A PUBLIC POLICY PRACTICE NOTE

Disease Management

December 2007

American Academy of Actuaries'
Disease Management Work Group

American Academy of Actuaries

This Practice Note was prepared by the Disease Management Work Group of the Health Practice Council of the American Academy of Actuaries[1]. The work group was charged with developing a description of some of the current practices used by U.S. health actuaries in 2007 with respect to work involving disease management programs.

Practice Notes from this work group describe what the work group believes to be the common practices of U.S. health actuaries. This Practice Note discusses some common approaches to evaluation in the disease management field. We make no representation of completeness; other approaches may also be in use. It should also be recognized that while this Practice Note provides guidance, it is not a definitive statement of generally accepted practice. Events occurring subsequent to the date of publication of this Practice Note may make the practices described herein irrelevant or inappropriate.

This Practice Note has not been promulgated by the Actuarial Standards Board, nor is it binding on any actuary.

The members of the work group responsible for this Practice Note are: Ian Duncan, chairperson; Robert Parke (former chairperson); Michael Blakeney; Andy Bren; Kevin Dolsky; Charles Fuhrer; Scott Guillemette; Donna Kalin; Arthur Lewis; Sandra Loyal; Wendi McNeilly; Catherine Murphy-Barron; David Nelson; Nancy Nelson; Donna Novak; Tim Robinson; Marcia Sander; Chuck Smith; Steele Stewart; Richard Tash; Michael Thompson; Howard Underwood; Greger Vigen; Margaret Wear; and Steve Wright.

Comments on the appropriateness of the Practice Note, frequency of updates, and substantive disagreements, are welcome. They should be sent to the Academy's State Health Policy Analyst, Geralyn Trujillo, at StateHealthAnalyst@actuary.org, or American Academy of Actuaries, 1100 17th Street NW, 7th Floor, Washington, DC 20036.

[1] The American Academy of Actuaries is a national organization formed in 1965 to bring together, in a single entity, actuaries of all specializations within the United States. A major purpose of the Academy is to act as a public information organization for the profession. Academy committees, task forces and work groups regularly prepare testimony and provide information to Congress and senior federal policy-makers, comment on proposed federal and state regulations, and work closely with the National Association of Insurance Commissioners and state officials on issues related to insurance, pensions and other forms of risk financing. The Academy establishes qualification standards for the actuarial profession in the United States and supports two independent boards. The Actuarial Standards Board promulgates standards of practice for the profession, and the Actuarial Board for Counseling and Discipline helps to ensure high standards of professional conduct are met. The Academy also supports the Joint Committee for the Code of Professional Conduct, which develops standards of conduct for the U.S. actuarial profession.

Table of Contents

Introduction and Purpose

This Practice Note discusses disease management programs. The Disease Management Association of America (DMAA) defines disease management (DM) as "a system of coordinated health care interventions and communications for populations with conditions in which patient self-care efforts are significant." DM programs may also be referred to as chronic care improvement programs. This topic is particularly challenging to measure and analyze because it involves the estimation of events that have not occurred. We recognize that actual evaluation practice varies substantially based on the size and business objectives of the client, population, and type of illness being reviewed.

This Practice Note provides a framework and describes key considerations for evaluating a DM program's impact on the cost and utilization of medical services. Clinical and humanistic outcomes, while they form part of most DM outcomes reporting, are outside the scope of this Practice Note.

Introduction

Escalating health care costs and an increasing public focus on health care quality are prompting employers and insurers to reassess the value and effectiveness of their medical management procedures. Many are looking at DM programs as a way to improve treatment of major chronic diseases, as well as the quality of life of employees/insureds, while reducing the need for and the costs of medical care. The improved health of participants in well-executed DM programs (such as programs aimed at managing diabetes and asthma) is clear and well documented. However, there is often a gap between favorable clinical results and a clearly identifiable financial impact. Many disciplines, including the actuarial profession, are establishing how to address the complex analytical issues inherent in assessing the financial impact of these programs.

Sponsors implement DM programs to improve health and productivity. They can also potentially save money for the ultimate payer (e.g., state Medicaid plans, employers, and other plan sponsors). A precise calculation of economic outcomes or return on investment (ROI) would take into account the clinical and humanistic components and the effects on other employer-provided programs (e.g., paid time-off, disability, and workers' compensation). However, such outcomes are outside the scope of this Practice Note. Therefore, for practical purposes, ROI is often measured by factoring in only the cost of the DM program and the program's impact on the cost and utilization of medical services (referred to as the "effect on medical costs" throughout this document). If the savings generated from more efficient use of the health care system are greater than the cost of the program, the program is considered to have generated a positive ROI. For payers using an outside DM vendor, the cost of the program would include the vendor's fees plus the internal cost of incorporating and running the vendor's program. For payers who choose to implement their own DM program, the costs would include developing and operating the program.

Definition of a DM Program

The DMAA defines DM as "a system of coordinated health care interventions and communications for populations with conditions in which patient self-care efforts are significant. It:

- Supports the physician or practitioner/patient relationship and plan of care;
- Emphasizes prevention of exacerbations and complications utilizing evidence-based practice guidelines and patient empowerment strategies; and
- Evaluates clinical, humanistic, and economic outcomes on an ongoing basis with the goal of improving overall health."

DM programs may also be referred to as chronic-care improvement programs.

State of the Industry

From its beginnings, as small programs operated by health plans or the pharmaceutical industry, to the current large, outsourced chronic-care improvement programs, DM's history is one of ongoing change. Its current state, in 2007, is no different. Experimental models include a significant internet education component, provider group-based care, integration of cases and DM, and wellness elements. Early exuberance over significant financial results gave way to growing skepticism, as plan sponsors failed to see the apparent savings reflected in their overall financial results. Plan sponsors turned to actuaries, who responded by acquiring expertise in this new area, while applying traditional actuarial techniques (e.g., data reconciliation and validation, trend and cost analysis, financial analysis). At the same time, the industry has responded by codifying its practices (for example, DMAA's publication of its guidelines for DM outcomes evaluation in 2006). While there is no "standard" financial measurement methodology within the industry, the increasing research and publication in this area may serve as an additional source of beneficial guidance for actuaries in their practice.

