AN INTRODUCTION TO CARE MANAGEMENT INTERVENTIONS
AND THEIR IMPLICATIONS FOR ACTUARIES

Ian Duncan, FSA FIA FCIA MAAA

Introduction

Unlike all other papers in this series about care management interventions and their implications for actuaries which are analytical, objective and based on research, this, the last paper in the series, gives me an opportunity to make some personal reflections on the state of care management evaluation.

When I began this research in 2003, I had little idea of the scope and duration of the work that it would involve. More than three years is a long time in which to be engaged in a single project, albeit part-time and with the assistance of many volunteer assistants, co-authors, reviewers, and others along the way. In total the study has generated 8 papers and about 100,000 words. Many practitioners, both actuaries and non-actuaries, have downloaded one or more of the papers from the Society of Actuaries Web site, and have used some of the principles we have discussed in their work. Some of the terms that we have coined in the course of the study, for example "migration bias", have found their way into day-to-day discussion of disease management (DM) outcomes. The popularity of the papers vindicates the Society of Actuaries Health Section's and the Committee on Knowledge Extension Research's decisions to support the research.

The state of the union

At the same time that we have been engaged in the study, the world of DM has not stood still. Others have written and researched on similar topics. It is interesting to consider what has been achieved and what has *not* been achieved in the last three years. The Disease Management Association of America (DMAA) published "Guide to Disease Management Outcomes Evaluation" in 2004. Far from establishing once and for all methodology and principles to be followed by practitioners, it is widely-agreed, including I believe by DMAA, that the guide falls short of the needs of the industry in this area. Accordingly, DMAA has convened another work group in 2006 to tackle the subject again. The findings of this work group, entitled "DMAA Outcome Guidelines Report" were published in December 2006. Because it is an industry consensus document, the DMAA work group report makes a number of recommendations that readers of these papers will be familiar with. In addition, the guidelines identify a number of potentially controversial issues, many of which are deferred for future consideration. While these guidelines may help practitioners, DMAA, as the industry trade association, will always be perceived by purchasers as representing an industry viewpoint. The professional actuarial associations, Society of Actuaries and American Academy of Actuaries, on the other hand have a reputation for being objective. Recommendations from the professional actuarial bodies, therefore, will carry more weight, particularly given the increasing involvement of actuaries in the performance and review of studies.

The American Academy of Actuaries released its paper "Disease Management Programs: What's the Cost?" in 2005, and has been working for several years on a

1

Practice Note for actuaries practicing in the field. Its release is not imminent, and the slow pace of progress indicates that this is not an easy topic in which to gain consensus.

Some peer-reviewed papers have been published. Ron Goetzel and others published a review of the peer-reviewed literature ("Return on Investment in Disease Management: A Review"; Health Care Financing Review, 26(4) Summer 2005 1-19). Ariel Linden, a well-known researcher in this field, has published a paper that has attracted considerable attention ("What Will It Take for Disease Management to Demonstrate a Return on Investment? New Perspectives on an Old Theme" American Journal of Managed Care, 2006, 12:217-222). This paper addresses what the author calls "number needed to treat", and which may also be called (as we do in Paper 4 of this series) the economics of DM. In addition, this paper draws attention to the need to identify a causal relationship for any savings estimated or measured, particularly tying such savings to the underlying inpatient admission experience of the population (where the major portion of savings are to be found).

With the exception of these papers, the literature has not seen many peer-reviewed publications. What is particularly puzzling is the absence of practical papers that examine the biases in measurement and the impact that these have on outcomes which is similar to the approach we take in Paper 8 of this study.

The papers that were written and published early in the life of the project have not been revised to take account of newly-published articles. The good news, however, is that (as suggested above) there isn't a large amount of relevant research that has been published in the past 3 years, so the "state of the union" after 3 years of research isn't materially different to what it was when I began this project in 2003.

Over the last three years there has been an increase in actuarial involvement in DM outcomes studies and audits. The fundamental building blocks of studies, rigorous reconciliation of data and understanding of per member per month costs and trends, lend themselves to analysis by actuaries. We also suggested in Paper 7 of this series that a relatively new technique in the actuarial arsenal, but one gaining wide acceptance – risk adjustment – also has a role to play in ensuring equivalence between populations.

To the extent that actuaries enter what was previously an area dominated by other health professionals and make a contribution, this study will have made a contribution.

What have we learned along the way?

This section is a summary of some of the key conclusions from the research.

