

WORKING P A P E R

Measuring and Reporting the Performance of Disease Management Programs

SOEREN MATTKE, GIACOMO BERGAMO,
ARUNA BALAKRISHNAN, STEVEN MARTINO,
NICHOLAS VAKKUR

WR-400

August 2006

Prepared for CorSolutions, Inc.

This product is part of the RAND Health working paper series. RAND working papers are intended to share researchers' latest findings and to solicit informal peer review. They have been approved for circulation by RAND Health but have not been formally edited or peer reviewed. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. **RAND**® is a registered trademark.

CONTENTS

FIGURES.....	V
TABLES.....	VII
SUMMARY.....	IX
ACKNOWLEDGMENTS.....	XI
1. INTRODUCTION.....	1
2. THE CONTEXT AND RATIONALE FOR A DISEASE MANAGEMENT REPORT CARD.....	2
2.1 PERFORMANCE PROBLEMS OF THE HEALTH CARE SYSTEM.....	2
2.2 DISEASE MANAGEMENT AS A POTENTIAL APPROACH TO CONTAIN COST AND IMPROVE QUALITY	3
2.3 THE DEMAND FOR TOOLS TO ASSESS THE VALUE OF DISEASE MANAGEMENT PRODUCTS	4
2.4 PURPOSE OF THIS REPORT.....	6
3. DESIGN APPROACH TO THE DEVELOPMENT OF A DISEASE MANAGEMENT PERFORMANCE REPORTING SYSTEM	7
3.1 OVERALL APPROACH.....	7
4. A CONCEPTUAL FRAMEWORK FOR DISEASE MANAGEMENT.....	11
4.1 INTRODUCTION.....	11
4.2 TOWARDS A CONCEPTUAL FRAMEWORK FOR DISEASE MANAGEMENT.....	11
4.3 THE CHALLENGE OF ATTRIBUTION.....	14
5. IMPLEMENTATION OF THE FRAMEWORK.....	21
5.1 INFORMATION, EDUCATION AND EMOTIONAL SUPPORT.....	21
5.2 SELF-EFFICACY.....	22
5.3 KNOWLEDGE.....	25
5.4 HEALTH RELATED BEHAVIOR.....	26
5.5 PROCESSES OF CLINICAL CARE.....	30
5.6 MORBIDITY REDUCTION.....	35
5.7 REDUCTION OF UTILIZATION AND DIRECT MEDICAL COST.....	40
5.8 IMPACT OF HEALTH ON PRODUCTIVITY.....	47
1. MEASURING ABSENTEEISM	49
2. MEASURING PRESENTEEISM	49
2.1. PERCEIVED IMPAIRMENT	50
2.2. COMPARATIVE PRODUCTIVITY/PERFORMANCE/EFFICIENCY	51
2.3. ESTIMATES OF UNPRODUCTIVE TIME	52
ASSESSMENT	52
3. COST ESTIMATION METHODS	53
SALARY CONVERSION METHODS	53
INTROSPECTIVE METHODS	55
FIRM-LEVEL METHODS	56
ASSESSMENT	57
5.9 HEALTH-RELATED QUALITY OF LIFE.....	57
5.10 PATIENT AND PROVIDER SATISFACTION.....	58
6. ISSUES IN CREATING A REPORTING FORMAT FOR THE MEASUREMENT SYSTEM...	61
6.1 MEASURE CALCULATION.....	62
6.2 MEASURE INTERPRETATION.....	62

APPENDIX.....	63
A. MEASURING SELF-EFFICACY.....	63
B. CLINICAL MEASURES WITH GUIDELINES.....	75
C. ESTIMATION PROCEDURE FOR RISK ADJUSTMENT.....	88
D. SURVEY INSTRUMENTS FOR HEALTH-RELATED PRODUCTIVITY.....	90
BIBLIOGRAPHY.....	97

FIGURES

Figure 4.1 - A conceptual framework of disease management..... 13

Figure 4.2 - Methods of Attributing Observed Effects to an
Intervention 16

Figure 4.3 - The difference-in-differences method..... 19

TABLES

Table 5.1. Clinical Process Indicators.....	32
Table B.1 Coronary Artery Disease.....	75
Table B.2 Congestive Heart Failure.....	77
Table B.3 Hypertension.....	80
Table B.4 Chronic Obstructive Pulmonary Disease.....	81
Table B.5 Asthma.....	83
Table B.6 Diabetes mellitus.....	86
Table D.1 Summary of Worker Productivity Measurement Instruments*.....	90
Table D.2 Detailed Properties of Worker Productivity Instruments.....	94
Table D.3 Content of Worker Productivity Instruments.....	95

SUMMARY

After a brief respite, health care costs in the United States are rising dramatically again, and are expected to continue growing faster than the overall economy. In the current year spending on medical care will amount to approximately \$1.77 trillion, an amount that represents 15.3% of GDP, a higher share than in any other country, and nearly double the share in 1980. By 2010 spending is expected to grow by 50% to \$2.64 trillion (CMS 2002). This trend has left policymakers and employers, as purchasers of care, actively searching for ways to contain medical cost.

Despite these high costs, the quality of care delivered by the health care system remains low. For example, a recent RAND study found that care for broad range of medical conditions was provided in accordance with current standards only about half of the time (McGlynn et al. 2003). Vulnerable patients with chronic conditions are especially affected by such problems and also consume the lion's share of the health care resources, rendering their care an obvious target for improving effectiveness and efficiency of care.

One promising approach to bridge the gaps in the health care system and improve management of patients with chronic conditions is a class of interventions known generally as disease management. Although the concept of disease management offers great promise, it has not yet been empirically demonstrated that these programs are able to reduce cost and improve care. There is also no industry-wide reporting standard for disease management vendors to measure and report their performance to clients and potential clients in a scientifically sound and comparable fashion. An initial step towards such a standard was recently made by American Healthways in collaboration with Johns Hopkins University.

In this report, we built upon this and other existing measurement systems to develop a comprehensive and scientifically sound Disease

Management Report Card. We start by introducing a conceptual framework for disease management and then assess which of the components of the framework are relevant enough for disease management clients to be covered in a report card. Next, we review existing measures for the selected components and determine which are valid and reliable enough to be used for disease management performance reporting. We address feasibility and other operational issues in the implementation of the proposed measures. The report also addresses two important analytical issues in constructing measures for a report card, which are methods to attribute observed changes to the intervention and sampling strategies to reduce potential bias.

This report presents the first comprehensive methodology for measuring and reporting the performance of disease management programs that is built on a conceptual framework covering the relevant components of the intervention, uses statistical techniques to attribute changes in the health, utilization, and behavior of the target population to the intervention and a sampling method that avoids common sources of bias. The methodology would allow a disease management vendor to fairly and credibly demonstrate its performance to current and future clients. Widespread adoption of standardized reporting would facilitate client choice and potentially stimulate ongoing improvements in the delivery of services under these programs.

ACKNOWLEDGMENTS

The authors would like to thank Elizabeth McGlynn, Elizabeth Malcolm, Susan Straus, Joy Moini, Cheryl Damberg and Michael Seid for their contributions to this report and Michelle Bruno for her help in preparing it.

1. INTRODUCTION

The purpose of this report is to describe the methods and rationale used to develop a standardized performance measurement and reporting system (report card) for disease management programs. This report card reflects a need for the industry to improve the ability to demonstrate value to a broad range of current and future clients and was developed jointly by RAND and CorSolutions. The conceptual framework for disease management that has been developed under this project and the measurement system that operationalizes this framework are designed to be applicable to any disease management vendor.

This report is composed of several sections. We first describe the context and rationale for a disease management report card, including its primary audience and intended use. Next, we describe the analytic approach to creating the report card. The fourth section sets out a conceptual framework to capture the essential components of disease management, which is operationalized into a measurement system in the fifth section. The sixth and last section sketches issues related to the communication of the selected measures, such as formation of composite measures and aggregation rules.

2. THE CONTEXT AND RATIONALE FOR A DISEASE MANAGEMENT REPORT CARD

2.1 PERFORMANCE PROBLEMS OF THE HEALTH CARE SYSTEM

Public and private purchasers of health care, and health plans as their agents are increasingly worried about rapidly rising health care costs and their impact on public finances and business competitiveness and are searching for ways to contain this cost growth. After a brief respite during the 1990s, double-digit growth rates in cost returned recently: During 2002, the average increase in health premiums paid by the nation's largest employers rose 13.7%, an amount far exceeding the 2.5% annual rate of inflation. In 2003, premiums increased yet another 13.9% (NSBA, 2003). In 2004, the United States will spend approximately \$ 1.8 trillion on health care and spending is expected to reach \$ 2.6 trillion by 2010 (CMS, 2002).

In spite of high expenditures, the quality of care that patients receive is still found wanting so that purchasers question the value that they receive. In a recent study published in the *New England Journal of Medicine* McGlynn and co-authors found medical care was provided in accordance with current medical standards only 55% of the time, with no significant variation between preventive care, care for acute illness, or for a chronic condition (McGlynn et al., 2003). Numerous other studies have documented similar results (e.g. Institute of Medicine, 2003; Clark et al., 2000; Legorreta et al., 2000; McBride et al., 1998).