As the industry's evolution brings more measurement challenges, the principles of this Practice Note, together with the Actuarial Standards of Practice, will serve to guide our development.

Measurement Methodologies

A. Causality vs. Correlation

In evaluating a DM program, it is important to understand how that program, and any realized financial or clinical improvements, are related. Obviously, the actuary would measure the results of the intervention, and eliminate any effects that are coincident with, but not due to, the intervention.

> *Causality* indicates that a measured outcome is the direct result of a particular event. Causality requires a very rigorous statistical methodology to demonstrate proof, which is usually impractical in the evaluation of DM programs.
>
> *Correlation* implies that the savings were strongly associated with the DM program, but that there's no evidence to demonstrate the program directly caused the outcomes.

For industry users of DM programs, correlation is generally sufficient demonstration of a DM program's value. It would be expected that any study would address and eliminate other potential sources of the measured outcome (for example, plan changes, changes in medical practice, and changes in the underlying risk-pool or the influence of other programs). A soundly structured measurement methodology that incorporates the concepts in this Practice Note may be used to demonstrate a correlation between DM and achieved savings or improved clinical outcomes.

B. Choice of Methodology

Practitioner opinions differ over classification of methodologies. However, it may be useful to determine whether any evaluation study is performed with or without a comparison group. While

it may be possible to construct a valid non-comparison-group study, it is more difficult to demonstrate validity in such a case.

Comparison Group Methodologies

The existence of a comparison group (sometimes called a "control group") is one of the requirements for a valid outcomes measurement methodology (see Wilson, T.W. and MacDowell, M. 2003.) Several types of control groups are frequently used in outcomes studies (see Duncan, I 2005).

- *Randomized control* compares equivalent samples drawn randomly from the same population, in which one group is subjected to the intervention and the control is not. References in the clinical or academic literature to control group studies are often referring to randomized control studies. However, other types of controls exist, as discussed below.
- *Geographic control* compares equivalent populations in two different locations.
- *Temporal control* (also known as the "adjusted historical control design") compares equivalent samples drawn from the same population but at different points in time, specifically, before and after the intervention program.
- *Product control methodology* compares samples drawn from the same population at the same point in time, but differentiating between members who have different products, such as HMO vs. PPO, or Insured vs. ASO.
- *"Patient as their own control" (aka "pre-post cohort methodology")* follows a closed-group (cohort) over time. This method differs from the "temporal" method described above, in which the intervention and comparison populations are re-sampled in each period to ensure equivalence.
- *Participant vs. non-participant studies* compare the experience of those who voluntarily elect to participate in a program with the experience of those who choose not to participate. This method is difficult to justify under most circumstances. The participants represent a group with risk factors potentially different from those of the non-participants; we already know they differ in terms of the important, but immeasurable, factor of willingness to take control of their own health by engaging in a program.

Non-Control Group Methodologies

- "*Services avoided" methodology* compares requested services with approved services. It's commonly used for services in which pre-authorization is sought or where a specific condition is treatable by a specific procedure.
- *External benchmark methodology* compares the experience of the measurement population with that of an external benchmark.
- *Evidence-based methods* combine a benchmark with a "patient as own control" method. The improvement in a patient's clinical measures is recorded and combined with the benchmark financial value of the clinical improvement. The benchmark values are often derived from studies in the literature, and may not directly apply to the population being measured.

Evaluation Considerations

Evaluation requires the comparison of a DM program's actual result with an expected result without the DM program. The study period must account for all factors that potentially affect the population in that period, including trends, risk, and population utilization.

A. Equivalence

A key consideration in evaluating the impact of a DM program is ensuring that the comparison group (for example, "pre-DM period," in a pre/post design) and the intervention group (or "post-DM period" in a pre/post design) are equivalent. This equivalence is necessary to ensure that any variances that occur between the two groups, once the DM program is implemented, are because the program is being measured. In order to maintain equivalence, it is usually prudent to carefully consider all potential confounding factors that might influence the experience of the underlying groups being measured.[2]

- Eligibility requirements
- Demographics
- Risk profiles
- Benefit structure
- Disease prevalence
- Disease Duration
- Member persistency
- Provider contracting

Health actuaries who are experienced at pricing and underwriting are accustomed to evaluating equivalence and the potential impact on financial outcomes of non-equivalence. Many of the techniques used in pricing and underwriting also apply in outcomes evaluation.

B. What to Measure

DM programs influence clinical, utilization, and humanistic outcomes. Other measures, such as reductions in HbA1c scores or the percentage of smokers in the plan, are typically monitored in order to demonstrate improvement due to DM. Analysis of key cost drivers for specific diseases, which involves measurement of clinical outcomes of these diseases, is outside the scope of this Practice Note, which addresses financial measurement only.

C. Financial Outcomes

Traditionally, financial outcomes of a DM program are reported as a ROI. ROI is defined, in this usage, as dollar savings in medical expenses resulting from the DM program, divided by program cost. While ROI indicates savings relative to cost, it does not convey the absolute magnitude of savings. A program may have a high ROI but a relatively low dollar savings, and thus be of lower financial value to a sponsor than a program with lower (but still positive) ROI and higher absolute savings. The use of the program cost in the denominator also makes

[2] A discussion of these factors is provided in Appendix 1.

comparisons between programs difficult, because programs have different cost levels. ROI is used by financial executives and others as a basis for comparing programs; the evaluation of programs, however, requires consideration of a wider range of factors, including assumptions and methodology. Wherever possible, the dollar savings and program cost should be disclosed, in addition to ROI measures.

When services are provided by an external vendor, the denominator of the ROI calculation consists of vendor administration and implementation fees only. The sponsor organization's DM program costs typically include fees paid to an external DM vendor, internal and external implementation costs and fees, and ongoing internal expenses related to the program. Internal expenses are often not tracked and program costs and resulting ROI consider only external vendor fees for program administration. DM vendors use this measure of ROI to compare their fees with estimated savings. The more comprehensive view of program costs may be an important consideration in comparing multiple DM programs' ROI, because the sponsor may have varying levels of involvement in marketing, communications, measurement, and clinical support. Additionally, the level of internal cost may change a program outcome from an acceptable ROI to an unacceptable one.

