The Financial Effectiveness of Disease Management." (Unpublished; forthcoming, 2004).

[7] See, for example: "Crossing the Quality Chasm: The IOM Health Care Quality Initiative." Institute of Medicine, 2003, or Gingrich, N. Ph.D. "Saving Lives and Saving Money." The Alexis DeToqueville Institute, Washington, DC, 2003.

Measurement Principles

Actuaries who deal with measurement of intervention outcomes should be familiar with the following six principles when constructing, reviewing or comparing a study. The first three of these principles are taken from a paper by Wilson and MacDowell.[8] We have added three other principles of our own that we have found to be equally important in practical applications.

Reference Population
Any outcomes measurement requires a reference population against which to evaluate the statistic(s) of interest.

Equivalence
To ensure validity in outcomes measurement, the reference population should be equivalent to the intervention population. We discuss the meaning of "equivalence" in more detail below.

Consistent Statistics
The comparison needs to measure the same outcome variable(s) in the same way in the reference and intervention populations.

Appropriate Measurement
Avoid, if possible, extraneous, irrelevant or confounding variables (factors) in measurement. As an example, a DM program may be implemented to manage the medical admissions of chronic patients. The actuary could measure all admissions (medical and surgical) of all patients (chronic and non-chronic). However, the medical and surgical admissions of all patients will be affected by many different factors, some of which may be influenced by DM, while many will not. The chances of a broad analysis being confounded by these other factors and non-managed lives is far greater than a narrow study of medical admissions within the chronic population. We do not go so far as to recommend that the study follow only the members who enroll in a program, because that approach introduces other biases. However, by defining as narrow a population as possible, and as narrow a set of outcomes as possible, the effect of confounding will be reduced.

Exposure
As actuaries are well aware, the calculation of an actuarial statistic requires clear definition of the numerator and denominator. In actuarial calculations, the denominator is referred to as "exposure." Accurate calculation of exposure requires similarly explicit definitions of categories of member, measurement time-periods, and eligibility in those periods. Those members who meet these definitions should be included in the appropriate group in the measurement period.

---

[8] Wilson, T.W. and M. MacDowell. "*Framework for assessing causality in Disease Management: Principles.*" Disease Management, 6 (3) Fall 2003.

Reconcile the Results
DM companies frequently analyze only small (managed) sub-populations, and sometimes claim savings results that do not appear to be reasonable in the context of the entire population or health plan. The actuary should be prepared, therefore, to reconcile the outcomes of a small population and those of the entire health plan. More important, the actuary should be prepared to explain what factors are driving the health plan's overall trend upward, even when the outcomes from the DM program are favorable.

Study Design Issues

Outcomes are evaluated within the context of a study design. Examples of study designs are: randomized, historical control or observational. The application of the study design raises many issues, including methodological issues, measurement issues, data issues, issues specific to chronic populations, and claims issues.

In future papers in this series, we will examine how some of these issues can affect measured patient outcomes and the estimated cost-effectiveness of interventions, as well as techniques that may be used to mitigate their influence on a study.

Methodological Issues in Study Design
Ensuring equivalence in the reference population is an important methodological issue. As discussed above, a good study methodology should include a reference population. Reference populations are generally constructed by one of three methods: randomized selection from the overall population; non-random selection from the population, with or without adjustment; or by following patient experience over time (a methodology often referred to as "patient as their own control").

The practicality of the study design implementation is also a consideration. Although a randomized trial does not necessarily guarantee an equivalent population, it is considered to be the "gold standard" for clinical researchers. Even in a randomized trial, equivalence between the intervention population and control group still needs to be demonstrated. Achievement of randomization in DM evaluation studies is believed by health plans to be impractical or even, in some instances, to be forbidden by medical ethics or regulation. When randomization is not possible, every effort should be given to planning and executing a study in such a way that equivalence is demonstrated in the reference and intervention populations.

Individual versus population studies is the last methodological issue explored. Many studies that claim to employ a reference group use the patient (pre-intervention) experience as the reference and patient (post-intervention) experience as the intervention group. While this design may meet the criteria for a reference group, the reference group may not meet the criteria for equivalence.

Measurement Issues in Study Design
In this section we review questions such as what to measure and when to measure it.

Appropriate Outcome/Outcome Measure
Clinicians, patients, and researchers often disagree about what outcome measure is most suitable. Patient "outcomes" include medical costs, quality-adjusted life years, functional status, employment status, long-term clinical outcomes, prevention of high-cost events, and patient satisfaction/quality of life measurement. The result of greatest interest to the actuary is the financial outcome measured either directly via claims, or indirectly, for example in terms of admissions. Paid claims net of cost sharing are subject to a number of effects such as contractual arrangements, plan design features, primary/secondary payer responsibility, or new technology. Thus an alternative measure not affected by these factors, such as admissions, bed-days, or allowed charges may be a more stable variable for the purpose of outcomes tracking.