Patients with chronic conditions are particularly vulnerable to such quality of care problems. It is estimated that more than 100 million Americans currently suffer from a chronic condition such as heart disease, diabetes or asthma, while at least 40 million of those have two or more chronic conditions. Chronic diseases currently account for an estimated 70% of all U.S. health care expenses. However, surveys and research consistently document that the majority of chronically ill patients are not receiving effective therapy, possess inadequate disease

control, and are not satisfied with the quality of treatment received (Wagner *et al*, 2001b; Ornstein, 1999).

Several studies indicate that poor management is experienced by over 50% of patients with diabetes, hypertension, tobacco addiction, hyperlipidemia, congestive heart failure, asthma, depression and chronic atrial fibrillation (IOM, 2003a; Clark *et al*, 2000; Legorreta *et al*, 2000; McBride *et al*, 1998; Ni *et al*, 1998; Perez-Stable & Fuentes-Afflick, 1998; Samsa *et al*, 2000; Young *et al*, 2001). Additionally, a total of 18,000 Americans are estimated to die each year from heart attacks because they were not treated according to medical guidelines (Chassin, 1997; IOM, 2003b).

A particular concern of large employers is the effect of insufficient management of chronic conditions on the productivity of their workforce. As an illustration, the National Committee for Quality Assurance (NCQA) estimated that absenteeism alone would be cut by 21.8 million days annually if all Americans suffering from asthma, depression, diabetes, heart disease, and hypertension were treated by top performing health plans, i.e. those with quality scores at the 90th percentile of their HEDIS indicators, the equivalent of adding the output of roughly 104,520 workers full-time for a year (NCQA, 2002). Beyond its effect on lost time, ill health negatively impacts productivity. A recent study showed, for example, that 76.6% of the productivity loss caused by chronic pain was due to reduced performance at work rather than absence from work (Stewart *et al*, 2003).

2.2 DISEASE MANAGEMENT AS A POTENTIAL APPROACH TO CONTAIN COST AND IMPROVE QUALITY

Disease management, "a system of coordinated healthcare interventions and communications for populations with conditions in which patient self-care efforts are significant" (<http://www.dmaa.org/definition.html>, September 20, 2004), is a promising approach to improve the effectiveness and efficiency of care and has become increasingly popular with private and, more recently, public purchasers.

Disease management started with pharmaceutical companies, who used their drug dispensary databases to identify patients with chronic conditions and offered educational services to them (Bodenheimer, 1999; Burns, 1996). The underlying assumption was that improved patient education would lead to greater involvement of patients in their care, better compliance with medication regimens and consequently more efficient care. Later, health plans started to implement in-house disease management programs, independent vendors appeared who offered disease management services to health plans, and even some provider organizations introduced smaller-scale programs (Bodenheimer, 2000).

Today, there are upwards of 200 commercial disease management firms and numerous internal health plan and provider-operated programs, which typically target chronic conditions such as diabetes, asthma, and congestive heart failure. While the industry remains very heterogeneous in terms of ownership, size and scope, some common features have emerged: According to the Disease Management Association of America, a "Full Service Disease Management Program" must include the following six components (www.dmaa.org/definition.html):

1. Population identification processes
2. Evidence-based practice guidelines
3. Collaborative practice models to include physician and support service providers
4. Patient self-management education (may include primary prevention, behavior modification programs and compliance/surveillance)
5. Process and outcomes measurement, evaluation, and management
6. Routine reporting/feedback loop (may include communication with patient, physician, health plan, and ancillary providers, and practice profiling)

2.3 THE DEMAND FOR TOOLS TO ASSESS THE VALUE OF DISEASE MANAGEMENT PRODUCTS

A key concern of purchasers is whether disease management can in fact deliver on the promise to improve the effectiveness and efficiency of care. Some initial research shows encouraging results. For example, it has been shown that improving guideline adherence in chronic

illnesses requires substantial patient participation and a growing body of literature demonstrates a positive effect resulting from systematic efforts to increase patients' knowledge, skills, and confidence in managing their conditions (Callahan, 2001; Norris et al, 2001). Recent meta-analyses involving more than 41 intervention studies demonstrates that by improving individual treatment processes, the quality of treatment for chronic illnesses such as depression and diabetes can be substantially increased (Renders et al, 2001; Callahan, 2001; Von Korff et al, 2001). Interestingly, this study found that only those interventions possessing a strong patient-oriented component resulted in noticeable improvements to patient outcomes. As another example, Group Health Cooperative reported that an in-house disease management program its 18,000 diabetics lead to improved treatment while costs decreased (McCulloch et al, 1998 & 2000). Limited evidence from other research projects also supports a positive effect of disease management interventions on quality and costs (McAlister et al, 2001; Norris et al., 2002; Wagner, 1998a).

But there is so far insufficient evidence from research studies to conclusively determine that disease management is a generally effective concept, the conditions under which it is effective, or whether some programs and vendors are superior to others. A particular problem is that most of the available evidence has been derived from interventions in the trial phase that were designed and carried out in an academic setting, thus limiting the generalizability of findings. So far, only three studies addressed the effects of large, population-based disease management programs, two of programs operated by integrated delivery systems (Fireman et al., Sidorov et al.) and one by a third-party vendor (Villagra and Ahmed). This paucity of evidence combined with the limitations of some methods used to estimate the effects of interventions, have lead to skepticism about the industry among actual and potential clients and to demands for greater accountability and transparency (Linden 2003). In addition, some general skepticism remains in the provider community about commercial vendors of disease management (Bodenheimer, 1999).

To counter this credibility gap, disease management vendors in collaboration with researchers have searched for more rigorous evaluation methods that could be used to report their performance back to clients. Most notably, American Healthways, a commercial disease management vendor, in collaboration with Johns Hopkins University has proposed a unified reporting framework. However, an industry-wide reporting standard that would allow clients to compare the performance of different programs and vendors and to identify a program that fits their particular needs best has not yet emerged.

2.4 PURPOSE OF THIS REPORT

The purpose of this report is to advance the debate about performance measurement and reporting for disease management programs by proposing a RAND/CorSolutions methodology that will allow a comprehensive evaluation of the impact of disease management interventions. The proposed methodology is designed to overcome some of the limitations of existing evaluation methods with respect to comprehensiveness of the measures set, sample selection and attribution strategy, enabling disease management vendors to demonstrate the value of their programs based on a comprehensive set of scientifically sound measures. The current report will present a list of measures and the rationale for selecting them.

The primary audience for this report card is current and future clients of disease management vendors, i.e. private and public purchasers of health care and health plans. Other potential users could be intermediaries of the clients, such as benefit consulting firms, policymakers and researchers.

3. DESIGN APPROACH TO THE DEVELOPMENT OF A DISEASE MANAGEMENT PERFORMANCE REPORTING SYSTEM

3.1 OVERALL APPROACH

The final goal of this project is to arrive at a reporting system that is based on a comprehensive set of measures but organizes all selected measures into a few dimensions to facilitate communication. Even the brightest human being can only hold a few pieces of information in short-term memory when making a decision. Cognitive psychologists suggest that about five to seven bits of data can be utilized when making a decision. Further, hierarchical structures that organize specific details within a general framework facilitate the use of information in three ways. First, hierarchies facilitate comprehension. Second, hierarchies help people memorize information and retrieve that information for later use. Third, hierarchies communicate importance. The framework used for the performance metrics, thus, should have few categories and should organize information in a way that is useful for decisionmakers.

There are two different strategies for creating frameworks. The first approach, which might be called "bottom-up," starts with the individual measures that are available and creates summary categories that maximize the number of measures used. This can either be done quantitatively, using factor analysis or other methods designed to identify patterns in data, or it can be done qualitatively by obtaining expert opinion. The second approach, which might be called "top-down," starts with conceptualizing the construct that the measures ought to cover and then identifies measures that capture those components.

The bottom-up approach is more frequently associated with research or decision analysis. This approach has the advantage of trying to use all available information. Since the approach is empirically driven, another advantage is the opportunity to identify patterns in data that might otherwise have escaped notice. The disadvantage of this approach, particularly if done quantitatively (e.g., using factor analysis), is

that it may produce results that are difficult to interpret and may not be valued or easily understood by the intended audience.

The top-down approach is more structured because it starts with intuitively plausible categories that reflect our understanding of the essential components of disease management intervention. The disadvantage of this approach is that there may be categories for which no or few measures currently exist. For this project, we opted for a top-down approach, primarily because disease management tries to affect a wide variety of clearly distinct categories, reaching from clinical processes over direct medical cost to employee productivity.

The first step was therefore to develop a conceptual framework of disease management that reflects our understanding of the essential elements of this intervention: the desired outcomes and the pathways designed to achieve those outcomes. The categories of this framework define the universe of measurement in its broadest sense, i.e. measures outside of those categories will not be considered. However, there will almost certainly be categories that will not be included in a performance report or for which no suitable measures can be found.