D. Time Periods for Measuring Financial Outcomes

Because of the shifting nature of enrollment and dropout in U.S. health benefit plan membership over time, financial outcomes are often best assessed in serial 12-month reporting periods. The most comprehensive studies are ongoing -- capturing and examining data at multiple points. When establishing the measurement time frames, the prudent actuary ordinarily considers the attributes noted above and their potential effects on financial outcomes. A widely held theory, as it relates to the effects of DM interventions, is that the longer a member is enrolled in the DM program, the more dramatic the results. In light of these basic attributes and opinions, the actuary may, nevertheless, consider establishing realistic timeframes that capture the materiality of the expected outcome(s). The actuary typically may compare results between 12-month periods. Because of seasonality, it is advisable to avoid comparison between periods of unequal duration. The periods need not be adjacent, but should be reasonably close together. When allowing for run-out, the actuary may choose to test the data for degree of completeness at different numbers of months after the close of the period, and apply completion factors if necessary.

E. Regression to the Mean

Regression to the mean confounds the impact of the intervention and is the most important and challenging issue in accurately measuring DM programs. The basic challenge is simple: Many sick patients get better in the normal course of events as they manage their chronic illnesses, while others do not or even get sicker. Regression to the mean is observed as either an increase or a decrease in individual consumption of health care resources, depending on the individual's initial situation. Within a larger population, the mix of individuals experiencing increases and decreases in consumption will result in the population increasing or decreasing resource consumption. Within a population experiencing natural increases (or decreases) in resource consumption, measuring the additional gain provided by a DM program is challenging.

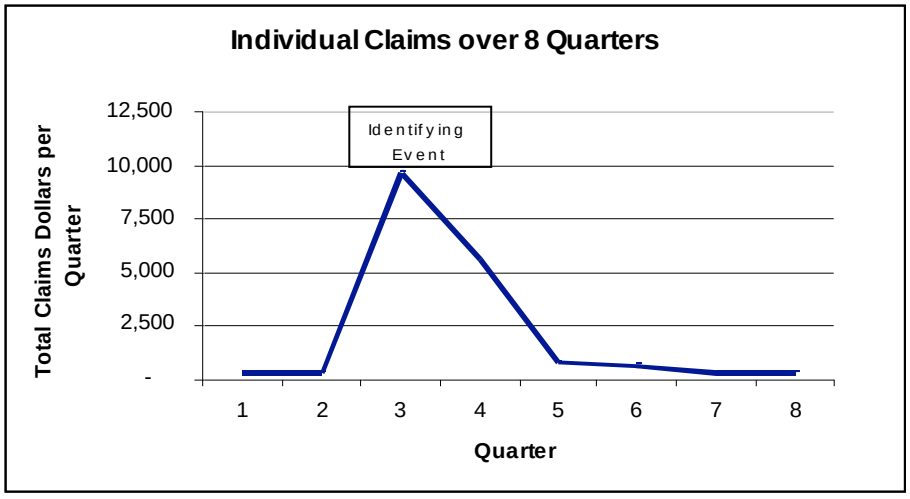
Members are often identified as being eligible for a DM program during a high point of individual medical utilization. Therefore, future costs (e.g., post hospitalization, if the pre-cost

includes the identifying hospitalization) are typically lower because of a return to lower levels of utilization and cost that results from the natural course of treatment and recovery. This change occurs with or without active intervention by a DM program, and is referred to as "regression to the mean." The example above is for a single member; regression to the mean also occurs at the population level. It is possible to select a population that minimizes regression, by offsetting the costs of members whose costs *increase*, against the reduction in cost of those whose costs *decrease*. When comparing groups of members, regression to the mean may bias the evaluation of a DM program, if the way members are identified is not consistent for the two groups.

For several years, regression to the mean was not recognized by some disease program managers, which contributed to the publication of significant first-year cost savings. A typical evaluation method was to compare the per-member, per-month (PMPM) cost of a cohort during the intervention period to the cohort's baseline period experience, after adjusting for trend, where the baseline period members were selected because they were high utilizers. This methodology builds in regression to the mean and therefore overstates savings.

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

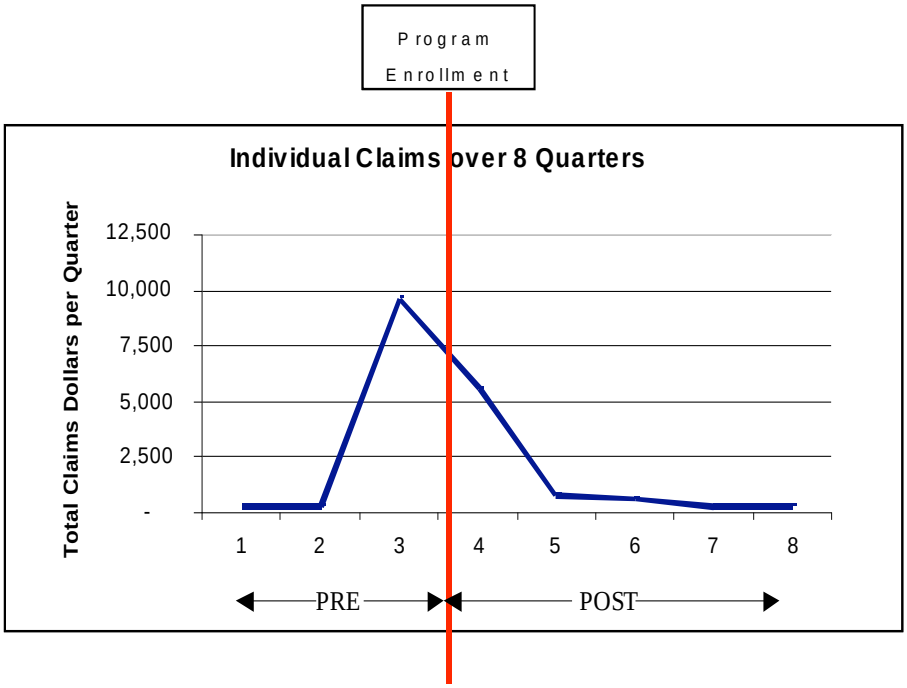
Figure 1



Depending on when this individual's experience begins to be measured, 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, tracking the experience after the program starts will show lower cost. The reduced cost may be incorrectly 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 some time after the identifying event. 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 (illustrated here), regression may be present in a population (although this is not necessarily the case, depending on how the population is identified). A population consisting entirely of individuals identified at the point of hospitalization, as in the example above, will exhibit regression to the mean. A population consisting of members at different stages of their cost cycle may not.