Timing of the Study: Determining "End Points" and "Starting Points"
In most clinical trials, patient "exposure" to a particular treatment begins at a defined time and ends at a pre-determined time, based on risk profile. However, a population measurement involves a single start- and end-date for the entire population. During the period of measurement, different members will have different risk profiles—some will be recently diagnosed, diseases will have progressed; some members will have had recent "events" (such as a hospitalization) and others not. Measurement of exposure and risk are fundamental building blocks of actuarial science, so the appropriate classification of members over time is an area where actuaries may be able to make a contribution to outcomes measurement.

Total Medical Costs Versus Disease-Specific Medical Costs
Most care management strategies focus on specific diseases. It is challenging to separate the medical costs by disease entity, for two reasons. First, since there is not always consistent coding of the medical claims on which evaluations rely, and claims may be coded to maximize reimbursement rather than ensure comparable outcomes, isolating the costs related to a single disease may prove impossible.

Second, members enrolled in disease management programs often suffer from more than one chronic disease. Where a particular chronic member should be classified is a challenge: should the member be classified according to the primary diagnosis on a claim, or according to the most frequently encountered diagnosis, or the most expensive diagnosis? From a financial perspective, a DM program is usually implemented to reduce costs, not disease-specific costs, so measurement of overall cost savings is appropriate.

Data Issues
As actuaries are all-too aware, drawing financial conclusions from data requires attention to data quality and interpretation. Many of the measurement issues in study design concern sources and uses of data.

Three common sources for data are incurred claims data, medical records, and survey data. The source of the data can affect measurement reliability. One characteristic of many intervention programs is the limited availability of machine-analyzable data. This is in part due to clinical training, which emphasizes extensive note-taking, and is resistant to a program design that emphasizes automation and homogeneous definitions, as are required for machine-analyzable data.

The timing of data collection and evaluation is also an issue. The financial pressures on both for-profit, publicly traded health plans and not-for-profit plans demand very quick evaluation of outcomes. This constraint, together with high membership attrition rates limits a Managed Care Organization's (MCO's) ability to continue a program and to track outcomes for a period of months or years. It also argues for proxy methods of interim measurement, based for example on admission data, work volumes or clinical improvement measures. There is a hypothesis (not tested, as far as we are aware) of "recidivism" (the tendency of the measured outcome to reverse over the long-run) in case and disease management. An intervention program may appear to achieve cost savings over a six- to 12-month period, but in fact, costs are simply deferred to a later period.

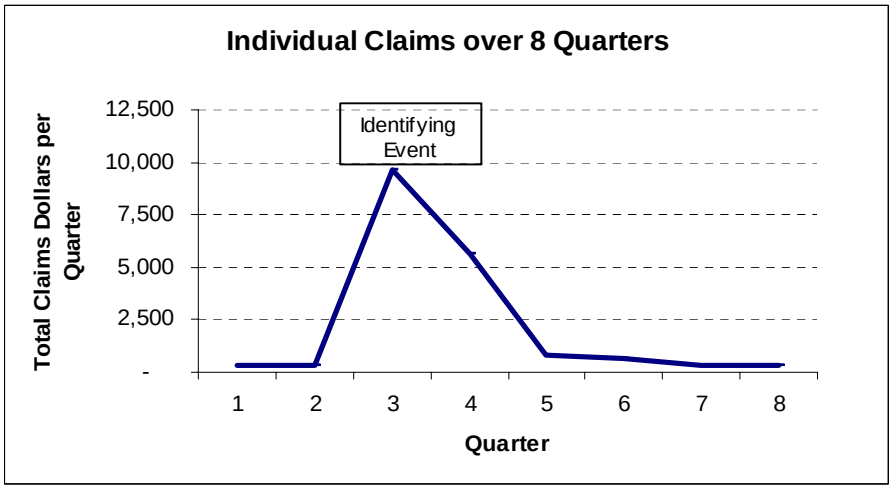
The issue of definition, which members to track for evaluation purposes, will be covered in more detail below.

Measurement Issues Specific to Chronic Populations
This section discusses certain issues specific to chronic condition populations that affect Disease Management evaluations.

Regression to the Mean
Many before-and-after evaluations that use the patient as the unit measurement (so-called "patient as their own control" designs) ignore the phenomenon that the outcomes of patients in period t+1 (evaluation or measurement period) are very often influenced by their state in the prior period t. Specifically, a high percentage of high-cost patients in period t are no longer high-cost in period t+1.