1. The most important objective in any care management outcomes study is to ensure comparability between the intervention and comparison populations. The existing DM literature tends to encourage a belief that there are two "threats to validity" in studies: selection bias, which will be observed when participants are compared with non-participants, and regression to the mean. But as we showed in Paper 6, regression to the mean is an *individual,* not a population concept. As discussion throughout our research suggests, the identification and correction of regression to the mean is a much larger and

more complicated issue than some of the literature suggests, particularly when definitions of who is included in a population may not be clear.

2. Our first two papers in the study were essentially background papers for those with little exposure to the topic. Paper 3 summarized a considerable body of literature on a number of care management interventions including DM, and provides readers with a useful summary of the range of savings claimed in the peer-reviewed literature.

3. Paper 4 presented an important concept – the economics of DM. One of the questions often asked about DM savings outcomes is whether these are plausible; application of a simple economic model to the underlying population will give users a range of likely outcomes and the sensitivity of those outcomes to different program components. More importantly, understanding the key variables of the financial model and their contribution to the overall savings calculation will allow analysis of individual variables that can be directly measured (for example, the enrollment rate).

4. Population studies, the most popular form of study in DM evaluation, may achieve comparability if the populations being studied do not change much from period to period. A major challenge for actuaries is to demonstrate this stability. Fortunately, actuaries understand the issues involved in ensuring comparability over time and the implication for per member per month costs when comparability is not achieved. Actuarial tools such as risk-adjustment exist that make assessment of risk profiles over time and demonstration of equivalence simpler.

5. As discussed in Paper 5, the actuarially-adjusted historic (pre-post) design, which is the most prevalent in the industry, offers a reasonable compromise between validity and practicality. Many would wish to use more scientifically-pure methods, but, as we discussed, these are seldom achievable. Instead, the popularity of the actuarially-adjusted historical control method in the industry is testament to the fact that a well-executed study is viewed as being reasonably reliable.

6. While the fundamental methodology does not vary much between practitioners, the assumptions and methods used to deal with data issues do vary considerably. Definitions matter. We covered in Paper 6 many of the issues that are usually considered in a study – exclusions, inclusions, timing, etc.

7. We discussed some of the issues that arise in measurement in Papers 7 and 8. The methodology, we recognize, is not perfect. With more research into its biases and publication of actual results, we would perform better and more accurate studies.

8. Paper 7, published in stand-alone form in the North American Actuarial Journal in October 2006, highlights the issue of chronic identification and its impact on chronic prevalence and trends. In a population study the issues of *what* claims codes identify a chronic population, *when* those codes have to be observed, how frequently and over what time period, are crucial. As an industry we have only begun to scratch the surface of these issues, but it is probably the single most important issue for the industry to focus on in the future.

9. It is important to understand the impact or "value" of different assumptions on the final results of a study. It is surprising to me that most of the discussion in the DM literature remains at a theoretical level when many practitioners have access to data sets and could simply test out some of the issues that they debate. The industry would benefit from it. It would make the current methodology more robust and would reduce the need for the

industry to search for alternative methodologies. In Paper 8 we examined some of the sensitivities of the results calculated using one such methodology for one client, under different assumptions. Much more of this type of analysis needs to be published, to gain knowledge about the methodology.

Recent Developments in Methodology

Medicare Health Support
One important recent development in care management (although not directly in methodology) was the introduction in 2005 of the Medicare Health Support program. This program applies to Medicare fee for service members with diabetes and/or heart failure. Section 721 of the Medicare Modernization Act of 2003 (MMA) (the same act that brought us Medicare Part D coverage for prescription drugs and expanded accessibility to health savings accounts) authorized development and testing of voluntary chronic care improvement programs, now called Medicare Health Support, to improve the quality of care and life for people living with multiple chronic illnesses. The Centers for Medicare and Medicaid Services (CMS) awarded eight different programs to disease managers in different regions. (One vendor has subsequently withdrawn from the program.) Assessment of program results involves a comparison of the managed population with a randomized control group. This program will finally provide the industry with the answers to two questions:
1. Does care management "work" (that is, produce a statistically-significant difference in financial and clinical results in the managed population)?,
2. Potentially more important, how do the financial results measured by the randomized control methodology differ from results measured by a standard industry methodology (such as the actuarially-adjusted methodology described in Paper 6)? While this comparative analysis is not part of the program, many researchers are anxiously awaiting the opportunity to perform just such a comparative analysis.