More recent research on DM evaluation shows that an appropriately identified population consisting of members in all stages of their disease, when compared with a population defined by the same criteria in another period, will demonstrate little regression to the mean.

F. Selection / Participation Bias

Selection bias is a statistical error or bias that may emerge from a study as a result of the actual selection criteria used to determine the group studied. Selection bias may cause measurable outcomes to appear more favorable than they actually are. For DM programs, the most common form of selection bias is self-selection, when the study group is composed of individuals who are offered a choice and decide to participate in a program. For such study groups, outcomes may be influenced significantly by the participant's direct interest in the program or outcomes being studied, i.e., either cost or health care resource use. For the same reason, it may be prudent to avoid measurement populations that comprise patients referred by an outside source, unless the members are included in the measurement population because they meet other objective criteria.

Data Sources and Considerations

A. Data Integration Issues

Data often must be aggregated from a number of different sources in order to evaluate a DM program. Traditionally these sources include detailed medical and drug claims data, such as that obtained from health plans or third-party administrators (e.g., data incorporated from CMS 1500s (professional claims) and UB-92s (hospital claims)) and from prescription benefit management companies. Other types of data, such as lab results and self-reported health status, are increasingly being incorporated to more accurately measure the effect of the DM intervention. When multiple data sources are utilized, care is required to ensure that eligibility data (e.g., unique member identifiers) is appropriately coded for each source and that all data sources tie together correctly. For example, pharmacies may inadvertently use the subscriber ID instead of the dependent ID when filling a prescription. If a unique member identifier is not utilized across all data sources, it is important to ensure that data from all sources is appropriately tied back to the correct member.

B. Methodology Application Issues

In order to avoid creating a bias in the DM financial analysis, it is ordinarily prudent to apply the same criteria to both the baseline and measurement periods (or to the intervention and comparison populations, if the design is not an historical comparison), with respect to handling of the data, identification of populations, adjustments to the underlying data, and the decision to include or exclude certain categories of claims. It is usually assumed that the claims data is generated using the same standards and coding practices in each period, although this is not always true (for example, where a health plan changes from capitation to fee-for-service reimbursement). The data is normally subjected to validation tests in order to demonstrate its completeness and consistency.

In order to ensure that the methodology is consistently applied to both the baseline data and the post-intervention data, a variety of issues that can affect the comparison -- some related to the DM intervention and some not -- are typically considered. A list of such issues is provided in Appendix 2.

Trend and Its Role in DM Calculations

A. The Typical DM Measurement Calculation

Although there is no official "industry standard" methodology for calculating DM savings, a majority of evaluations are conducted using a population-based, pre-post approach.

A simplified version of a typical approach is presented below:

1. DM Measurement Population PMPM
   - Calculate baseline year PMPM for members meeting criteria for DM population eligibility.
   - Calculate intervention year PMPM for members meeting criteria for DM population.
   - Adjust baseline year PMPM for DM population to the intervention year by multiplying by an estimate of the trend that would have been experienced had no program been in place.

   Total Population Trend
   - Calculate baseline year PMPM for total membership.
   - Calculate intervention year PMPM for total membership.
   - Calculate cost trend between base year and intervention year for total membership.

In many evaluation studies, the trend experienced in the non-chronic population is used as the estimate of the chronic population's trend.

2. Calculation of Savings
   - Apply total population cost trend to base year PMPM for DM population to produce trended baseline year PMPM.
   - Compare each trended baseline year PMPM for DM population to actual PMPM of DM population in intervention year.

If an estimate of the return on the investment in the program is required, divide total estimated savings by the amount paid to the vendor (plus any internal program expenses).

B. Choice of an Appropriate Trend Adjuster

For the typical DM measurement calculation, which involves comparing population experience in two different periods, a trend adjuster is applied. The use of a trend adjustment is an acknowledgement that PMPM costs in health care change between periods (usually, but not always, increasing). Actuaries are accustomed to measuring PMPM cost and utilization trends and are familiar with this phenomenon. What actuaries may be less familiar with, however, is the differential experience of sub-populations (for example, the chronic population or the diabetic population) within the overall population. Sub-population trends may be different from those of the overall population.

Some practitioners resist the use of a trend adjustment because it often appears to be correlated with large savings estimates. This objection confuses the need for a trend adjustment (which can be objectively established from utilization and unit-cost experience) with the actual trend

adjuster used (which may be inappropriate for the particular population being measured). This section discusses issues involved in choosing an appropriate trend adjuster.

The choice of a trend adjustment is an important decision. There are three common types of trend adjustments that can be considered.

1. *Aggregate trend*. The ultimate goal of those buying health management programs is much lower trends. One basic method applies a trend to the cost from the previous period. The trend used can come from two sources. One source is an external credible benchmark, such as published financial reports (adjusted for benefit design or other changes). The other, for very large populations, is the historical trend in costs over time. Unfortunately, both of these trends are subject to the many challenges described in the rest of the Practice Note. Therefore, DM programs usually use much more complex approaches.

2. Although the aggregate trend is rarely used as the only measure of results, it does provide a check total. Buyers do not accept high ROI estimates from more complex methods if there is not a corresponding reduction in the trend in overall claims.