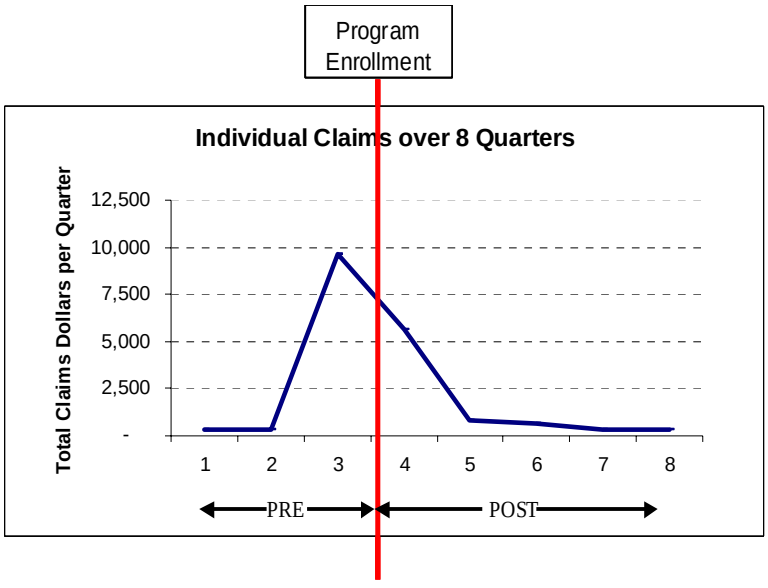
The graph in Figure 1 illustrates the phenomenon of regression to the mean at the level of the individual member:

Figure 1



Depending on when this individual's experience begins to be tracked for the purpose of measurement, regression to the mean may be captured in the claims data. For example, if the identifying event for a DM program is the hospitalization claim that occurred in Quarter 3, and this claim is included before the start of the DM program, the tracking of the experience after the program start will show lower cost. The reduced cost may incorrectly be attributed to a DM program, when, in fact, the cost reduction is the natural course of the individual's illness and claims experience. This phenomenon is illustrated in Figure 2. In this example, an individual member is identified (through claims) and enrolled in a program. The experience before the member's enrollment (the enrollment is indicated by the vertical line) is included in the "Pre" experience; the experience after enrollment is included in the "Post" experience.

Figure 2



In addition to its effect at the individual level in the "patient as their own control" type studies, regression to the mean has implications for population studies. It is often assumed that, because individual member level regression (as illustrated above) is present, the entire population experience will exhibit the same phenomenon. This is not necessarily the case, however. A group of individuals identified through a sentinel event (such as a hospitalization) will exhibit regression to the mean; an entire population, consisting of members identified at different times, may or may not exhibit regression.

Figure 3 illustrates the more general impact that regression to the mean (claims increasing as well as decreasing) may have on an analysis. Note that Figure 3 differs from Figure 1 and Figure 2, which show an individual's claim cost pattern over time, because Figure 3 shows the claims experience of an entire population over two years. Only members who were eligible and had claims in Year 1 are included in this analysis, so new members or members who had no claims in Year 1 are excluded.

In Figure 3, in which data are for the continuously enrolled members of a managed care plan for the two years 1997 and 1998, members are allocated into categories based on their cost-category in Year 1 ("Historic Period"). The members of this population are drawn from a health plan with limited managed care interventions: pre-authorization, some concurrent review and in-hospital case management, but no outpatient case management or disease management.

Figure 3

$' 000	Distribution of members and claims				
		Projection Period			
Historic Period Group	Historic Period Cost	$0 - $2	$2- $25	$25+	Projection Period Cost
Low $0 - $2 87%	$324	$ 327 90% *90%*	$ 5,368 10% *64%*	$ 46,836 0% *40%*	$ 831
Moderate $2 - $25 12%	$5,658	$ 668 55% *10%*	$ 6,599 40% *34%*	$ 47,811 5% *40%*	$ 5,398
High $25+ 1%	$49,032	$ 847 26% *0%*	$ 9,609 46% *2%*	$ 58,489 28% *20%*	$ 21,017
TOTAL	$1,230	$ 355	$ 5,851	$ 49,377	$ 1,581

Source: Lotter Actuarial Partners data; 200,000 continuously enrolled members of an HMO; Baseline year; 1998; Projection period is 1999.

One percent of members have historical costs in excess of $25,000, with an average paid claim cost of $49,032. The outcome of each category is shown in Year 2 ("Projection

Period"). Ninety percent of Year 1 low-cost members remain in the same category in Year 2, with approximately the same average cost. The second line under the projection period distribution of members and costs indicates the source of that period's membership in the prior year: for example, 64 percent of the intermediate group of members in Year 2 come from the prior year's low-cost members. Regression to the mean is illustrated by the outcome of the one percent of members who were high-cost in Year 1: 26 percent of these members are low-cost in Year 2, and 46 percent of these members are in the intermediate group. Only 28 percent of the members continue to experience high costs in Year 2, while nearly three-quarters of members have costs less than $25,000. The average cost of the high-cost members declines from $49,032 to $21,017 from Year 1 to Year 2.

The "Moderate" cost group in Figure 3 consists largely of chronic patients. Note that in this example, if the population tracked is the Year 1 "Moderate" cohort, the average cost is observed to fall 4.6 percent from $5,658 in the baseline year to $5,398 in the intervention year, in the absence of any interventions. If the population tracked is the Year 1 moderate population compared with a similarly defined Year 2 moderate population, costs increase 3.4 percent, from $5,658 to $5,851.

