PREFACE

This book grew out of a series of papers that was developed with grants from the Society of Actuaries Health Section and the Society of Actuaries Council for Knowledge Extension and Research, together with additions from other actuaries and researchers in the field of measurement of chronic care and other care programs. The author is grateful for the peer review and valuable input on the original papers by the project oversight group:

Bryan Miller, FSA, MAAA, Blue Cross Blue Shield of Kansas City (Chairman)
John Cookson, FSA, MAAA, Milliman USA
Stacey Lampkin, FSA, MAAA, Anthem, Inc.
John Stark, FSA, MAAA, Anthem Inc.
Margie Rosenberg, PhD, FSA, MAAA, University of Wisconsin-Madison
Ronora Stryker, ASA, MAAA, Society of Actuaries
Stephen Siegel, ASA, MAAA, Society of Actuaries

We are also grateful to the Society of Actuaries for their permission to use the original copyrighted papers as the foundation for this book.

The valuable assistance, support and input from the Society of Actuaries Health Section (Karl Volkmar, FSA, MAAAA, Chairman) and the Committee on Knowledge Extension Research (Curtis Huntington, JD, FSA, MAAA, University of Michigan, Chairman) is also acknowledged. Also, a number of reviewers have read earlier drafts of the manuscript and their suggestions are gratefully acknowledged.

Sheryl Coughlin, PhD
Henry Dove, PhD
J. Ramsay Farah, MD, MPH, FAAP, FACPM, CPE, CMRO,
Stephen Siegel, ASA, MAAA

One reviewer deserves special thanks. Iver Juster, MD, ActiveHealth Management Inc., diligently reviewed the entire text and his thoughtful contributions are reflected throughout.

A number of co-authors and research assistants have contributed to different chapters.

Chapters 2-4: Henry Dove, PhD, Yale University, Division of Health Policy and Administration

Chapter 4: Koren Odierna, Yale University, Division of Health Policy Research and Administration Assistants Elizabeth Heckinger, UCLA

Chapter 9: Mita Lodh, MBA, PhD, FSA, MAAA Schrammraleigh Health Strategy, Gregory Berg, PhD, and David Mattingly, BA, McKesson Health Solutions.

Chapter 10: Thomas Messer PhD, ASA, MAAA, Solucia Inc., J. Ramsay Farah, MD, MPH, FAAP, FACPM, FASAM, CPE, CMRO, Kyahn Kamali, MHS, CPHIMS and Jeffrey Harner, MS, Nationwide Better Health Inc.

Chapter 11: Robert Bachler, FSA, FCAS, MAAA, Munich Re Healthcare and Iver Juster, MD, ActiveHealth Management Inc.

Chapter 12: Rebecca Owen, FSA, MAAA, Solucia Inc.
Henry Dove, PhD Yale University, Division of Health Policy and Administration

Chapters 13-14: Sheryl Coughlin, PhD MHA, B. App. Sci., Solucia Inc.

Chapter 15: Christian Birkmeyer, MS, Solucia Inc.
Manjula Singh, MA, PhD, Solucia Inc.
Appendix: Vincent Kane, FSA, MAAA, DxCG Inc.
Anju Joglekar, PhD, DxCG Inc.

Gail Hall, FSA, MAAA of ACTEX Publications should be commended for taking a risk on an unknown author and a topic that is still outside the actuarial mainstream. I am grateful to Marilyn Baleshiski for her help in editing and laying out the manuscript.

Readers familiar with the Society of Actuaries papers will recognize that they form the basis for Chapters 2 through 12. Chapters 5 and 10 are, however, entirely new material. We have added new material on wellness and employer worksite health programs, a growing, related, area, in Chapters 13 and 14.

A note on terminology: this book purposely aims to cover the broad topic of Care Management Interventions. As discussed in Chapter 2, there are many different types of interventions that may be included under this broad title. Later in the book, we focus specifically on one type of intervention, Disease Management. This is for several reasons, including the significant interest on the part of actuaries and purchasers in this type of intervention. Despite this focus, the techniques and methods discussed in later chapters may be applied (with appropriate modifications for circumstances) to different types of interventions.

We also acknowledge the contributions of Munich Re Health Care, Highmark Inc., McKesson Corporation and Nationwide Better Health Inc., without whose support the analytical work would not have been completed.