3. *Price trend*. Under almost all circumstances, prices for services (unit costs) are increasing from one period to the next. Because price changes for services can be measured (either on average or for each specific service), costs from the prior period can be adjusted for price changes.

4. *Utilization trend*. In some cases, there may be demonstrable increases in the utilization of services for particular chronic populations over multiple time periods. If so, such a change in utilization can be reflected in a trend estimate.

Medical cost trends are influenced by many factors that affect cost and utilization of health care services. Using the total population trend to measure DM program impact has the advantage that it captures many of the standard trend considerations including:

- Change in membership: age/sex distribution or health status
- Change in contracted rates or contract arrangements (e.g., fee for service vs. capitation)
- Change in covered benefits (e.g., cost sharing)
- Physician practice patterns
- Catastrophic claims
- Regulatory changes

However, the total population trend does not adjust for disease-driven trend considerations experienced by special DM populations such as:

- Treatment advances: Surgery, biotechnology
- Diagnostic advances
- Changes in screening recommendations

Disease-driven factors can lead to higher or lower trend in health care costs for special populations, leading to possible bias with traditional ROI approaches that use total population trend. In a commercial population, the relatively low prevalence of chronic disease, while a consideration, may not be material. For a Medicare or Medicaid population it is likely to be of more significance.

In choosing an appropriate trend adjuster, the sensible actuary will normally be aware of the possibility of migration bias, a factor that occurs when members migrate from one population to another over time. The extent of this phenomenon will depend on the duration of the measurement and the method of identification of the chronic population. (See Bachler, Duncan, and Juster, North American Actuarial Journal, October 2006.) In addition, the actuary may recognize differential utilization of different services within different populations: for example, chronic populations which, although they are high *absolute* utilizers of drugs, spend relatively less per capita on drugs than non-chronic populations (because they are high utilizers of costly inpatient services). Therefore, consideration is often given to adjusting a trend derived from a non-chronic population.

Operational Issues

Multiple programs
While, for some years, care management programs have been aimed at discretely identifiable populations (e.g., DM, catastrophic case management, end-stage renal disease), the current trend is toward more patient-centric integrated programs (i.e., in which a care manager may direct a number of different interventions to the target population). At the same time, traditional chronic DM is being deconstructed into segments that may be managed by different vendors or a combination of vendor and internal staff. These new models pose significant challenges for measurement.

The typical DM measurement methodology (i.e., historical population control) works well at the population level; it is difficult to apply in situations where multiple programs or vendors may manage a defined group at different times. We are likely to see the application of alternative methodologies to accommodate increasingly complex measurement needs.

Use in Medical Management

a. Identification of disease states that would benefit from DM

b. Identification of treatments that eliminate the need for DM

c. Program structure issues (what they are and how to control for them)

Other Considerations

Validation of Results

As discussed throughout this Practice Note, program financial evaluation is a complex topic. The wide range of existing methods and results creates the potential for major understatements or overstatements of results.

The reported ROI savings for apparently comparable programs ranges widely, with results from breakeven to 7:1 being observed. If the analysis doesn't manage the major problem of regression to the mean, savings estimates will be highly inflated. It is crucial to understand the method for identifying the population and identification of chronic members. The various selection biases must be managed. Examples of practices that lead to inflation of results include:

- Selection of a study population in such a way that regression to the mean is inevitable;
- Bias in the selection of the study population (such as those that elect to enroll only);
- Use of an inappropriate trend adjuster;
- Exclusion of members with no claims from the denominator (e.g., members for whom the plan sponsor paid program fees);
- Published savings results that cannot be reconciled with the overall trend in the population, cost PMPM, or admission rates;
- Failure to remove members that do not continue to meet criteria to qualify as diseased;
- Lack of consideration of changes in the medical care environment, such as movement of drugs to generic status or changes in treatment standards;
- No consideration of difference in the average duration since onset or identification of the disease status between the base and the measurement periods;
- No variation in the trend adjustment assumption by type of service (e.g., inpatient, outpatient, physician, drug);
- (In programs in which evaluation is performed for individual conditions) Migration of members between DM programs between the base and measurement periods (For example, if a member is identified as diabetic during the base period but subsequently develops cardiac disease, how are claims for this member evaluated during the measurement period?);
- Failure to consider the impact of large, outlier claims (Depending upon the population size, large claims may distort results significantly. They may cause results to be overstated or understated, depending upon whether they occurred in the base or measurement period.); and
- Differences in the ability to access the managed population over time (For example, is the number of diseased members with valid phone numbers consistent over the various measurement periods?).

Statistical Validity

Depending on the size of the overall population, the targeted DM population may not be large enough to be statistically valid. In addition, any measure of the financial impact from DM programs is subject to uncertainty. As a result, a definitive determination of the financial impact of a DM program is sometimes difficult. However, actuarial techniques (credibility weighting, for example) provide guidance in this area.

Future Developments

Given the importance of medical costs, there is a continuing transformation in how DM programs are created, implemented, and measured. New systems capabilities and actuarial techniques continue to evolve as well. The actuary working in this area may have to move beyond the historical practices described in this material to meet the new environment. These changes include:

- Active health management programs, such as smoking cessation and obesity management tied to health risk assessments (HRAs) and stronger Web support.
- Strong operational and systems integration of health management, DM, and complex case management in large organizations.
- Increasing use of comprehensive disease registries, based on clinical results such as lab tests. This will change the methods for identifying patients and eliminate some of the measurement challenges over time. Many health plans and other health organizations are also developing comprehensive patient electronic records to make consistent data available over time. These records provide both disease registry and intervention data. At this time, however, no consistent standard has been developed.
- Behavior change and sales techniques are being used to increase participation in DM programs.
- Physicians are attempting to expand their role in health management outside the doctor's office.
- Various states are implementing programs aimed at smoking or childhood obesity.
- New analytic tools, such as episodes-of-care groupers, are just starting to be applied to DM programs.
- Changes in how participants are qualified to participate in the program, such as automatic enrollment versus requirement of consent.
- Level of communication between the DM program, the health plan, and the primary physician, regarding the member's health care.