Identifying Patients

The above discussion of regression to the mean argues against use of "patient as their own control" as a comparison group. A frequently used alternative is the "Population" approach, in which all members who meet the identification criteria in a baseline period are considered the comparison group, and all members who meet the same set of identifying criteria (irrespective of whether they were included in the baseline population, are enrolled in the program, etc.) are considered to be the intervention population. Very precise criteria should be established to identify chronic patients, and determine when they are included in the study. This method of identifying a comparison population relies on uniformity of the distribution of members with respect to the cost of their disease. Some members will be experiencing declining costs, as in the example above, while other members will be experiencing increasing costs as they experience a health-related event. Provided the distribution of member risk-status is similar in each year, this population approach will result in equivalent populations.

Establishing Uniform Risk Measure for Comparability

Different patients present widely differing combinations of co-morbidities, conditions, and other risk factors, in addition to different risk profiles at different times. Evaluation of outcomes requires a method for ensuring equivalence between populations. Specifying and identifying patient co-morbidities and risk factors continues to be a challenge of clinical epidemiology. Many of the risk factors that need to be considered in ensuring consistent risk-profiles are the same risk factors that actuaries use for pricing and underwriting health care coverage.

Claims data are subject to certain problems that can make them less reliable than medical record review or patient interviews for identifying chronically ill members and assigning a risk status to them. Patient interviews and chart reviews are impractical and subjective.

Objective, consistent definitions should be established that identify the population from which the target management candidates will be drawn, and whose experience will be tracked for financial outcomes measurement purposes. Identification criteria can influence the financial outcome of a program.

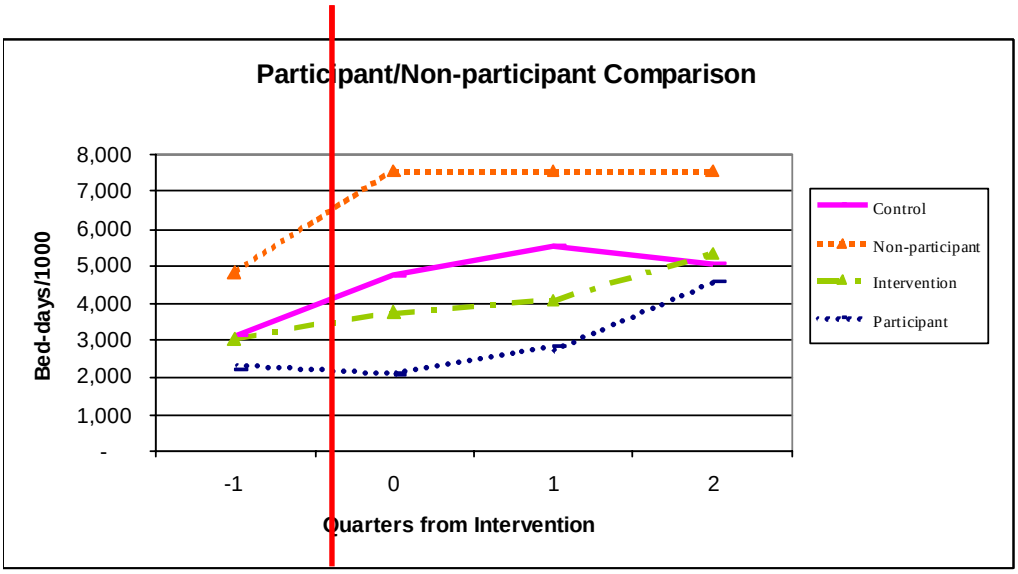
Patient Selection Bias

If randomized trials are not performed, there is always a potential problem of selection bias. Authors are divided about whether it is possible to adjust for bias. For example, Karen Fitzner, et. al., in "*Guide to Disease Management Program Evaluation*" (DMAA, Washington, DC, 2004) reviews different methods used by authors to avoid bias and confounding. All of these methods have in common two elements: the existence of bias is known and its extent is quantifiable. In the circumstance in which bias is suspected, but its extent is unknown, it appears to us that adjustment is difficult, if not impossible.

One of the most common sources of bias in evaluation is a study design that limits evaluations to those members who enroll in a voluntary program. By definition, those members who elect to enroll in a voluntary program are a different risk-profile to those members who do not elect to enroll. The following chart, taken from the authors' unpublished data, tracks outcomes over time of different sub-populations from a chronic disease population subject to a disease management intervention. Unlike most DM programs, the chronic patients included in this study were randomized first, prior to enrollment. Thus the outcomes of the intervention and control groups (absent random fluctuations) represent a robust comparison for evaluation. Members were initially randomly assigned (prior to the start of the program) to intervention (75 percent) and control (25 percent). The point in time at which identification and assignment to the intervention is performed (start of the intervention program) is indicated by the vertical line. Participants were recruited from the intervention group, and the control group was untouched. The unit of outcomes measurement reported below in Figure 4 is bed-days per 1,000 per year.

Readers who are familiar with health plan bed-days per 1000 per year statistics will find these levels high. Remember, however, that these statistics are for a sub-set of the population, the chronic members only.

Figure 4



Two different member-outcome states are reported in Figure 4: members who enrolled in the program (Participants) and members who declined to enroll (Non-Participants). Members whom the nurse interventionists were unable to reach (No contact) are included in the Non-Participant group. Outcomes are compared with those of the control group. The effect of the intervention is shown by the difference between the Control and Intervention groups, and represents the reduction in total bed-days seen over the three quarters.