INTRODUCTION

Howard Bolnick, FSA , MAAA, HONFIA

Health actuarial practice has been a growing and dynamic part of the profession for many decades. When I began working in the area in the 1970s, indemnity insurance was the only game in town: data were quite limited, and actuarial tools were basic. All this changed with the advent and growth of managed care. There were new types of insurance arrangements that required more robust data, and, new actuarial tools were needed to successfully manage a growing variety of managed care plans, each with its own distinct characteristics and its own distinct health actuarial needs. The profession successfully responded to these changes. Health actuaries broadened their scope of practice and developed the tools and analyses needed to successfully support this new environment. Health actuaries remained leaders in a changing environment by using our background, unique skills, and creativity to become a key resource in the success of managed care plans.

Throughout its existence, managed care has continued to evolve. Older prescriptive approaches to care management are being replaced with newer supportive approaches to population health management such as disease management programs and wellness programs. As these new types of care management evolve, they clearly pose a new challenge to health actuaries. To sustain our leadership role, health actuarial practice will need to continue to broaden its scope and create new tools to support the changing environment.

Ian Duncan's new book, *Managing and Evaluating Healthcare Intervention Programs,* continues this tradition of actuaries responding to a changing environment. This book is a thoughtful, well written, and well-researched study that provides actuaries, senior managers, financial managers, and others interested in the topic with a wealth of information, careful analyses, and a strong intellectual basis for expanding actuarial and financial leadership to population health management.

3 ACTUARIAL ISSUES IN CARE MANAGEMENT EVALUATIONS

3.1 INTRODUCTION

This chapter 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**, addressing basic principles that should be considered in any evaluation; **Study Design Issues**, exploring issues that arise when assessing or planning a study; and **Risk Factors**, covering factors that influence the inherent risk in a population being managed and which therefore influence the measured outcomes.

As managed care has evolved, actuaries have tended to perform their 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. Much of the debate in care management evaluation concerns methodology, but this is just one of a larger set of issues concerning the validation of financial outcomes.

We believe that in the future actuaries will be involved in three important areas: the economics of care management programs, risk adjustment and predictive modeling, and financial outcomes evaluation.

3.1.1 The Economics of Care Management Programs

One factor common to the seven care management programs described in the previous chapter 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 will be discussed in detail in Chapter 4), 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. We will address this issue in more depth in Chapter 6.

3.1.2 Risk-Adjustment and Predictive Modeling

Risk-adjustment and predictive modeling are processes for comparing different populations, providing insights on 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 professional actuarial literature[1]. Predictive modeling is the convention used in Disease Management (DM) to identify, categorize or prioritize 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 self-management with the aid of telephonic intervention). Interest is growing in using risk adjustment or similar techniques (for example, propensity scoring) in the process of the assessment of outcomes. This topic is discussed in more detail in Chapter 7.

3.1.3 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 are discussed in detail in Chapter 7. As important as the choice of a methodology, however, are the adjustments made to achieve comparability between the reference and the intervention population. (Throughout this text the terms "reference population" and "comparison population" or "comparison group" are used interchangeably.) Many of the issues 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 (Weingarten et al. [209]). A similar survey of financial outcomes found mixed results (Krause [111]). Since it is an axiom of the managed care industry that "higher quality" leads to lower cost [2], the apparent inconsistency in these two studies should be of concern to all who work within the care management industry, and requires further analysis.

[1] See for example, Cumming et al. [35], Duncan, Dove et al. [47] and Duncan & Robb[53].

[2] See for example, IOM [34] and Gingrich [71].

Evaluating intervention programs has proved to be difficult. Unlike in clinical outcomes, where one can measure clinical improvement, what is being measured in financial outcomes 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. Health plans and other healthcare professionals generally believe that conducting randomized trials is impractical or even illegal, with the result that 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.

3.2 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 were proposed in a paper by Wilson and MacDowell [219]. We have added three other principles of our own that we have found to be equally important in practical applications (numbers 3.2.4 through 3.2.6 below).

3.2.1 Reference Population

Any outcomes measurement requires a reference population against which to evaluate the statistic(s) of interest, even if that reference population is the intervention population at an earlier point in time (pre-program).

3.2.2 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 later.

3.2.3 Consistent Statistics

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

3.2.4 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). The medical and surgical admissions of all patients will however, 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. By defining as narrow a population as possible, and as narrow a set of outcomes as possible, the effect of confounding will be reduced.

3.2.5 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 defined 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. In our (practical) experience performing care management evaluations, many of the problems that arise in studies do so because of difficulty in defining who is exposed, and when.

3.2.6 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. Some in the industry refer to this process (and the specific factors that are recommended) as the application of "Plausibility Analysis."