The actuary needs to understand these changes to provide the much-needed validation and measurement of results.

APPENDIX 1:
Factors That May Be Considered When Demonstrating Equivalence

Equivalence is normally demonstrated both in the chronic population and (if its experience is used to calculate trend or another external benchmark) the comparison population.

(i) Eligibility Requirements

The underlying group may be any group that purchases (or implements) a DM program (an employer group, health plan, or government program, for example). Within this group, a sub-group may be defined, such as all active employees or all members with a certain condition. It is important that eligibility criteria be established for the underlying group and subgroup because, for evaluation purposes, failure to establish clear eligibility can result in confounding or uncertainty in the resulting financial outcomes calculations.

The basis for eligibility of a program will be membership in the sponsoring group. Whatever criteria apply to eligibility for insurance coverage will apply to eligibility for the program; when a member ceases to be eligible for coverage, for example, eligibility for inclusion in the program (and measurement for outcomes) will cease. In addition to coverage eligibility criteria, a program sponsor or provider will often impose more stringent program participation conditions, electing to include, for example, only members aged over 4 or under 65.

On occasion, participation in a program will depend on its purchase by eligible employers. Employer purchase at different times will result in the addition of groups of members at different times, rather than on a common start-date; for program outcomes that depend on a baseline group or cost for comparison, the addition of these members can result in non-comparability between the intervention and baseline populations. Members sometimes become eligible for inclusion in a program on a self-referral basis. A member who has diabetes, for example, may be referred to a diabetes management program by a physician, even though the member does not qualify for the program as a result of satisfying the necessary claim criteria. While these members are generally eligible for services (the management population) they are generally excluded from outcomes measurement (measurement population) because their identification is not objective and they cannot be matched with comparable members in the comparison population.

(ii) Demographics

The group covered and the control group are defined by any factor that could materially affect the results. Typical factors include the following:

1. Age —Studies often focus on a particular age category (e.g., under 21, over 65). If, however, a wide range of ages has been studied, results for broad age ranges are often shown separately.
2. Gender
3. Geography
4. Time Period
5. Benefit and claims payment protocols — Differences that materially affect claims costs are normally identified, be they differences in plan of benefits, the prevalence of other payer liability, fee schedules, or claims precertification.

6. Type of illness affecting people included in the study
7. Portion of the total number of people with an illness who were included in the study — Were all the people with a certain condition included in the study, or only the people who volunteered for the DM program?
8. Number of people covered — Was it a small study with the chance of random events materially affecting the results, or were there thousands of people covered ensuring credible results?
9. Type of participation — Were all the members present for the entire study, or were new entrants allowed in for part of the study period?

(iii.) Risk Profiles

Changes in risk over time for the population being measured can contribute to calculated financial outcomes. The actuary may want to normalize for these changes by applying risk scores (risk adjustment) using a validated risk model (such as one of the commercially-available models). See the SOA risk adjuster study for further details. (http://www.soa.org/research/files/pdf/Risk_Adjusters.pdf)

(iv.) Benefit Structures

Changes in benefit designs can also contribute to calculated savings, either positively or negatively. Some actuaries have found that using allowed amounts rather than paid amounts in the calculations helps to eliminate some of this effect. For example, if there is a shift in benefits, where members take on more of the financial burden of health care costs, allowed amounts in the calculations would account for this.

(v.) Disease Prevalence

A chronic population, whether consisting of a single disease or multiple chronic diseases, will exhibit changes in disease prevalence over time. This disease prevalence may arise from different sources, which the actuary may wish to consider carefully:

- Changes in underlying clinical disease prevalence;
- Changes in identification methodologies (and related claims issues);
- Changes in the composition of the population (e.g., the addition of a large new employer group);
- A change in the severity of the population (for example, increasing co-morbidities); and
- Statistical false-positives (members who met the chronic identification criteria in a prior period but who no longer do so in the current period).
-

Depending on the source of change in prevalence, different methods may be required to ensure equivalence between populations over time.

Different identification rules may be applied to address this problem. DMAA (2006) contains a typical set of identification criteria. NCQA also publishes a set of criteria (HEDIS) that may be considered, although the HEDIS definitions require data that go beyond claims, which makes their application problematic. With any definition, a balance typically is struck between

sensitivity (finding all the members who *may* have the condition) and specificity (excluding those members who are "false positives" and who may be identified through rule-out services).

(vi.) Disease Duration

The severity of a condition and the related costs of treatment may increase with increased duration since initial diagnosis of the condition. Equivalence may be difficult to assess if the mix of members at different duration since first diagnosis differs between populations. This difference may be further confounded by factors such as the existence of a management program that has identified and managed members in the past, and the effect of identifying events, which often result in short-term fluctuations as member costs regress to the mean.

A mathematical analysis that may be helpful to model the underlying processes is the Markov or Disease State (transition) model. While this type of analysis has considerable appeal to the modeling of chronic disease and its co-morbid conditions, few clients will have sufficient longitudinal data to allow a full analysis.

One simple approach that retains some of the durational analysis but is not limited by data availability is to group members in any analysis period into "select" (incident, or those newly-diagnosed) and "ultimate" (prevalent, or those who had prior claims for the condition).

Some more sophisticated analyses are possible and examples are discussed in Appendix 2.

(vii.) Member Persistency

Persistency poses a challenging question in a DM environment. A variety of approaches are being used. As programs evolve from disease management to health management, with a variety of contact points and delivery media (phone, mail, and email), the persistency definition is becoming increasingly complex. There are two primary approaches with a wide variety of underlying details. You can either begin with a closed group of known chronic patients or you can compare a current year of chronic patients to the next year of chronic patients. Each approach has its strengths and weaknesses, including several major potential problems which create an automatic regression to the mean. Under both approaches, the patient is ordinarily traced:

- First, by membership within the insurance product;
- Second, by employment, COBRA, or retiree medical benefits with the employer;
- Third, by enrollment with medical plan, product, medical group, etc.; and
- Finally, by the common population over a specified time frame.

(viii.) Employer Incentives and Their Effect on Participation Rates.