Plausibility
Practitioners have also contributed to advances in outcomes measurement, although the techniques have not been published in the peer-reviewed literature. Al Lewis, president of the Disease Management Purchasing Consortium International, recommends the use of what he calls "Plausibility Factors". These factors are not a substitute for a calculation of savings, but rather, a method for accounting for causality (See Paper 5) that would make the calculated savings possible ("plausible"). Plausibility analysis requires the calculation of the following statistic (the plausibility factor) for the entire health plan:

Disease-specific Admissions/1000 (Program Year)
Disease-specific Admissions/1000 (Baseline Year)

A reduction in the plausibility factor is cited as evidence that the program has achieved its goals and lends support to any calculated savings, because the reduction in admissions per 1000 for a specific condition reflects an attempt to identify just the issue that the program should be affecting.   Comparison between the admission rates for the entire health plan population avoids issues of chronic identification and migration bias, which we discuss in Paper 7, although it introduces other potential errors due to population changes over time. While the use of condition-specific admissions makes sense theoretically, the practical issues with claims-coding and primary and secondary diagnoses make us uncertain as to
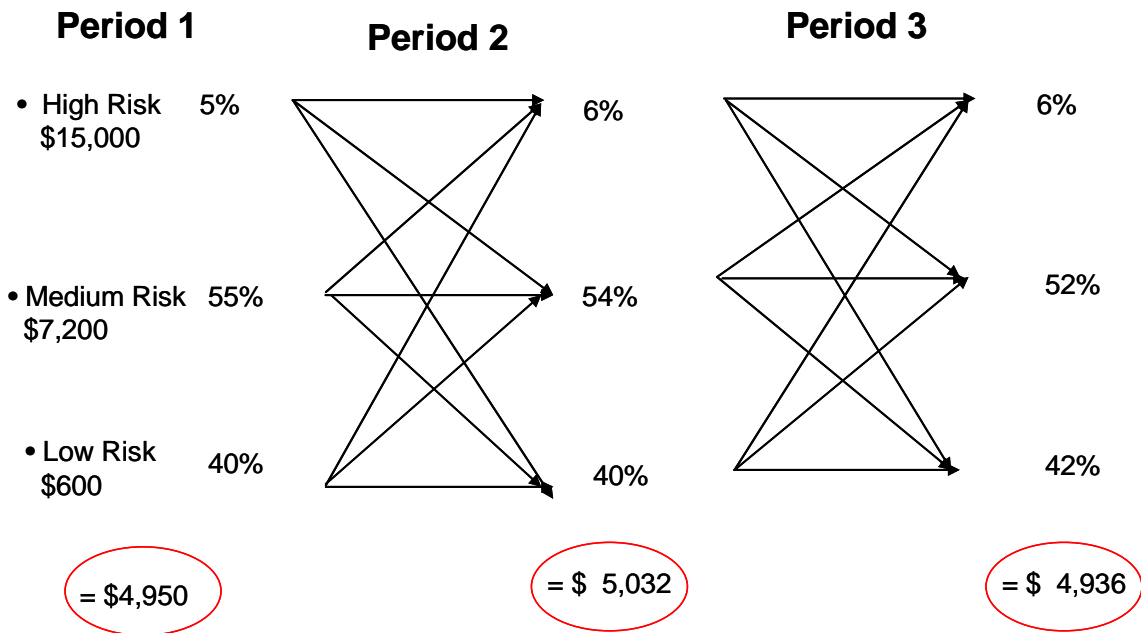
whether the application of this method, without adjustments, is sufficiently free of bias to make the method accurate.

Duration Adjustment
A technique that is promoted among some practitioners is the use of duration adjustment. This technique recognizes that utilization and cost of chronic patients changes over time, often increasing as the patient becomes older or acquires other conditions. In a typical DM evaluation (as we have noted in the course of this study) the chronic population "matures" between the baseline and intervention period simply because one more year passes between the baseline and intervention period. Maintaining the same mix of "duration" is thus difficult. In a typical DM evaluation, only three years of data are available, making it difficult to analyze duration since initial diagnosis. Nevertheless, it is possible to separate "select" (incident, or newly-diagnosed members) from "ultimate" members (those identified with the condition before the baseline period), analyze these members separately, and apply a weighting to the results of each group, when the mix changes significantly.