Comparison between the Intervention and Control groups is appropriate, because the members are assigned to these two populations based on objective criteria. Comparison between the participating (self-selected) population and the control group is not appropriate because of the selection bias inherent in the participation process.

The Intervention group consists of two sub-groups: Participants and Non-Participants. Overall outcomes of the intervention group compared with the control group (the difference between the two middle lines) indicate reduction in bed-days. It is important to note that the beginning (pre-program) utilization of the Intervention and Control groups (two middle lines) is the same, consistent with the random (unbiased) allocation of members between the intervention and control groups. Utilization of the Participant and Non-participant sub-groups is significantly different (in particular, the Non-participant group has higher beginning utilization), indicating selection bias. The patients who enrolled in the program (resulting from the ability of the health plan to reach the member, and then the patient's willingness to enroll when reached) represent a different experience group than those who did not enroll. Specifically, the non-participants had higher

utilization than both the participants and the control group, indicating the effect of the enrollment bias.

Patient Drop Outs
Members may drop out of a follow-up study for a number of reasons: voluntary exit, termination from the health plan, transfer to a different group or product, or death. These factors can affect the outcomes. Within the enrolled group, the follow-up with different members is also potentially anti-selective; some patients will stay in a telephonic intervention program for the prescribed duration, while others will drop out because they are feeling better, or for other reasons.

General Versus Specific Population
Some interventions are used only on an extremely selected, and therefore small, subset of potential enrollees; thus sample size can be problematic unless very large populations are available. Large-case management interventions, for example, tend to be applied in a very small subset (often less than ½ percent) of the population. The co-morbidities, outcomes and cost of these members are highly variable, making it difficult to apply standard study designs. At the same time, the effect of the intervention, while significant at the individual level, may be too slight relative to overall claims to allow its effect to be measured in the entire population. A measurement methodology that is appropriate for a chronic population (where the prevalence of disease is often five percent or more in a commercial population) may not be appropriate in a large-case management population with a prevalence of ½ percent.

Claims Issues in Study Design
Most evaluations will be based on administrative claims. This section discusses five considerations relative to claims: fixed time periods, member eligibility, claims run-out, outliers, and special problems with claims data.

Fixed Time Periods
Epidemiologists sometimes consider one year's data inadequate for outcomes evaluation because with continuous identification and program enrollment, all patients do not have equal "exposure." In addition, because of the time taken for claims to be completed (see below), the amount of time taken to perform a rigorous evaluation of a program will be long, even if the time period is restricted to one year's incurred claims. For chronic disease management programs, however, there are usually a sufficient number of members with the condition that a "spread" of risk conditions will be assured, allowing for stability in measurement over time. Actuaries are used to calculating exposure, even when a member is eligible for less than one year, so this factor should not be a problem. Short exposure periods must, however, allow sufficient time for the "process" aspects of a program to be completed: data collection, chronic member identification, communication, enrollment, and patient education.

Enrollment Issues/Eligibility
Actuaries know that eligibility files of most managed care organizations are frequently incomplete, making it difficult to identify patients. The timeliness of new member

enrollment, or terminating member disenrollment should be factored into any study, since annual disenrollment rates exceed 20 percent in many plans. The drop out effect of member disenrollment is further complicated by members who terminate in one plan or product, but who reappear in the health plan under a different member identifier (because they have joined a new group, are covered by a spouse, or changed products).

Claims Run-Out

Analysts must wait for physicians and other providers to submit claims; however, there is usually a lag of several months in claims submission. In addition, when claims are disputed as to eligibility, subrogation or primary payment, claims that are initially processed may be re-adjudicated or reversed, making it difficult to draw conclusions from immature claims data. While actuaries have techniques for handling immature data, these techniques generally depend on data that reflect a stable underlying operational state. By definition, the introduction of care management introduces change to the operations of the health plan, potentially rendering projections based on the prior state invalid. Customers of medical management programs often want to see immediate results, and are not at all comfortable with the idea that they will be paying for a program where results will not be credible or stable for upwards of two years.

Outliers

Actuaries are familiar with the potentially distorting effect of outlier claims—atypical cases that may distort overall study results. In a DM program, outliers may be members with unusual conditions, individual large claims, or both.

Special Problems with Claims Data

The quality of claims data has improved substantially in the last 10 years. Hospital data is still vastly more complete and accurate than claims submitted by physicians. Pharmacy data, useful for identifying many conditions or identifying conditions on a more-timely basis than hospital claims, may not be present in certain groups of patients. When chronic patients are identified through claims, it is important that the claims and coding on which the identification depend be consistent between groups and over time. Because there is no single agreed upon definition of administrative-claim based chronic disease criteria, there is room for difference of opinion, and therefore “false positives” and “false negatives” occur in the identification of chronic members.