DM programs are effective only if eligible members become and remain actively involved. Employer incentives such as reduction in health care premiums, income incentives, or increased contribution to an HSA or HRA account have consistently been observed to result in greater member engagement in these programs. Engagement is required for generating desired behavioral changes (e.g., improved diet, adherence to a treatment regimen) which, in turn, increases the potential for desired financial outcomes. The actuary might consider, both the cost of providing an incentive, and estimated savings when evaluating the financial impact of DM programs. While costs of implementing, managing, and paying for an incentive program are

immediate, there is generally a lag in realizing financial benefits because behavioral changes are gradual and their impact on health and utilization outcomes will be observed at a future date. In addition, it is prudent for the actuary to remember that incentives will apply to all members (including those who would have participated and been compliant without the incentive), which drives up the cost of the program without producing additional benefits.

APPENDIX 2:

Data Issues That May Be Considered When Constructing the Analysis Data Set

1. High Dollar Amount and/or Untargeted Claims

Certain claims and conditions may be excluded from the DM financial evaluation because they are unlikely to be directly related to the primary disease(s) being managed. It may be useful to perform the evaluation with and without exclusions. Examples of exclusions include trauma, maternity, mental health and substance abuse treatment, and transplants. With certain exceptions (e.g., progression of a chronic condition to transplant, or complications of maternity or substance abuse treatment due to chronic disease), these claims generally do not reflect the costs and utilization associated with the chronic diseases targeted by DM programs.

Because these are often high-cost, low-frequency events, a few such claims can significantly distort the average costs for either analysis period. As discussed above, certain categories of claims (e.g., traumas or transplants) may be completely excluded from the study. When such claims are included in the analysis, care must be taken to account for any changes to either the provider network or the provider contracts. Hospital payment outlier provisions in particular can have a significant impact on total dollars associated with catastrophic claims.

Identification of claims for exclusion may be done in different ways. One method is to simply identify high-cost members with total claims in excess of a defined "outlier" dollar threshold, and exclude them from the DM financial analysis. These outliers may be excluded only for the period over which they accumulated claims above the determined threshold. For example, if a member accumulates claims dollars over the outlier threshold amount in the baseline but not in the measurement period, he may be excluded from the baseline but included in the measurement period.

A more complex method is to identify members who have excluded conditions through the diagnosis and procedure codes available in the claims data. The members so identified, or possibly just the claims directly associated with the excluded condition, are then excluded from the analysis entirely, effective in the month of service of the identifying claim. Other methods might include exclusion of hospice or skilled nursing facility claimants, or exclusion of members who have been institutionalized for more than an agreed-upon time frame (typically 30 to 60 days).

A DM financial evaluation may focus on non-surgical admissions, as surgical admissions are often for elective or other non-elective procedures not related to the DM process. However, the decision as to which classes of surgical admissions to exclude ordinarily will depend on the chronic condition(s) being evaluated. For example, exclusion of gastric bypass surgery may be appropriate because related claims would inflate utilization and cost estimates and possibly skew interpretation of DM interventions. Conversely, exclusion of coronary bypass surgery may be inappropriate because it is desirable to include all costs associated with cardiac-related procedures in a cardiac DM program.

Admissions related to the managed chronic condition may be the primary focus in order to gauge whether the management program is improving the health of its members. However, this ignores other conditions that may be aggravated by the disease being measured. Focusing only on chronic admissions may also contain some self-selection bias. Further, conditions that have no obvious relationship to the targeted condition may be related to, or complicated by, the chronic condition. For example, some research suggests that cardiac ischemia may cause back pain, so it would be incorrect to exclude an admission for DRG 243 "Medical Back Problems."

Cost and utilization might also be measured by considering all admissions regardless of diagnosis. This will include admissions directly related to the target disease, as well as any other condition that may or may not be caused or aggravated by the disease. The disadvantage of this approach is that some conditions (e.g., maternity or trauma) may be included that have no direct relationship to the disease being studied, which may distort the results of the analysis. An advantage may be that the inpatient analysis will have more robust numbers to evaluate. A larger analysis data set may imply greater credibility of results. In addition, this method does not require a subjective determination of what conditions to include and exclude, but automatically reflects the impact on cost and utilization of any co-morbid conditions.

2. Changes in Provider Contracts

Changes to provider contracts often result in higher allowable charges for services, such as the annual update to the resource-based relative value scale (RBRVS). Such straightforward changes can be modeled through standard trend assumptions. The actuary may want to review the contract issues affecting the claims and, if necessary, apply other actuarial adjustments, such as exclusion of certain capitated encounters, or adjustments that allow for a change in reimbursement from fee-for-service to DRG code.

3. Payment for Capitated Services Reflected in Claims Data

Because capitated providers are not paid based on services rendered, it is often difficult to obtain adequate encounter data. This will affect the calculation of claim frequency as well as estimated fee-for-service equivalent costs. Efforts to improve the encounter data, or changes to the schedule of capitated services or the volume of membership under capitation, can all increase the difficulty of isolating the effects of the DM intervention.

4. Inclusion of Lab Data

Laboratory and radiology claims are often included in analysis of DM financial results. However, caution is advised when one is relying on such data to determine whether a member has been diagnosed with a particular chronic condition. Because such data will include tests for particular conditions that may prove negative, laboratory and radiology claims normally will only be used to assign members to chronic disease categories when test **results** are consistently available for both the baseline and the measurement period.

5. Changes in Average Length of Stay (ALOS)

A reduction in average length of stay may indicate that a DM intervention is reducing the severity of the condition of members admitted. However, changes to the ALOS may be due to other factors such as changes in the hospital reimbursement structure discussed above. In addition, a DM intervention may reduce the number of admissions, while the average severity of those admitted (as measured by ALOS) remains relatively constant. In this case, a comparison of overall inpatient days per 1,000 may be more appropriate.

6. Definitions of Re-Admission

Any comparison of the number of admissions per 1,000 requires a consistent definition of when an admission is considered to be a re-admission versus a new admission.