Transition States (the Markov Model)
Transition states are similar to a select and ultimate analysis, except that movement is allowed both upward and downward. The following diagram illustrates the issue:

Period 1   Period 2   Period 3

- High Risk   5%   6%   6%
  $15,000

- Medium Risk 55%   54%   52%
  $7,200

- Low Risk   40%   40%   42%
  $600

= $4,950   = $ 5,032   = $ 4,936

In this example, for simplicity, we assume a closed group, with members transitioning between states but neither leaving nor entering the group. Chronic members are distributed between High Risk, Medium Risk and Low Risk groups in Period 1, each of which has its own average cost per member per year. Different transition probabilities exist, tracking the tendency of members to transition to other states in the following year. Others who have used this type of analysis have expanded the number of states to include disease free, formerly diseased (those members who meet the identification criteria in one year but fail to qualify the following year) and newly-diagnosed. Termination is obviously another state not considered here for simplicity. It is possible to demonstrate that, under certain transition probabilities the system converges rapidly and results in a stable

distribution by state. A completely stable population will result in constant per member per year costs (absent any underlying trend effects such as changes in utilization or unit cost increases). In our example, year 2 costs increase by 1.7% from year 1 because of the change in risk-mix or profile, while year 3 represents a 2% reduction from year 2 costs, again because of change in risk-profile. In a typical DM outcomes study, this change in risk-profile will not be isolated and will, instead, contribute to the overall measured trend in the population.

The transition probabilities and time to convergence lends itself to further analysis and could help us to understand how closely any one-year state is to a stable state. Probabilities that indicate slow convergence of the model would lead one to conclude that the underlying hypothesis (that the baseline population is a reliable control for the intervention population, except for the utilization and unit cost trends) should not be accepted.

While no work has been published (to my knowledge) in the disease management field on this topic, it represents a potentially fruitful area for actuarial analysis.

Where to from here?

More work needs to be done to understand some of the areas we analyzed, and those discussed above. Other areas for future research include:

Chronic Identification
In Paper 8 we consider the effect on the measured results of changes in the way chronic members are identified. In Paper 7 we also demonstrated that *when* the member was identified as chronic, it can have a significant effect on trend, and thus, on the estimated savings from a program. Understanding the impact on a study of these issues is not just an actuarial task and will require involvement of clinical and actuarial researchers.

Transition States
I have discussed some of the implications of a transition state model above. If we understood chronic members' propensity to change states (particularly as their disease condition matures over time) we could perhaps do a better job of analyzing how and whether an intervention has changed that propensity.

What "works" in care management?
Those of us who are practitioners in this area have been focused, because of the needs of our employers and clients, on assessing the impact of a program, particularly on financial outcomes. This focus has often been on program results at the expense of attempting to discern the impact of different types of intervention within sub-populations. For example, a typical disease management program may include different types of interventions delivered to many different types of members (conditions; co-morbidities; level of severity and risk). Programs often co-exist, within a health plan, with case management interventions that apply yet more intensive management to a member's problems. My prediction for care management in the future is that we will see fewer, more intensive interventions targeted at smaller chronic populations, within integrated programs that include both intensive case management and broader population management (or wellness). This trend will increase our need to know what works, with whom. It will also increase the need

for more accurate predictive models to be able to identify those members who match the "target" profiles.

A "standard" methodology

The DM industry has struggled and failed for a number of years to agree on a standard measurement methodology. By default, however, most evaluations tend to be performed using a variant of the actuarially-adjusted historical control methodology. Given that a large percentage of industry evaluations are performed using a similar methodology, with variation being in the details (chronic definitions; timing; exclusions and inclusions), I have suggested above that a more potentially useful expenditure of the industry's resources would be in understanding the impact on the measured results of these definitions, as a pre-cursor to developing a common set of definitions. The industry has for too long struggled to respond to the demand for an absolute result (how much was saved), a problem that may soon be answered by the Medicare Health Support program, rendering industry efforts redundant. Instead, the industry should borrow a leaf from the National Council for Quality Assessment (NCQA) book and develop a set of measures *together with standard definitions* that health plans and those performing interventions could produce that would allow comparisons to be performed. I do not think that any user of NCQA's HEDIS measures would necessarily believe that these are an *absolute* measure of health plan quality, or, for that matter, that they are the only measures of health plan quality. But the measures, imperfect as they are, have the advantage of being standardized, produced by all health plans, and therefore comparable. The DM industry could perhaps learn from the experience of NCQA and develop similar measures (and definitions) that would allow valid comparisons between programs and vendors.

And in conclusion….

Thank you for your interest in this subject, and our study. I look forward to actuaries taking a prominent role in this, as in other aspects of managed care.