We will therefore assess for each category whether measures should be included in the reporting system by applying the following three criteria:

1. Is the category of sufficient interest to users of the reporting system to justify the cost and complexity of including measures for it? This criterion is satisfied if the category captures either an end result of the disease management intervention of obvious relevance, e.g. direct medical cost, or an intermediate result that is known to predict such end results. For example, clients may not want information on every step in a clinical pathway but would want to know how disease management changes processes that are proven to impact health outcomes.
2. For a category deemed relevant, are there established and scientifically sound measures available to capture performance?

3. Do data sources available to a typical disease management vendor or operator allow a sufficient number of measures to be constructed for the respective category?

For categories that are deemed relevant, i.e. those that pass the first criterion, but cannot be implemented, i.e. that don't pass one of the other two criteria, we recommend further development work.

Since this selection process may still result in an unreasonably large number of categories or measures within a category, a final assessment step will solicit input from the end users of the reporting system to assure that the final set of measures is comprehensive, parsimonious and balanced.

4. A CONCEPTUAL FRAMEWORK FOR DISEASE MANAGEMENT

4.1 INTRODUCTION

As we have discussed above, disease management presents a theoretically appealing model to improve the effectiveness and efficiency of medical care. Furthermore, the business model of today's disease management vendors appears to be economically viable.

It is not, however, universally accepted that disease management can significantly improve the functioning of the health care system. This skepticism results from limited empirical data and inadequate analytic methods. Thus, a comprehensive and scientifically sound standard methodology to measure and report the performance of disease management programs is vital for the credibility and growth prospects of the industry.

Initial steps in this direction have been made by American Healthways, a large disease management vendor, in collaboration with Johns Hopkins University. Their efforts produced a guideline and explicit methods for measuring medical cost reduction and quality improvement (American Healthways 2002). While their methodology represents an enormous improvement over what was previously used in the industry, it has two major limitations. First, their approach to estimating the cost savings attributable to disease management may overestimate the true effect, as we will demonstrate in detail below. Second, their set of clinical and utilization measures encompasses well-established indicators, but fails to provide a full account of how a given disease management program performs.

The RAND/CorSolutions project sets out a comprehensive and scientifically sound reporting framework that improves upon current reporting systems. As outlined above, we start by proposing a conceptual framework for disease management programs that will provide guidance for the selection of a comprehensive set of performance measures.

4.2 TOWARDS A CONCEPTUAL FRAMEWORK FOR DISEASE MANAGEMENT

The conceptual framework proposed here is in keeping with and builds upon the definition of disease management provided by the Disease

Management Association of America (DMAA) and the MacColl Institute for Healthcare Innovation's Chronic Care Model. Figure 4.1 displays the conceptual framework. Disease management is pictured as a patient-centered approach consisting of three building blocks:

1. The provision of actionable data to patients and their providers
2. Education, primarily of patients but to some degree of their providers about their disease and its treatment based on the latest medical evidence
3. The provision of social and emotional support to patients to enable them to act on the newly gained information.

Routine provision of these interventions makes patients more knowledgeable about their disease and their general health status and empowers them to act upon the newly gained knowledge. This affects both their own health-related behaviors, such as compliance with treatment regimes, diet and exercise, and their interactions with their providers, i.e. patients become a more active partner in health care decisions rather than passive recipients of care. This new role for patients, in combination with provider-directed data and education, influences the way patients receive health services as well as the specific services received. For example, treatment decisions may be better aligned with patient preferences and providers are reminded to conduct regular tests to monitor the patients' conditions.

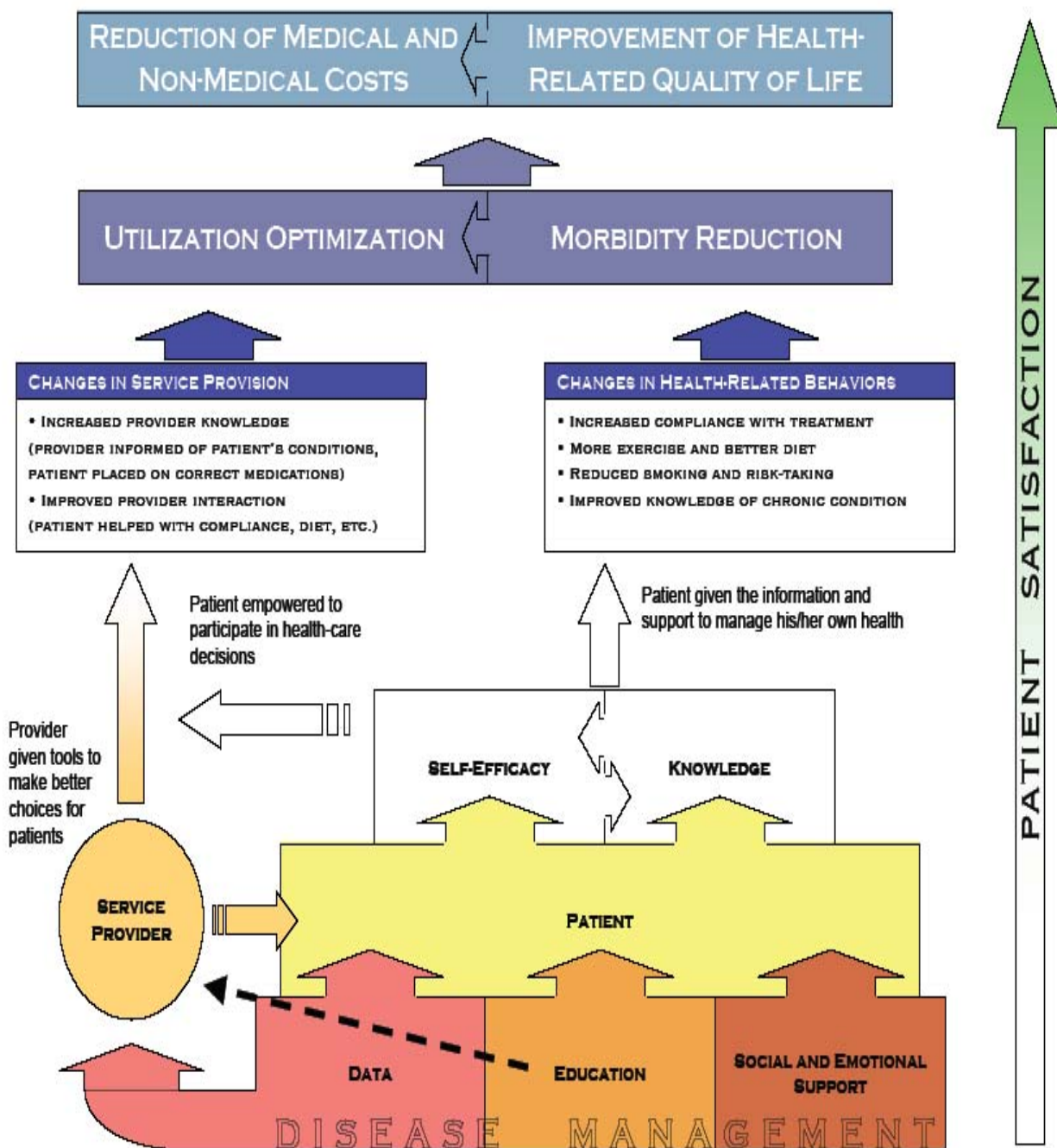


Figure 4.1 - A conceptual framework of disease management.

The combination of improved care and behavioral changes would lead to improved utilization of health care resources, where improved utilization does not necessarily mean the provision of fewer services but the provision of the most effective care in the most appropriate setting. This may occur through a reallocation of services from crisis response (ER, hospitalization) to proactive management (tests, outpatient visits, preventive care). It could also reduce the burden of disease by slowing the progression of chronic conditions and possibly even by reducing their prevalence. More efficient utilization and reduced morbidity would lead to both better health related quality of life for patients and a reduction in cost. The two components of cost reduction would be the reduction of direct medical cost and, for employers, a reduction in the non-medical cost of illness, such as loss of productivity.

4.3 THE CHALLENGE OF ATTRIBUTION

A key challenge for performance reporting of disease management programs is to attribute observed changes in measures to the intervention. In a research context, this is usually done by contrasting the results in the intervention group to those in a control group, but disease management programs that are implemented in an operational setting do not usually have a control group available that would allow assessing which portion of the results is attributable to the intervention and which portion is due to changes in such factors as inflation, demand for services, the mix and prevalence of diseases, and aging in the population. Figure 4.2 presents a broad categorization of attribution methods with increasing scientific rigor, but decreasing practicability outside of a research setting.

The choice of an attribution strategy will largely be driven by availability of data. Obviously, experimental and quasi-experimental approaches are the most rigorous way of attributing observed changes to the intervention, if a control or comparison group exists. In its absence, we would recommend the use of benchmark data as rough estimates

for secular trends and statistical adjustments to the degree possible, but acknowledge that appropriate data will not always be available. It should be emphasized that the categories are not mutually exclusive, i.e. they can be combined into an overall method. For example, even a quasi-experiment approach would typically utilize adjustments for patient mix and/or inflation to estimate cost savings.