3.3 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 relating to methodology, measurement, data management, validity, treatment of chronic populations and claims.

In Chapter 7, we 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. First, we will introduce two concepts discussed in more detail in Chapter 7: the concepts of methodology and causality.

3.3.1 Causality

Causality is an important concept in both scientific and commercial studies of DM outcomes. Just because savings are associated with a program does not necessarily mean that the savings are a result of the program. Attributing causality to an intervention program is a difficult problem, and one that has not been much studied in the field of DM outcomes. The scientific community demands a demonstration that a particular outcome has resulted from a particular cause. This requires proof that the DM intervention "caused" the specific outcome. Research to date has been focused on attempts to obtain an accurate estimate of savings, no matter the source. Because of the difficulties inherent in proving causality, commercial purchasers of DM

programs are usually satisfied with a weaker standard of proof: "demonstration" of savings, rather than proof of causality. Appendix 3.1 contains a more detailed discussion of causality for readers who are interested in studying the issue further. Business users, while demanding considerable rigor in other aspects of an evaluation, (such as validation and reconciliation of source data) may be satisfied with a "demonstration" standard, where "association" between cause and effect may be sufficient, rather than the stricter test of "causation."

3.3.2 What is a "Methodology?"

The dictionary definition of a methodology is: "a body of methods, rules, and postulates employed by a discipline**:** a particular procedure or set of procedures[3]". Methods on the other hand, are the tools, techniques and procedures that bring that vision to life. The calculation techniques or methods used by actuaries all contribute to the overall methodology. For example, a technique (such as an adjustment for age, or for trend) does not stand on its own but is rather an input to a methodology. In Chapter 7, we compare the characteristics, including validity, of 10 different methodologies for calculating savings results.

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. Such 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.

3.3.3 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, number of

[3] Merriam-Webster Unabridged Dictionary On-line.

work- or school-days lost to illness-related absence, 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 via alternative measures such as 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. 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 pe-

riod 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 or which members to track for evaluation purposes, will be covered in more detail later.

3.3.4 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 3.1 illustrates the phenomenon of regression to the mean at the level of the individual member:

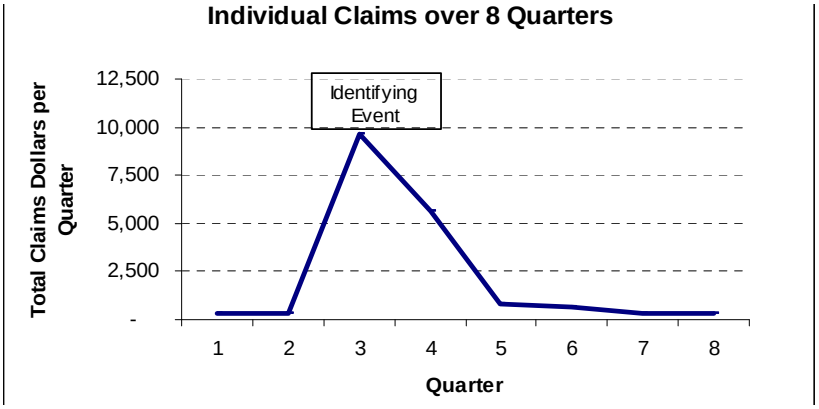

FIGURE 3.1

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 3.2. 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 in-

cluded in the "Pre" experience; the experience after enrollment is included in the "Post" experience.

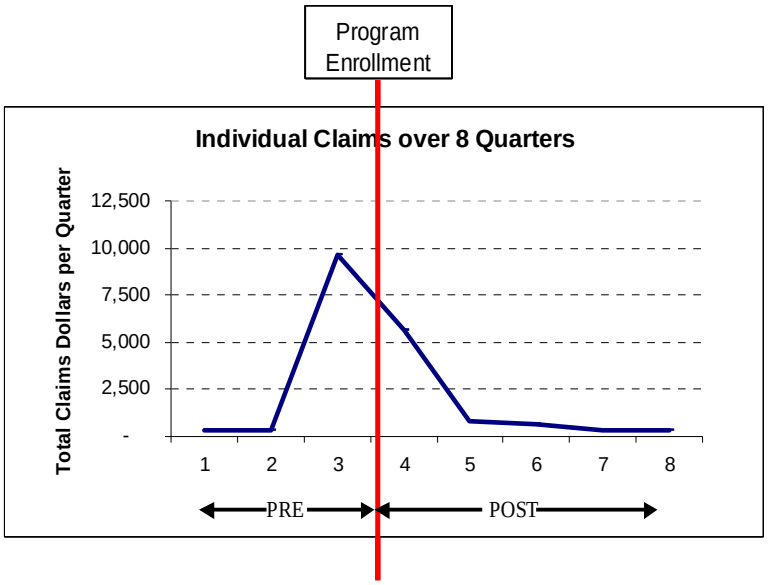

FIGURE 3.2

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. 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.