False positives are members identified as having a condition who do not, in fact, have the condition with which they are identified. False negatives are members who have the condition who are not identified through the identification algorithm. False positives in particular have an impact on financial outcomes measurement because, by definition, the false positive member does not have claims identifying the chronic disease in the intervention year (and is likely to be lower-cost than a member who does have the identifying claims). False negatives do not contribute in this way because they do not contribute claims costs.

Risk Factors

Now let's turn to the specific risk factors that must be considered in any measurement calculation and which should be reported in order to ensure comparability and reproducibility of results. In any study, various characteristics of the reference and intervention groups must be evaluated, so that the effect of the intervention on outcomes can be properly assessed. Statistical techniques (beyond the scope of this paper) may be used to adjust for differences, if needed. At the least, information should be given regarding the variable and its potential effect. The eight variables listed below are referred to by economists and epidemiologists as "confounding variables." Actuaries know them as "risk factors," and are accustomed to allowing for them in pricing or underwriting health insurance coverage.

1. Demographic variables
2. Exclusionary conditions that exclude certain members
3. Exclusionary conditions that exclude certain claims
4. Persistency
5. Chronic prevalence and risk classification
6. Severity of illness
7. Contactability
8. Operational Issues

At the very least, any study of outcomes needs to include reference to the values of these variables and the way in which they have been taken into consideration in the study design. When we consider that the issues implicit in a variable will be familiar to actuaries, we do not discuss the variable in detail. When a variable is not one of the "usual" variables that actuaries consider, we provide additional information. We will include more discussion of these variables in a later paper on an actuarial methodology for analyzing medical management outcomes.

Demographic Variables

Variation in the following variables (all of which are familiar to actuaries) can affect the result of any outcomes measurement exercise.

- Age
- Gender
- Medicare eligibility
- Other payer eligibility
- Other sources of services (either reimbursed or not reimbursed)
- Medical group election alternatives
- Product and benefits design/description

Exclusionary Conditions – Members

It is common in evaluations to exclude certain members from either the program or the evaluation, or both. "Exclusionary conditions" that exclude a member from the care program include:

- Conditions with severe privacy restrictions on either data or contacts (e.g., HIV/AIDS, Mental Health).
- Management of conditions for which the sponsoring organization has contracted with another vendor or the sponsoring organization provides for outside of the particular care management program (e.g., Mental Health, maternity, cancer).
- Conditions that imply that the member is not a good clinical candidate for care management (e.g., institutionalization, members in case management, and members with End-Stage Renal Disease).
- Conditions that imply that a member is not a good financial candidate for care management (e.g., program sponsor is the secondary payer, implying that any financial gains accrue to the benefit of a party other than the program sponsor).

Exclusionary Conditions – Claims

It is common to exclude certain claims from an evaluation in order to reduce confounding and/or "noise" from conditions that are either not manageable or are subject to fluctuations. Here are some typical exclusionary conditions to apply to claims in a savings calculation:

- Some categories of claims (maternity, mental health, cancer) are excluded from the measured outcomes experience because they involve conditions that DM does not aim to affect. (The conditions cited here are examples of conditions that are excluded from "traditional" disease management programs. There are specialist programs, however, that address these conditions. It would obviously not be appropriate to exclude these conditions for one of these programs.) Some authors argue that non-medical (surgical) admissions should also be excluded because these categories are particularly subject to supply-induced demand, making it difficult to compare populations over time or geography when these categories are included in the measurement. There is substantial literature[9] associated with the phenomena of supply-induced demand and practice variations, which is outside the scope of this paper.

Some categories of claims are excluded from calculations because they involve conditions that are difficult to predict and manage (for example: trauma, maternity, and

---

[9] Lucas, F.L., Wennberg, D.E., Malenka, D.J. *"Variation in the use of echocardiography."* Eff Clin Pract., 1999 March-April, 2 (2): 71-5.
Wennberg, D., Dickens, J. Jr., Soule, D., Kellett, M. Jr., Malenka, D., Robb, J., Ryan, T. Jr., Bradley, W., Vaitkus, P., Hearne, M., O'Connor, G., Hillman, R. *" The relationship between the supply of cardiac catheterization laboratories, cardiologists and the use of invasive cardiac procedures in northern New England."* Journal of Health Services Research & Policy, 1997 April, 2 (2): 75-80.
Wennberg, J.E. *"On patient need, equity, supplier-induced demand, and the need to assess the outcome of common medical practices."* Medical Care, 1985 May, 23 (5): 512-20.
Wennberg, J.E., Fisher, E.S., Stukel, T.A., Skinner, J.S., Sharp, S.M., Bronner, K.K. *"Use of hospitals, physician visits, and hospice care during last six months of life among cohorts loyal to highly respected hospitals in the United States."* BMJ, 2004 March, 13, 328 (7440): 607.

mental health), or because they constitute "noise" (or potential confounding) in a savings calculation.