7. Adoption of New Technology

New technology is providing additional sources of data. As any particular technology is implemented, however, the available data may be inconsistent and thus not directly comparable due to different data sources or different time periods. For example, electronic medical records may significantly improve the ability to monitor and measure a DM intervention, but these systems are still under development and their use is inconsistent among providers. Other technology may mean that data elements are provided from a different source or are no longer captured. A1C tests, for example, have typically been captured from lab data. However, newer office- and home-based versions of the A1C imply that members are no longer sent to the lab for these tests. Thus, lab data between different periods would not be comparable.

8. Data Quality

Because medical and drug claims are the primary data sources for the analysis, any inherent data limitations will have an impact on the ability to isolate the effects of the DM intervention. A common limitation results from provider capitation arrangements. Because capitated providers are not paid for each service rendered, utilization data is frequently under-reported and difficult to compare. Even with full claims data, information that could improve the accuracy of the comparison may not be available. For example, the UB-92 (hospital claim) form does not provide certain information (such as data on co-morbid conditions and demographic characteristics) as would be necessary to completely assess the risk profile of members admitted during the analysis period. Thus, it is difficult to determine whether the intervention improved the risk profile of the membership, or if other changes or random variation caused any perceived differences in the experience.

In any analysis, a clear understanding of the data is advised before any meaningful results can be obtained. This is particularly true with respect to utilization measures. If inpatient days are being calculated, the claim's header date is typically used for the calculation of length of stay (discharge date minus admit date), rather than the claim's line-level dates. Use of the header date eliminates double counting of days for a particular inpatient stay, as may occur due to interim billing and line-level dates. A similar understanding of the data is prudent with respect to utilization data for other service categories. For example, when a physician provides lab or

radiology services during an office visit, these services may be bundled into the one office visit encounter (even though other physicians may need to refer the member to an outside lab), or they may be segmented and counted separately.

The same amount of claims payment run-out is usually applied to both the baseline and measurement periods in any DM financial analysis. In general, a run-out period of six months for both periods will ensure that the data is mostly complete. If more recent incurred data is desired, or if claims payment patterns are relatively slow, completion factors may be used to improve the comparability of the data between the periods.

Denied claims typically are not included in a DM financial analysis. Knowledge of the health plan's claims payment system, and of how reversals are handled, will facilitate the correct accounting for claims that were originally paid and then reversed.

APPENDIX 3:
Relevant Actuarial Standards of Practice (ASOPs)

ASOP No 5, *Incurred Health and Disability Claims*
http://www.actuarialstandardsboard.org/pdf/asops/asop005_076.pdf
This standard gives guidance to actuaries preparing or reviewing financial reports, claims studies, rates, or other actuarial communications involving incurred claims within a valuation period under a health benefit plan. The standard applies when a study of the impact of DM requires the quantification of incurred claims.

ASOP No. 8, *Regulatory Filings for Rates and Financial Projections for Health Plans*
http://www.actuarialstandardsboard.org/pdf/asops/asop008_100.pdf
This standard sets forth recommended practices for actuaries involved in the preparation and/or the review of regulatory filings for health plans. This ASOP may apply if the impact of DM techniques on medical expenses affects the regulatory filings.

ASOP No. 19, *Appraisals of Casualty, Health and Life Insurance Businesses*
http://www.actuarialstandardsboard.org/pdf/asops/asop019_099.pdf
This standard gives guidance to actuaries who perform professional services with respect to appraisals of casualty, health, and life insurance businesses.

ASOP No. 23, *Data Quality*
http://www.actuarialstandardsboard.org/pdf/asops/asop023_097.pdf
This standard provides guidance relating to the quality of quantitative information used by the actuary. Good data quality is imperative when measuring the actual or potential effectiveness of a DM system.

ASOP No. 25, *Credibility Procedures Applicable to Accident and Health, Group Term Life and Property/Casualty Coverages*
http://www.actuarialstandardsboard.org/pdf/asops/asop025_051.pdf
This standard provides guidance to actuaries in the selection of a credibility procedure and the assignment of credibility values to sets of data, including subject experience and related experience. This ASOP may be relevant if the DM mechanism is being applied to a population too small to be measured effectively.

ASOP No. 31, *Documentation in Health Benefit Plan Ratemaking*
http://www.actuarialstandardsboard.org/pdf/asops/asop031_060.pdf
This standard provides guidance on documentation of ratemaking for health benefit plans. This ASOP would apply when the impact of a DM mechanism is being considered in setting premium rates.

ASOP No. 41, *Actuarial Communications*
http://www.actuarialstandardsboard.org/pdf/asops/asop041_086.pdf
This standard gives guidance to actuaries about written, electronic, or oral actuarial communications. In general, this ASOP covers the context and purpose of all the

communications between the actuary and the health plan . The most likely shortcomings would be in the area of clarity and completeness, where the communication will encompass health care as well as actuarial aspects.

ASOP No. 42, *Determining Health and Disability Liabilities Other Than Liabilities for Incurred Claims*
http://www.actuarialstandardsboard.org/pdf/asops/asop042_091.pdf
This standard gives guidance to actuaries determining health and disability liabilities other than liabilities for incurred claims. It is relevant to DM only in connection with the analysis of the financial impact of the DM mechanism. The ASOP emphasizes the importance of understanding the contractual basis on which the plan operates and how this is reflected in management and accounting reports that provide the source data an actuary uses. Again, if the DM company assumes some risk that the projected cost impact will be realized, that risk-sharing will ordinarily be considered in performing financial projections,

References

1. Wilson, T.W. and MacDowell, M. "Framework for Assessing Causality in Disease Management Programs," *Disease Management*, Fall 2003.
2. Duncan, I. "Evaluating Disease Management Savings Outcomes," Paper 5 in the Society of Actuaries series on Care Management Interventions, available at www.soa.org).
3. Bachler, R, Duncan, I, and Juster I, "A Comparative Analysis of Chronic and Non-chronic Insured Commercial Member Health Care Trends," North American Actuarial Journal, October 2006.
4. Duncan I. (ed). "Dictionary of Disease Management Terminology (2nd edition), DMAA 2006.