Intervention Data Only

In some cases, in particular if collection of the required data is tied to the intervention itself, attribution is not possible in the first observation period. In this case, one would report the data from the first observation period as baseline and use changes from the baseline for attribution.

Unadjusted Pre-Post Comparison

Under this method, pre-intervention data are required for the treatment group. The effect of the intervention is expressed as the difference between the performance along an indicator in the pre-intervention period and the intervention period(s). For example, direct medical costs incurred by the treatment group in the intervention year are simply subtracted from the costs in the baseline year to estimate savings. While the method is obviously easy to implement, making it the most common attribution strategy in disease management, it has severe limitations because it fails to account for factors such as changes in prices, secular trends in care practices and changes in the composition of the treatment group over time.

METHODS OF ATTRIBUTION

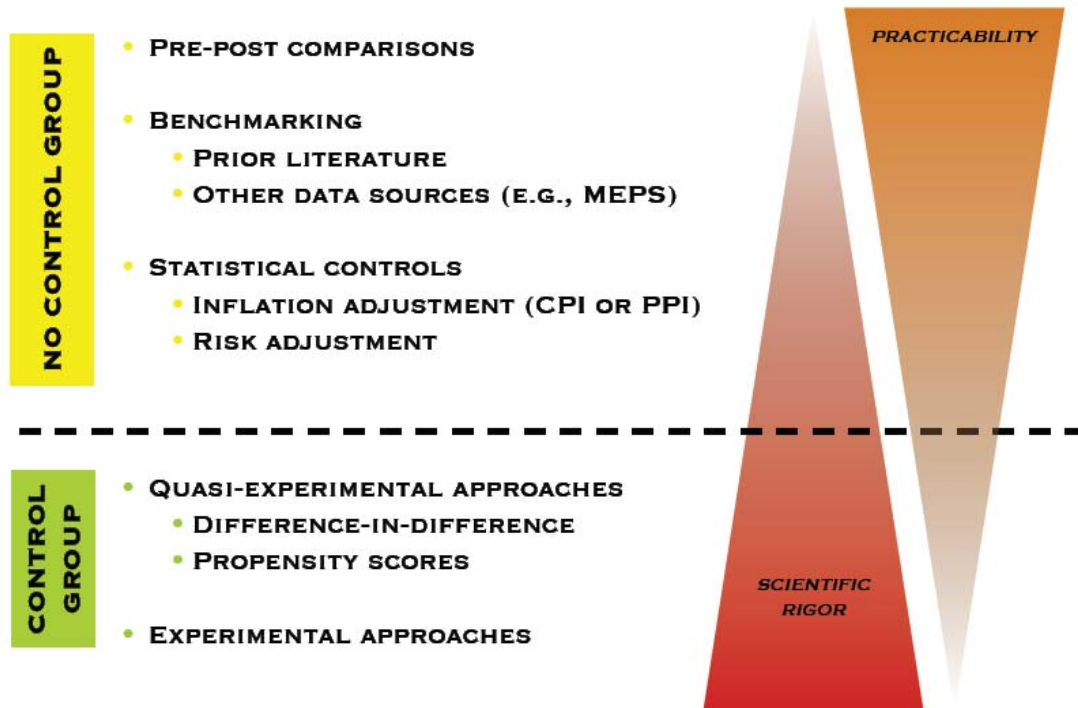


Figure 4.2 - Methods of Attributing Observed Effects to an Intervention

Benchmarking

Benchmarking uses estimates derived from other sources to put the results obtained through a pre-post comparison into perspective. For example, if the pre-post comparison suggested a five-percent cost decrease and a benchmark experienced a two-percent increase in the same period, the estimated effect of the intervention would be seven percent. Groups for benchmarking can be chosen based on such characteristics as medical condition, geographic area, age, and year. They are available through publicly available sources such as the Medical Expenditure Panel Survey (MEPS), research projects or commercial sources and can be based on surveys or claims data. Even rough estimates of historical performance in the intervention group could serve as a benchmark.

While falling short of a formal adjustment method, benchmarking allows in theory to account for various effects that may mask the true effect of the intervention, such as secular trends in spending and market- and disease-specific changes. It is able, however, to do so only if the benchmarking data have been derived from a population similar enough to the intervention group. This is usually not the case, as, for instance, research studies and the MEPS typically provide dated information and may not match the characteristics of the disease management population.

Statistical Adjustment

Statistical adjustment tries to account for the effect of observable and measurable characteristics to get a more precise and unbiased estimate of the treatment effect. Such adjustment can be as simple as adjusting direct medical costs for inflation to complex statistical modeling to correct for changes in patient casemix over time. An application of this attribution method is discussed in greater detail in Section 5.7 and Appendix C.

Quasi-Experimental Approaches

These designs require a control group that should be as similar as possible to the intervention group but do not require random assignment. Opportunities for such analyses sometimes arise through natural

experiments, such as staggered enrollment in a disease management program, or though the availability of administrative data for a population not exposed to disease management. If the natural experiment results in balanced treatment and comparison groups the analysis can proceed as if a randomized experiment was performed. If the treatment and comparison groups are not perfectly similar various analytical approaches are possible to adjust for the differences; we outline here the difference-in-differences approach to exemplify the general approach (Figure 4.3). Others are propensity scores or other case matching methods.

In the difference-in-differences approach, for example, direct medical costs are calculated for both the treatment and control groups in both the year prior to the DM intervention and the first year into the intervention. For both groups, the costs in the intervention year are subtracted from the costs in the baseline year, forming the first differences (A-B and C-D for the treatment and control groups, respectively, Figure 4.3). The difference within the control group (C-D) is then subtracted from the difference within the treatment group (A-B), forming the difference-in-differences estimate. As a measure of differential change the difference-in-differences estimate is able to account for secular trends in spending and utilization, an important contribution in light of the strong upward trend towards higher cost of medical care. Thus, it is able to attribute savings to the intervention, even if the simple pre-post comparison had suggested a cost increase.

Experimental Designs

The most unbiased estimate of the treatment effect is obtained through experimental designs in which study participants are randomly assigned to the intervention and control arms. Assuming adequate randomization, the simple pre-post comparison represents the treatment effect. However, this approach is rarely available outside of research

or demonstration projects, such as the Medicare Health Support Demonstration¹.

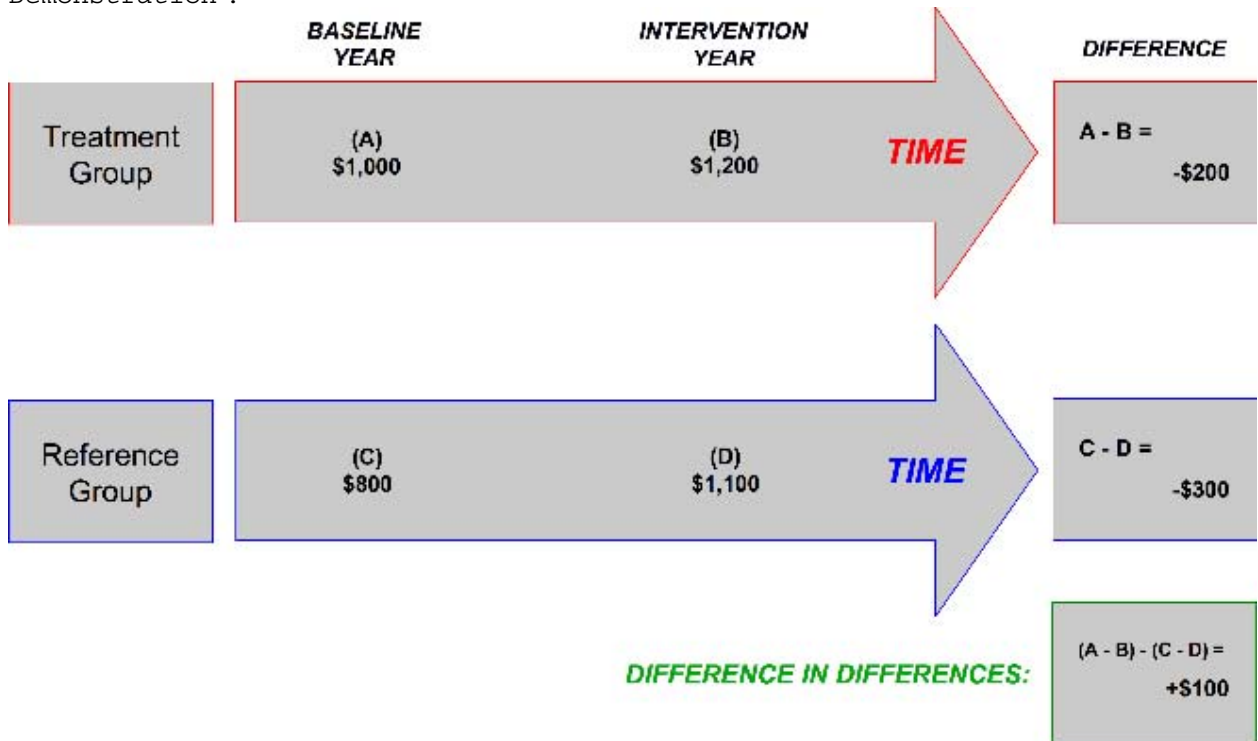


Figure 4.3 - The difference-in-differences method.