Persistency

It is important to understand the terms under which a member may enter/leave the underlying group. DM companies often work only with data on the chronic population (or even more narrowly, the sub-set of the chronic population enrolled in the DM program) and therefore do not have insight into overall enrollment trends in a health plan. Results may be affected by persistency of enrollment with the medical group, medical plan, product, employee tier, or employer. Re-enrollment frequency and identification of the member across products may be a contributor (either positively or negatively) to trend[10], as members enter or leave a group. Claims levels will also differ according to the availability of out-of-state and out-of-area coverage and the likelihood of services being provided in those settings. Finally, another important aspect of persistency is persistency in the care management program. Different DM companies have different rules about required length of enrollment in, and conditions for "graduation" from a program. Recording of this persistency (and availability of the data) is not consistent between companies and sometimes makes comparisons between programs and vendors difficult.

Chronic Prevalence and Risk

The basis of "risk" and savings opportunity in a DM program is chronic prevalence: after all, the more chronically ill people that are present in a population, the greater the opportunity to improve health and reduce costs. (See Paper 4 for more detail about the economics of DM.)

Chronic prevalence is defined[11] as:

$$\frac{\text{Number of individuals with the condition}}{\text{Total number of individuals in the population}}$$

This statistic is measured at a single point in time, and therefore the statistic value will vary when calculating chronic prevalence at different times. A more important consideration when comparing prevalence between populations is whether prevalence includes or excludes duplication (members who have more than one chronic condition are counted only once, or are counted each time they have a condition).

The target chronic diseases are typically Heart Failure, Diabetes, Ischemic Heart Disease, Asthma, and Chronic Obstructive Pulmonary Disease (COPD). The technical definition of any one of the target chronic diseases, which is a combination of occurrences of specific claims codes, is crucial. Unfortunately, there is no uniform definition of chronic disease in use for either candidate identification or outcomes measurement purposes.

---

[10] "Trend" is used here in the actuarial sense, meaning the rate of increase in per member per month claims.

[11] See: Ian Duncan (ed): *Dictionary of Disease Management Terminology*. Washington, DC. Disease Management Association of America, 2004.

Chronic condition definitions change over time, as new codes are added to code sets,[12] and new sources of data become available. In any comparative study, objective criteria should be used that are easily applied, do not require manual intervention to perform, and can be readily implemented. Examples of disease definitions are:

- HEDIS (Health Plan Employer Data and Information Set) definitions
- Proprietary disease definitions, for example, those inherent in algorithms, such as D_xCG, ACG, etc. These may be more appropriate for a particular situation, but make comparability difficult.
- Some definitions depend on drug claims data or laboratory values, which are not uniformly available, making consistency and comparability difficult.
- Some disease definitions require clinical intervention (chart review). Others depend on self-reported data that are usually subjective, difficult and expensive to collect.

For a robust set of claims diagnosis-based definitions, see Ian Duncan (ed) *Dictionary of Disease Management Terminology* (DMAA, 2004, op cit.)

A patient's primary diagnosis (the condition, problem or other reason for the encounter that is chiefly responsible for the services provided) is usually more rigorously coded on claims than secondary diagnoses (other conditions or problems that affected a patient's treatment). However, in order to identify the chronic member and the member's co-morbidities, all diagnoses should be used. The accuracy of and rules for assigning ICD codes on medical claims by medical records technicians or billing personnel contributes to some of the "false-positive" issues.

The clinical view of a chronic disease is that once diagnosed, the disease continues for life. However, the concept of "once chronic, always chronic," which is consistent with the clinical view of chronic disease, is not always confirmed by the data. Some health plan members who meet a set of objective criteria for identification as a chronic condition member in Period 1, may not meet the same criteria in Period 2. This definition of "statistical false positive" is not the same as that used by clinicians or epidemiologists.

Given that identification of chronic condition is usually performed based on claims, we believe that our definition has merit in the DM outcome measurement context (although we do not lose sight of the fact that both clinical and statistical false positives and false negatives may be present in any measurement). Ideally, a set of identification definitions would be sufficiently sensitive to identify all members with the condition (limited false negatives). At the same time, the definitions should limit the number of false positives. Discussion of sensitivity and specificity is beyond the scope of this paper.[13] There is an

[12] Common examples of claims code sets include: ICD (International Classification of Diseases, 9th or 10th Edition Clinical Modification (ICD 9/10 CM); CPT-4 (Current Procedural Terminology, 4th Edition); HCPCS (Health Care Common Procedure Coding System); and NDC (National Drug Code).

[13] For definitions of Sensitivity and Specificity, see Ian Duncan (ed): "*Dictionary of Disease Management Terminology*" (DMAA, 2004).

inverse relationship between sensitivity and specificity. It is likely that a wide net will catch "false positives," or members who are identified as meeting the disease definition, but who do not actually have the disease.

Some obvious adjustments to the identification methodology can be made to ensure that members who appear to have a claim for a particular condition, but who, for example, may have had a test to "rule-out" the condition, are appropriately excluded. An example of a more difficult set of issues is those members who are identified through drug claims in one period, but not in a second period: was the member a "false-positive" or did the member's employer (or the member) switch drug coverage?