Table 3.1 demonstrates the more general impact that regression to the mean (claims increasing as well as decreasing) may have on an analysis. Note that Table 3.1 differs from Figure 3.1 and Figure 3.2, which show an individual's claim cost pattern over time, because Table 3.1 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 Table 3.1, 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.

TABLE 3.1

$'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	$5,368	$46,836	$831
		90%	10%	0%	
		90%	64%	40%	
Moderate	$5,658	$668	$6,599	$47,811	$5,398
$2 - $25 12%		55%	40%	5%	
		10%	34%	40%	
High	$49,032	$ 847	$9,609	$58,489	$21,017
$25+ 1%		26%	46%	28%	
		0%	2%	20%	
Total	$1,230	$ 355	$5,851	$49,377	$1,581

Source: Solucia Inc. 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 Table 3.1 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, transparent, and 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. At the same time, it should be remembered that clinician identification is not perfect either: conditions are not simple "binary" events (disease/non-disease). Rather, there are degrees of clinical disease and at times "fuzzy margins" at which it is as yet impossible to establish objectively whether an individual "has" the disease. In addition, in the absence of a unified Electronic Medical Record, not all of a patient's interactions with the medical system will be available in partial records, making identification less certain.

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, Fitzner et al., [58] review 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, in a voluntary program, those members who elect to enroll are a different risk-profile to those members who do not. Figure 3.3, taken from our unpublished data, tracks outcomes over time of different subpopulations 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 per-

cent). 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 3.3 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.

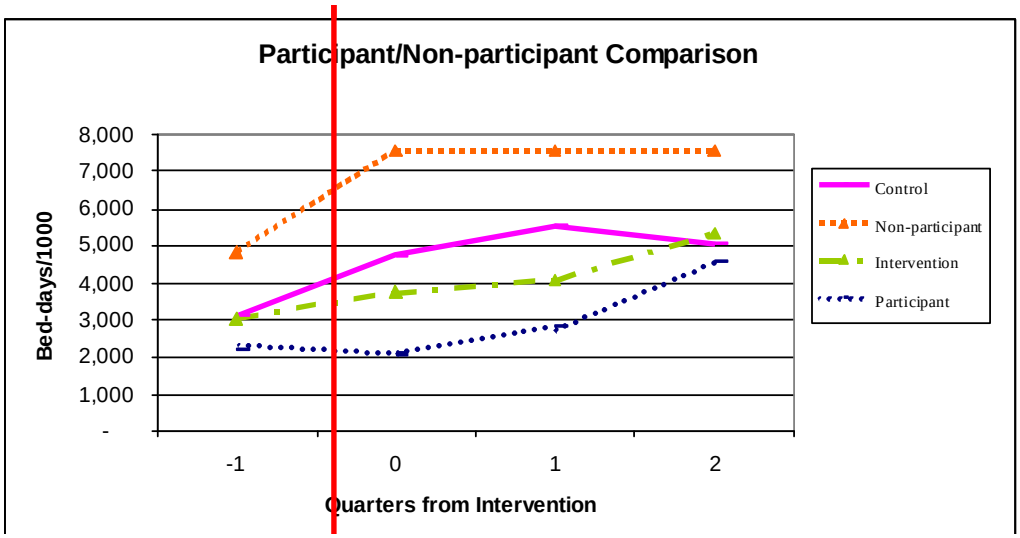

FIGURE 3.3

Two different member-outcome states are reported in Figure 3.3: 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. Enrollment bias will be present in a program for a number of reasons:

- The sickest patients may be more difficult to contact if they are hospitalized or undergoing some other form of treatment;
- Moderately sick Commercial patients may be *easier* to reach because they are not actively at work;
- Patients who are less severely ill may be more likely to already be in control of their care and therefore will welcome working with a nurse; and
- Patients who are more severely ill are more likely to suffer depression which, in turn, inhibits their ability to self-manage and likelihood to enroll in a program.

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.

3.3.5 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 mature (see below), the amount of time needed 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 calculate 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 results in 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 when 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 depends 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 create this problem because they do not contribute claims costs.