¹ Data from such randomized trials, however, can and should be used to estimate the bias that less rigorous evaluation methods introduce by comparing results derived from the randomized trial to hypothetical results that the intervention would have yielded, if there had been no control group.

5. IMPLEMENTATION OF THE FRAMEWORK

A comprehensive measurement system should capture the essential components of the conceptual framework within the bounds of parsimony to minimize the burden of collecting and reporting data. As mentioned above, we will therefore assess for each of the categories whether measures should be included in the reporting system by applying the following three criteria:

1. *Relevance*: Is the category a key element in decision-making? Are the results of sufficient interest to users of the reporting system to justify the added cost and complexity of including measures for it?
2. *Measures Availability*: For a category deemed relevant, are there established and scientifically sound measures available to capture performance along this category?
3. *Feasibility*: Do data sources available (or potentially available) to CorSolutions allow constructing a sufficient number of measures for the respective category with reasonable effort?

The evaluation criteria are being applied in a hierarchical manner, i.e. if a criterion is judged not to be met, the following ones are not being discussed. If we recommend measures for a category, we briefly comment on operational issues.

5.1 INFORMATION, EDUCATION AND EMOTIONAL SUPPORT

Relevance

For a disease management intervention to function well the core elements of the intervention, i.e. the provision of information, education and support to patients and also to providers, must be well designed and executed. Management of a disease management program is likely to find reports about the operation of their program of interest for internal quality improvement, but clients may be principally interested in whether the intervention achieves its intended results. And given the paucity of research, it is unknown how well particular

program components predict relevant end results. We therefore recommend that these categories not be included in an external performance reporting system.

5.2 SELF-EFFICACY

Relevance

Self-efficacy is a person's judgment of his or her ability to perform a behavior in a given situation (Bandura, 1977, 1997). Unless people believe that they can produce a desired effect by their own actions, they have little motivation to act or persevere in the face of obstacles. Self-efficacy beliefs influence the behaviors individuals choose, how much effort they invest in those behaviors, how long they persist at those behaviors in the face of barriers, and the level of accomplishment they realize (Bandura, 1977). Although self-efficacy perceptions are not the only determinant of behavior change - individuals must also have the appropriate knowledge, skills, and incentives (Bandura, 1977) - they are often found to be the best predictor of people's actions (e.g., see McKusick et al., 1986).

A growing body of literature clearly demonstrates that self-efficacy is a consistent, independent predictor of an individual's intentions and actions with regard to the initiation and maintenance of healthy behavior (see Bandura, 1997 and Schwarzer, 1992) for reviews). In the Appendix, we briefly review some of the literature that demonstrates the importance of self-efficacy in smoking cessation, weight management, physical activity, and adherence to treatment regimens, four behaviors that are targeted by disease management interventions, and discuss the measurement of self-efficacy with regard to these behaviors. Before that, however, we make some general points about the measurement of self-efficacy.

Self-efficacy is not only an important intermediary outcome; it is an important outcome in and of itself. In organizational training, self-efficacy is considered an important outcome measure of learning (Kraiger, Ford, & Salas, 1993). Likewise, in health behavior change, increases in self-efficacy are a direct indicator of treatment success. If treatment is successful, patients should become increasingly

confident of their ability to change maladaptive patterns of behavior and to implement and maintain new healthful ones. Moreover, measurement of this construct is consistent with the theoretical foundations on which many disease management interventions are built. One of the main theoretical bases of CorSolutions' interventions, for example, is social cognitive theory, in which the notion of self-efficacy plays a pivotal role. Measures of self-efficacy, in addition to traditional health outcomes, provide a comprehensive assessment of treatment effects.

Self-efficacy is not a global perception that functions across behaviors and situations. Self-efficacy judgments are specific to the behaviors that must be enacted in the situations in which they occur (Bandura, 1977, 1986, Hofstetter et al.1990, Murphy et al. 1995, O'Leary 1985). Accordingly, Bandura has argued that self-efficacy measurement should be specific both to the situations in which the behavior will take place and the level of challenge in that situation (Bandura, 1991).

Single-item measures of self-efficacy that do not require individuals to consider the situations under which behavior is to occur are inconsistent with Bandura's (Bandura, 1977, 1997) conceptualization of self-efficacy. Such measures are also ineffective predictors of behavioral change (e.g., see Forsyth & Carey, 1998). To achieve high predictive validity, it is necessary to measure self-efficacy with multiple items representing a variety of contexts in which the behavior is to occur and different gradations of difficulty. For example, self-efficacy to resist overeating should not be reduced to a single judgment about one's ability to resist overeating. One's ability to resist overeating may depend on whether one is alone or in the company of others, how one is feeling (e.g., anxiety and depression are known to promote overeating), and the time of day under consideration (e.g., controlling one's eating in the evening is often harder than during the day). Details such as these must be represented in measures of self-efficacy to achieve reliable prediction of behavior.

Measures Availability

Self-efficacy measures consistent with Bandura's conceptualization have been developed in each of the areas of interest to disease

management interventions. We propose to include measures for the following four areas in a disease management report card. These measures are described in detail in Appendix A.

Smoking Cessation

Weight Management

Exercise and Physical Activity

Treatment Adherence

Feasibility

All suggested measures would require dedicated data collection and a modification of the currently used data collection protocols for obtaining data on self-efficacy. Self-efficacy scales would be included in the telephonic intervention for all high acuity patients. We anticipate that administration of the self-efficacy scales will take approximately 2-3 minutes per scale.

Operational Issues

Self-efficacy should be measured in all areas at the time of patient enrollment patients or shortly thereafter. Beyond this baseline measurement, self-efficacy should be measured every two months, using a staggered measurement schedule so that a patient completes only one or two self-efficacy scales per assessment. This staggered measurement schedule reduces the time needed to measure self-efficacy during any one phone call, and assures that a few measurements are obtained for each patient on each self-efficacy scale during the average length of enrollment in the high acuity program (i.e., 6-9 months). Questions regarding self-efficacy for treatment adherence would be asked of all patients. Questions regarding smoking cessation, healthy eating, and physical activity would be asked only for those patients who are receiving interventions for these behaviors (e.g., if a patient is not actively engaged in weight management, self-efficacy for weight management need not be measured).

For consistency, all of the self-efficacy instruments included in Appendix A can be completed using the same response scale. We recommend

the 5-point scale ranging from "not at all confident" to "extremely confident" used in the smoking cessation self-efficacy scale by Velicer and colleagues (Appendix A).

We recognize that the instruments we have suggested may be too lengthy for practical purposes. In some cases, shorter instruments have been tested or recommended (i.e., smoking abstinence and exercise self-efficacy) but in other cases (weight management and treatment self-efficacy) they have not. Even the 9-item smoking abstinence instrument might be somewhat too long. For these cases, we recommend that research to validate shorter instruments. This can be done by administering the full instruments to a large sample of people and using a method such as principal components analysis to determine what items best comprise a valid, shorter instrument (e.g., see Hodgins, Maticka-Tyndale, El-Gueybaly, & West, 1993). Once a smaller set of items has been selected, the scale should be validated by assessing its association with relevant health outcomes (e.g., reduction in smoking; increase in physical activity).

5.3 KNOWLEDGE

Relevance

Improving participants' knowledge about their health is an important avenue through which disease management programs try to achieve behavior change. In that, it would seem appropriate to determine whether and to what degree knowledge improves as a result of disease management interventions. However, we do not recommend including such measures in a report card for two reasons. First, knowledge is regarded as a distal factor predicting health behavior change (in comparison, for example, to self-efficacy) (Bandura, 1977). Therefore, knowledge fails to meet our criterion of being established as a direct or proximal predictor of desirable end results. Second, the cost of identifying or developing knowledge measures for all of the conditions that disease management addresses would be enormous. Moreover, measuring whether the interventions enhance patient knowledge is best performed using equivalent forms rather than administering the same test repeatedly. Administering knowledge tests also increases the cost of the

intervention. Third, administering a knowledge test is evocative of the Socratic method, which is inconsistent with the participative approach that is typically used in disease management interventions. We recommend that nurses ask questions regarding knowledge of how to manage their conditions in routine exchanges with patients, but we do not recommend incorporating formal knowledge tests into a report card.

5.4 HEALTH RELATED BEHAVIOR

Relevance

Changing health-related behaviors is a central goal for a patient-centered approach like disease management. Thus, measuring those changes conveys important information to disease management clients about how well the intervention affects patients' risk profiles. In addition, many health-related behaviors have a proven link to better health outcomes.

Measures Availability

Disease management programs commonly report measures that reflect health-related behaviors. We propose the following 10 measures that are all commonly used in this field.

Some important gaps in measurement remain. For example, an important goal for disease management is to improve patients' diet both to bring it in line with specific requirements of their condition (e.g., salt intake restriction in CHF, Bread Unit schedule in diabetics) and to improve health in general (e.g., weight loss programs). Many well-established dietary schemes exist, such as the DASH-diet for hypertension, but standardized measures that track compliance with such diets are lacking.