The minimum period of data required for consistent identification is related to the issue of disease definition: what minimum duration of data is required to accurately identify a member's disease state? Frequently, members are assigned to chronic categories based on the prior 12 months' claims experience. But is 12 months likely to result in a more accurate identification than six or 24 months of claims history?

In addition to chronic prevalence, risk and opportunity are affected by disease stratification and severity (e.g., Type 1 or 2 diabetes; Class III or Class IV Heart Failure[14]; or high-, moderate-, low-risk chronic members).

Disease Severity

Disease severity clearly affects a patient's claims cost, and therefore the potential for savings. Disease severity is more difficult to capture from claims data. The available data sources are those discussed previously in the "Data Issues" section of this paper, plus (once a patient is contacted and enrolled) self-reported data.

Many practitioners want to use risk-adjustment methods to assign a risk or severity "score" to a patient based on claims and diagnosis information. This method has some promise, where the risk-adjustment algorithms are "open" and can be replicated. However, the objective of DM is to influence the patient's disease state, achieving its aims through improvements in measures such as medication compliance and test scores (see Paper 5). These are the same variables that affect the patient's score, and there is potential for confounding. We will discuss further in a future paper.

---

[14] Type 1diabetes was previously called insulin-dependent diabetes mellitus or juvenile-onset diabetes. Type 1 diabetes may account for five percent to 10 percent of all diagnosed cases of diabetes. Type 2 diabetes was previously called non-insulin-dependent diabetes mellitus or adult-onset diabetes. Type 2 diabetes may account for about 90 percent to 95 percent of all diagnosed cases of diabetes. Class III Heart Failure results in marked limitation of physical activity. Patients are comfortable at rest. Less than ordinary activity causes fatigue, palpitation, dyspnea or anginal pain. Class IV Patients have cardiac disease resulting in inability to carry on any physical activity without discomfort. DM companies differ in their number of member risk levels, and how members are assigned to those levels. While useful clinically, the assignment of Heart Failure classes requires clinical assessment and cannot be performed from administrative claims, making this type of stratification of limited usefulness for claims-based outcomes assessment.

A consistent definition in terms of patient severity is critical for patient classification. Severity can have two meanings, however, that do not necessarily produce the same member stratification: financial severity (those members at high risk of adverse financial outcomes) and clinical risk (those at high clinical risk). There is little consistency in terms of patient risk classification in DM programs; vendors use different approaches that include some or all of financial or clinical risk, and gaps in care; some vendors classify into high-, medium-, and low-risk classes, and others classify into high and low only. Some vendors classify the high group as the top 10 percent of patients (ordered in terms of risk), while others use different percentages. Clearly, comparison between programs that define risk and determine the strata differently is difficult and subject to potential misinterpretation. In Paper 4, we will discuss an approach to risk classification that addresses these issues.

A related issue is the need for a consistent definition in terms of the likelihood to benefit from DM. Unlike chronic disease, which (when assessed from claims history) is reasonably objective, likelihood to benefit is a subjective concept. Despite this, many programs assess candidates and select them based on likelihood to benefit from the program. If outcomes are assessed for the selected population only, the selection process clearly introduces a non-random bias. However, if outcomes are measured for an entire, objectively identified population, this bias may be avoided.

"Contactability"

It is not sufficient to identify a chronic population; chronic care management aims to involve the member in taking responsibility for the member's own healthcare. Disease Management programs depend on the ability of the manager to reach and engage the member. If the health plan does not have good contact data, or if the members choose to ignore contacts by the health plan or its representatives, the program will not succeed. Populations will vary with regard to this risk factor (HMO and other gatekeeper type plans, which require "positive" enrollment generally have better contact data than indemnity-type plans, for example), and the "contactability" of a specific population, or of a population at a period in time, should be disclosed in order to ensure transparency and comparability of results.

Operational Issues

Because all DM programs are not the same, it is important for comparative purposes that operational statistics be reported. The following are examples of measures that should be reported for any program that is being evaluated:

- Number of eligible (health plan enrolled) members;
- Number of chronic patients identified, and timing of the identification;
- Number of chronic patients "available" (valid contact information; not on a "do not call" list);
- Number of chronic patients contacted;
- Number of chronic enrollees in DM program;
- Length of time the member is involved in the intervention;

- A definition of “graduation” (intervention program completion by the member) and member graduation rates, and;
- Methodology applied to compare the reference population to the intervention population should be clearly specified. Even within a particular methodology, different results may be obtained by the use of different assumptions with regard to variables, making full disclosure critically important.

CONCLUSIONS

In this second paper in the series, we have discussed a number of the definitional, measurement and data issues that an actuary should be aware of when performing a program evaluation. Many of these issues will be familiar to actuaries in other contexts, for example anti-selection or underwriting risk factors. Actuarial familiarity with the issues will increase the potential value of actuaries to an organization interested in objectively evaluating a program.

In Paper 5 we will review commonly used measurement methodologies, and then in Paper 6 lay out an actuarial-adjustment methodology for evaluating program management outcomes, including specifics of measures to address confounding and other issues mentioned here.

First, however, in Paper 3, we review the literature on clinical outcomes and cost savings reported in studies of different types of program.