For all participants:

- Smoking²

² Definitions adapted from the Canadian Tobacco Use Monitoring Survey (CTUMS) (<http://www.hc-sc.gc.ca/hecs-sesc/tobacco/research/ctums/term.html> accessed September 30, 2004)

- Rate of daily smokers
- Proportion of daily smokers who quit successfully for less than a year
- Proportion of daily smokers who quit successfully for more than a year
- Exercise³
 - Proportion of participants engaging in at least 30 min of at least moderate activity daily if medically possible⁴. Alternatively, this measure could be specified as the average number of days that participants engaged in 30 minutes or more of at least moderate activity?
 - Proportion of participants engaging in at least 60 min of at least moderate activity daily if medically possible⁵. Alternatively, this measure could be specified as the average number of days that participants engaged in 60 minutes or more of at least moderate activity?
- Obesity
 - Proportion of overweight participants ($25 < \text{BMI} < 30$)⁶
 - Proportion of obese participants ($\text{BMI} \geq 30$)⁷
 - Proportion of overweight and obese participants who lose at least 10% of their body weight over a year⁸.

³ The targets have been defined for the general population. Exercise targets for participants with advanced disease would have to be operationalized differently.

⁴ Target recommended for general health by the IOM (Food and Nutrition Board, Institutes of Medicine, Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids (Macronutrients), Washington, D.C., 2002)

⁵ Target recommended for weight control by the IOM (Food and Nutrition Board, Institutes of Medicine, Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids (Macronutrients), Washington, D.C., 2002)

⁶ As defined by the CDC (<http://www.cdc.gov/nccdphp/dnpa/obesity/defining.htm> accessed September 30, 2004)

⁷ As defined by the CDC (<http://www.cdc.gov/nccdphp/dnpa/obesity/defining.htm> accessed September 30, 2004)

Two additional health-related behaviors that would be interesting to include in a report card are compliance with dietary restrictions and medication regimens, but it appears difficult to measure those with the reliability and validity required for a reporting system. Objective measurement approaches for medication compliance have been introduced for drug trials, such as unannounced pill counts or electronic devices that record when a drug container is being opened, but don't seem suitable for the disease management context. Medication adherence is thus commonly measured by asking patients whether they take their drugs, which is prone to error, e.g. because patients may genuinely believe that they comply but misunderstood the instructions or because patients may give socially desirable answers. The latter error might be particularly problematic in the disease management setting, as participants might become more likely to affirm their compliance, if their disease management nurse asks them repeatedly about it. More complex survey tools have been developed, such as the Medication Adherence Self-Report Inventory, but not yet widely tested (Walsh et al. 2002). Thus, there is so far no widely accepted, valid and reliable method to track medication compliance over time based on self-reporting, as it would be required for a report card. A possible approach would be to combine patient self-reports, analysis of drug claims for refills patterns and medication adherence self-efficacy.

Documenting adherence to dietary restrictions based on self-reporting is fraught with similar measurement problems. It also requires asking participants about their compliance with several different restrictions, depending on their constellation of diseases (e.g., low cholesterol, low salt and a bread unit schedule). To avoid asking directly for compliance but to document eating patterns based on self-report, Food Frequency Questionnaires have been developed that ask

⁸ Target recommended by the U.S. Surgeon General (http://www.surgeongeneral.gov/topics/obesity/calltoaction/fact_advice.htm accessed September 30, 2004)

patients how often they consumed certain food items⁹. But their reliability has also been questioned (Schaefer et al. 2000). An even more involved method is a food diary that requires patients to document their food consumption on a daily basis. Further research would be necessary to determine which method could realistically be applied to disease management population and how to construct measures from the data.

Congestive heart failure (CHF)

- Proportion of participants who measure their weight daily¹⁰. Alternatively, this measure could be specified as the average number of days that participants measured their weight?

Diabetes mellitus

- Proportion of diabetics who self-monitor blood glucose (SMBG) at least daily¹¹

Feasibility

Because these measures rely on patient self-report, they require dedicated data collection. Most of the required data elements are currently collected by disease management operators or could be collected with limited additional effort.

Operational Issues

⁹ See for example, <http://www.jhbmc.jhu.edu/weight/forms/foodfrequency.pdf> accessed October 7, 2004.

¹⁰ As recommended by the 2002 ACC/AHA guideline (http://www.acc.org/clinical/guidelines/failure/hf_index.htm accessed September 30, 2004)

¹¹ According to a recent guideline, this is a minimum requirement. More frequent tests are required for type I diabetics, type II diabetics on insulin and patients who modify their regimen. However, the role of SMBG in type II diabetics who are stable on diet alone is not established. (Goldstein DE, Little RR, Lorenz RA, Malone JI, Nathan DM, Peterson CM. Tests of glycemia in diabetes. Diabetes Care 2004 Jan;27(Suppl 1):S91-3.)

The nurses soliciting the required information from disease management participants would have to strictly adhere to the pre-defined format of the questions to ensure reliable data collection.

5.5 PROCESSES OF CLINICAL CARE

Relevance

Measures of clinical processes of care, or process measures, convey whether care is provided in accordance with established medical standards, usually by reporting the percentage of opportunities in which the appropriate clinical process is delivered. While they represent an intermediate result of the intervention, reporting them to clients would be desirable for two reasons. First, process measures are more sensitive to changes in care than outcomes measures and will thus be able to capture an effect of the intervention earlier and with greater statistical power. Second, a key requirement in the rigorous development of process measures is that there is empirical evidence or at least professional consensus that the selected care processes have a meaningful effect on outcomes, so that the required link between improved processes as intermediate results and outcomes of care is present.

Measures Availability

Given the great interest in quality of care, a large number of process measures have been developed in recent years. For the most common conditions, there is even consensus emerging about which process measures to use. To some degree this consensus is implicit, i.e. most researchers and measures developers in a given field keep measuring the same construct with slight variants of measures. In some areas, there have been explicit consensus processes. For example, the National Diabetes Quality Improvement Alliance, a voluntary collaboration of organizations¹² that are concerned about the care of diabetes patients,

¹². These organizations include: Agency for Healthcare Research and Quality; American Academy of Family Physicians; American Association

has formally converged on a core list of nine measures that address the most important aspects of good diabetes care¹³.

As the number of measures that could potentially be considered for inclusion in a disease management report card is quite large, we started by reviewing the measures for six major chronic conditions that are contained in three measurement systems: the measures currently used by CorSolutions, the measures proposed by the American Healthways/Johns Hopkins University collaboration and the applicable measures of RAND's QA Tools system, a comprehensive quality measurement system with 439 process measures. The six conditions are:

- Coronary artery disease (CAD)
- Congestive heart failure (CHF)
- Hypertension
- Chronic obstructive pulmonary disease (COPD)
- Asthma
- Diabetes mellitus

We assessed for each unique measure, i.e. a measure that was designed to capture a distinct process of care, whether the clinical evidence supporting the measure was adequate and whether it was plausible to hold a disease management program accountable for the performance on this measure.

We found substantial overlap between the three systems in that 17 (25%) of the 69 potentially applicable measures were included in all three systems and 19 (27%) were included in two systems, albeit usually with slightly different operational definitions. Based on our review, we recommend including 53 clinical process measures for the six conditions

of Clinical Endocrinologists; American College of Physicians; American Diabetes Association; American Medical Association; Centers for Disease Control and Prevention; Centers for Medicare and Medicaid Services; Joint Commission on Accreditation of Health Care Organizations; National Committee for Quality Assurance; National Institute of Diabetes and Digestive and Kidney Diseases; The Endocrine Society; U.S. Department of Veteran Affairs.

¹³ <http://www.nationaldiabetesalliance.org/Final2004Measures.pdf>
accessed October 15, 2004

in a disease management report card. All conditions except hypertension are well covered by the recommended set of indicators, as each condition has at least 9 measures that capture aspects of prevention, diagnosis and treatment of the respective condition. To comprehensively reflect clinical processes in care for hypertensive patients that are under control of a disease management program, additional research would be needed, as only one applicable measure for treatment of hypertension could be identified¹⁴. On the other hand, one could argue that measuring blood pressure control alone would be sufficient for quality measurement purposes. Table 5.1 lists those measures. A summary of how those measures map into the three systems and the rationale for our recommendations can be found in Appendix B.

Table 5.1. Clinical Process Indicators

¹⁴ A search of the National Quality Measures Clearinghouse also failed to yield additional applicable measures.

Coronary Artery Disease [n = 12]

Prevention
Proportion of participants who receive smoking cessation counseling
Proportion of participants who receive screening for diabetes
Proportion of participants who receive a flu vaccination
Proportion of participants who receive pneumococcal vaccination
Proportion of participants who receive depression screening
Diagnosis
Proportion of participants who receive fasting lipid level
Proportion of participants who receive LDL screening
Proportion of participants who receive LV function test after AMI
Treatment
Proportion of participants with beta blocker usage
Proportion of participants in compliance with antiplatelet therapy
Proportion of participants who receive lipid lowering therapy
Proportion of participants who receive ACEI/ARB

Congestive Heart Failure [n = 11]

Prevention
Proportion of participants who receive flu vaccination
Proportion of participants who receive pneumococcal vaccination
Proportion of participants who receive warfarin
Proportion of participants with atrial fibrillation and/or prior thromboembolic event who receive warfarin
Proportion of participants who receive depression screening
Proportion of participants screened for depression and referred for follow up if at risk
Treatment
Proportion of participants with beta blocker usage
Proportion of participants with vasodilator usage
Proportion of participants with LV EF measurement
Proportion of participants on ARB/ACEI who receive annual creatinine checks
Proportion of participants on ARB/ACEI who receive annual potassium checks

Hypertension [n = 1]

Treatment
Proportion of participants who receive depression screening

Chronic Obstructive Pulmonary Disease [n = 9]

Prevention
Proportion of participants who receive flu vaccination
Proportion of participants who receive pneumococcal vaccination
Proportion of participants who receive depression screening
Diagnosis
Proportion of participants who receive spirometry testing
Treatment

Proportion of participants on bronchodilator
Proportion of participants with steroid inhaler use
Proportion of participants who receive oxygen therapy O2 saturation in below 88% at rest
Proportion of participants on bronchodilators who receive ipratropium
Proportion of participants who receive spacer use or proper MDI instructions

Asthma [n = 11]

Prevention
Proportion of participants who receive flu vaccination
Proportion of participants who receive pneumococcal vaccination
Proportion of participants who receive depression screening
Diagnosis
Proportion of participants who receive spirometry testing
Proportion of participants on theophylline with a daily dose of >= 600mg who receive routine theophylline level checks
Treatment
Proportion of participants on beta agonist or anticholinergics
Proportion of participants who receive inhalable steroids for uncontrolled asthma
Proportion of participants who receive appropriate use of long-term control medication
Proportion of participants who receive prescription of rescue inhaler
Proportion of participants with moderate to severe asthma in compliance with contraindication to beta-blockers
Proportion of participants who receive proper instructions of MDI use or spacer

Diabetes Mellitus [n = 10]

Prevention
Proportion of participants receiving lipid testing
Proportion of participants having annual foot exam by physician
Proportion of participants who receive flu vaccination
Proportion of participants who receive pneumococcal vaccination
Proportion of participants who receive ASA prophylaxis
Proportion of participants who receive depression screening
Diagnosis
Proportion of participants having dilated eye exams annually
Proportion of participants having microalbumin testing
Proportion of participants receiving biannual HbA1c testing
Treatment
Proportion of participants who receive ACEI/ARB for albuminuria

Feasibility

A total of 29 of the 53 proposed measures (55%) could potentially be constructed from claims data. Thus, any disease management vendor with access to and processing capabilities of insurance claims data could implement those measures. The remaining 24 measures (45%) require data elements that are not typically available from claims and therefore require dedicated data collection. Since many of those measures, or variants of them, are typically being reported disease management operators, we believe that all of them can be derived from data that are currently collected or could be collected with reasonable effort.

Operational Issues

We recommend that all process measures should take the form "number of eligible patients receiving a given care process" divided by "number of patients eligible for the process" rather than as relative improvement over a baseline value for two reasons. First, baseline performance will only be available for the measures that can be constructed from claims data, while data for the remaining measures will only be available for intervention years. Second, expressing performance as relative change would fail to bring out measures with high compliance rates, because relative improvements will decrease the closer one get to full compliance, which is also referred to as ceiling effects.

5.6 MORBIDITY REDUCTION

Relevance

Reducing the burden of disease, in particular for patients with chronic conditions, is an explicit goal for disease management. Measuring patient outcomes to evaluate whether disease management achieves this goal is consequently of great relevance for current and potential clients of disease management vendors. Morbidity reduction can be assessed proximally with measures of disease control in patients with chronic disease or proxy outcomes (e.g., glycemic control in diabetics), more distally with true outcome measures for those patients (e.g., lower

extremity amputation rates in diabetics) and in the very long run with measures for the incidence and prevalence of chronic conditions (e.g., diabetes prevalence). We would recommend focusing on measures of disease control for disease management reporting for three reasons. First, the more distal measures will not be very sensitive to the intervention, since the required end points occur at a very low rate in a given observation period, leading to power problems. Second, an effect of disease management on those measures can only be expected after years, and precise expectation for those time horizons would have to be derived from empirical evidence. Third, factors other than the quality of the disease management intervention, in particular baseline patient risk, would influence those measures, requiring elaborate risk adjustment procedures.

Measures Availability

There now exist commonly used and widely accepted measures of disease control for several chronic conditions and we recommend incorporating the 10 measures in a disease management report card. All of them are either currently reported by CorSolutions, part of the American Healthways/Johns Hopkins University measurement system or recommended as criteria for disease control by leading scientific organizations.

However, we could not identify any suitable existing measure for disease control in COPD and asthma. Also, lacking are measures for exercise capacity that would be very important in assessing beneficial effects for patients with CHF and COPD. Exercise capacity has traditionally been assessed with semi-quantitative, self-reported measures, such as number of blocks a patient can walk, but those have obvious reliability issues (Enright 2003). The American Thoracic Society has now officially endorsed a standardized six-minute walk test to measure treatment response in patients with moderate to severe cardiac

and pulmonary disease¹⁵. The measure has been extensively tested and validated and would be well suited to track the effect of the disease management intervention, but it would require testing patients at their providers and reporting the results back to a disease management operator on a regular basis, which does not seem to be feasible.

Coronary Artery Disease:

LDL cholesterol at target level (<100 mg/dl)¹⁶

Blood pressure at target level (<140/90 mmHg)¹⁷

Admission rate for angina without procedure¹⁸

Congestive Heart Failure:

Blood pressure at target level (<140/90 mmHg)¹⁹

30-day hospital re-admission rate²⁰

Hypertension

Blood pressure at target level (<140/90 mmHg)²¹

¹⁵ 2002 ATS Statement
(<http://www.thoracic.org/adobe/statements/sixminute.pdf> accessed October 4, 2004)

¹⁶ This is the currently recommended target by the National Lipid Association (<http://www.lipid.org/clinical/articles/1000015.php> accessed September 28, 2004)

¹⁷ This is the target currently recommended by the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7)
(<http://www.nhlbi.nih.gov/guidelines/hypertension/express.pdf> accessed September 28, 2004)

¹⁸ This measure is part of the AHRQ QIs
(http://www.qualitymeasures.ahrq.gov/summary/summary.aspx?doc_id=4632&string=cad accessed September 30, 2004)

¹⁹ This is the target currently recommended by the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7)
(<http://www.nhlbi.nih.gov/guidelines/hypertension/express.pdf> accessed September 28, 2004)

²⁰ Recommended by the Canadian Cardiovascular Outcomes Research Team/Canadian Society of Cardiology
(<http://www.ccort.ca/CCORTCCSCHFabridged.asp> accessed September 28, 2004)

Chronic obstructive pulmonary disease (COPD)

There are several measures for disease severity that have been established for COPD. A common pulmonary function parameter is the FEV₁, reflecting airway resistance, which is of great importance for determining the stage of the disease in the initial diagnosis²². Patient self-reported measures have been developed to complement those test results, such as the Baseline Dyspnea Index (BDI) and the Medical Research Council (MRC) scales²³. More recently, a multidimensional index has been proposed, which integrates pulmonary function and other test results, patient symptoms and functional capacity and general health status measures²⁴. But none of those parameters have been supported as measures of disease control so far. In fact, the use of pulmonary function tests to track COPD management has been explicitly questioned (Celli 2000). Thus, no well-established measure for disease control is available for COPD that could be recommended for a disease management reporting system. Further work would be needed to establish a measure.

Asthma

Both pulmonary function tests and patient self-reported health and functional status²⁵ are acknowledged to be of great relevance in the

²¹ This is the target currently recommended by the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7) (<http://www.nhlbi.nih.gov/guidelines/hypertension/express.pdf> accessed September 28, 2004)

²² See for example, The 2004 Pocket Guide of the Global Initiative for Chronic Obstructive Lung Disease (<http://www.goldcopd.com/> accessed October 1, 2004).

²³ Mahler DA and Wells CK. Evaluation of clinical methods for rating dyspnea. *Chest*, 1988, Vol. 93, p. 580-586

²⁴ Celli BR, et al. The Body-Mass Index, Airflow Obstruction, Dyspnea, and Exercise Capacity Index in Chronic Obstructive Pulmonary Disease. *NEJM* 2004, Vol. 350, p. 1005-1012

²⁵ The National Asthma Education and Prevention Program Expert Panel Report 2 recommends periodic assessment of functional status, signs and symptoms but does not provide a standardized assessment instrument.

monitoring of asthma (Celli 2000). However, no specific measure for these constructs has been proposed thus far and established as valid and reliable indicator of disease control so that we cannot recommend a measure for asthma control, but suggest developmental work in this area.

Diabetes mellitus

Adequate glyceic control (HbA1c <7.0%)²⁶

Poor glyceic control (HbA1c >9.0%)²⁷

LDL cholesterol at target level (<100 mg/dl)²⁸

Blood pressure at target level (<130/80 mmHg)²⁹

Feasibility

Most of the measures will demand dedicated data collection, because neither the required level of clinical detail nor test results are commonly available from administrative data. Per our initial assessment, the required variables are typically collected by disease management operators.

Operational Issues

The target levels in all listed measures must be reviewed on a regular basis to keep them in line with the latest recommendations. For

²⁶ This is the target currently recommended by the American Diabetes Association (http://www.guideline.gov/summary/summary.aspx?doc_id=4679&nbr=3413&string=diabetes accessed September 28, 2004)

²⁷ This is the criterion for poor control currently used by the National Committee for Quality Assurance (http://www.qualitymeasures.ahrq.gov/summary/summary.aspx?doc_id=457&string=diabetes+AND+ncqa accessed September 28, 2004)

²⁸ This is the target currently recommended by the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7) (<http://www.nhlbi.nih.gov/guidelines/hypertension/express.pdf> accessed September 28, 2004)

²⁹ This is the target currently recommended by the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure (JNC 7) (<http://www.nhlbi.nih.gov/guidelines/hypertension/express.pdf> accessed September 28, 2004)

patients with multiple chronic conditions, the most stringent target should be used, e.g. the target blood pressure for a diabetic with hypertension should be 130/80 mmHg. We recommend that all disease control measures should take the form "number of eligible patients meeting target" divided by "number of eligible patients".

5.7 REDUCTION OF UTILIZATION AND DIRECT MEDICAL COST

Relevance

As optimizing utilization and reducing medical cost are the very purpose of disease management interventions, the relevance of measures in this category is high for clients and potential clients.

Measures Availability

All Disease Management vendors report cost and utilization measures to current and future clients, typically disaggregated by types of service and patient condition. We recommend using the following measures that are similar to the ones currently used by CorSolutions and those proposed by American Healthways/Johns Hopkins University.

- Number of physician office visits per 1000 participants per year (overall, overall observed minus expected, by condition)
- Number of ER visits per 1000 participants per year (overall, overall observed minus expected, by condition)
- Number of hospital admissions per 1000 participants per year (overall, overall observed minus expected, by condition)
- Number of hospital days per 1000 participants per year (overall, overall observed minus expected, by condition)
- Number of drug claims per 1000 participants per year (overall, overall observed minus expected, by condition)
- Total medical cost PMPM (overall, overall observed minus expected, by condition)
- Total prescription drug cost PMPM (overall, overall observed minus expected, by condition)
- Total inpatient cost PMPM (overall, overall observed minus expected, by condition)

- Total ER cost PMPM (overall, overall observed minus expected, by condition)
- Total outpatient cost PMPM (overall, overall observed minus expected, by condition)

Feasibility

All recommended measures could be calculated from standard health insurance claims and enrollment data. Estimating expected cost requires software for statistical modeling and patient risk stratification.

Operational Issues

A. Data Cleaning

Like others that have addressed this issue (Linden 2003), we recommend that cost and utilization measures should be constructed for full observation years rather than shorter time periods to avoid seasonal effects. A run-off period of at least three months should be allowed after each observation year to ensure that the claims database is complete.

Also, disease management vendors should request information about the claims adjudication process from the originator of the data. There are three different ways how adjudication can be reflected in claims data:

1. Overwrite the initial claim with a corrected one
2. Add a new claim with the corrected amount to the claims file
3. Add a new claim with the amount to be adjusted for to the claims file

Those three adjustment procedures imply different ways of handling potential duplicate claims, i.e. claims with the same service date and identical provider, service and diagnosis codes. The first procedure should not lead to such duplicates, the second would imply that the claim with the latest file date should be used, and the third that the sum over such claims should be calculated to get to the paid amount.

B. Estimation of Cost and Utilization Reduction

The challenge of attribution is particularly salient for those measures, as they are at the core of performance reporting for disease management programs. Changes in utilization and spending can result from many other factors or confounders than the disease management intervention, such as changes in patient mix, available technology (e.g., new and expensive drugs and devices entering the market), benefits design (e.g., changes in co-payments or stricter utilization review by a health plan) and concurrent interventions (e.g., nurse advice lines offered by a health plan). As mentioned above, disentangling the effect of disease management from the effect of such others factors requires a comparison group that corrects for those factors. Without a comparison group attribution is challenging, further complicated by the fact that attrition may result in substantial changes from the pre-intervention population to the population of the intervention years. Year-on-year attrition rates for health plans are known to be as high as 30%, and, while job tenure tends to be longer, turnover can be substantial

We recommend to use statistical modeling adjust for the impact of patient mix, i.e. changes in disease severity and demographics, to control for at least one important set of confounding variables. The approach is described in full technical detail in Appendix C.

In short, we determine in each year whether a given patient has or has not received care for a given condition and how severe the condition was based on their medical claims and create so-called disease markers. Those disease markers will be derived from commercially available software, such as the ERGs or DxCGs. We estimate in the baseline year, i.e. the year prior to the disease management intervention, to what degree the presence of a condition of given severity, expressed by the disease markers, influenced the cost and utilization of this patient. In other words, the disease markers serve as multipliers that increase or decrease cost for a given patient compared to a baseline and the model estimates how large a multiplier each disease marker is. The multipliers are then applied to all patients in the intervention year(s), i.e. we calculate their predicted cost/utilization based on their diagnoses in those years. This approach allows adjusting for changes in the patient

population's demographic and casemix over time, even in the presence of attrition, and represents a substantial improvement of an unadjusted pre-post comparison.

However, two substantial limitations remain: First, the method can be described as reflecting the expected cost of the intervention-year population at baseline-year conditions. The baseline-year conditions are influenced, among others, by available technology, benefit structure and other cost-containment measures. To the degree that the conditions change, the estimate provided by the method may over- or underestimate the effect of the intervention. Second, to the degree that the disease markers are influenced by utilization patterns, the categorization of patients in the intervention year can already reflect an effect of the intervention. This would bias the estimates downward. Thus, the estimates derived by this approach are likely to constitute a lower bound of the actual effect and need to be interpreted in conjunction with information about secular cost trends.

C. Defining the Analytic Sample

Much of the controversy around disease management has resulted from measuring performance based on biased samples of patients (Lewis, 2003). Bias occurs if an unobservable or unmeasured characteristic of a patient makes it simultaneously more or less likely that s/he has a positive/negative treatment effect and a higher/lower probability of responding well to the intervention. For example, in the early days of disease management it was quite common to only include patients in the analysis who agreed to participate in the program. It is reasonable to assume that this participant cohort has a higher, unobservable or unmeasured, and intrinsic motivation to take care of their health. Thus, they are more likely to benefit from an intervention than the overall patient population and the described comparison strategy overstates the true effect of the intervention.

Also, some disease management vendors used to include only patients with the high cost in the baseline period in their analysis, as they concentrated their efforts on this group. Those may be patients who had been hospitalized in the base period. However, since high-cost events in

medical care tend to be non-recurring, a certain proportion of those patients would end up with lower cost in the next period regardless of the intervention, commonly referred to as regression to the mean. Consequently, this method tends to overstate the results of the intervention. Several other similarly problematic methods have been used, such as restricting the type of claims to be considered or using very short comparison periods. For example, some disease management vendors have restricted the claims included in their analyses to those that are specific to a disease, such as annual eye examinations for diabetics. This type of limitation ignores the impact that a specific disease has on other aspects of health and does not provide a true picture of the change in a person's health and their costs (Stone, 1999).

A solution has been suggested by American Healthways/Johns Hopkins University (American Healthways 2002). Their methodology suggests that each eligible patient should be included in the analysis (intent-to-treat), that every claim be counted and that a full year's worth of claims for each patient should be analyzed in the base and intervention periods (American Healthways 2002), thus eliminating many sources of bias.

One source of bias, however, remains possible, because the method includes patients only after they have revealed themselves as having the disease by filing a claim with the respective diagnosis, but retains them in the sample, even if they do not file a claim in the following period(s)³⁰. The consequence is that the savings estimate may be biased upward: no patient with zero claims is included in the base period, but some patients with zero claims are included in the intervention period. The estimated cost per patient will therefore be lower in the intervention period, even if nothing else changes and the magnitude of

³⁰ AH/JHU proposed three different possible populations. Our discussion and illustration specifically refer to their "continuous population." Their other two proposals, a "measurement period population" and a "dynamic population," also suffer from biases due to the fact that a person eligible for the DM intervention is only included in the analysis once they have a medical claim.

the bias depends on how many patients move between zero claims and positive claims between the two periods.

To address this bias or to get at least an idea of its magnitude, the Disease Management Purchasing Consortium (DMPC) recently proposed a diagnostic method, the DMPC diagnostic (Lewis 2003). It compares the cost in the respective high-cost categories in the base and intervention years, based on the argument that effective disease management should reduce the probability of patients ending up in this category and reduce cost even if they do. While the method has some plausibility because most medical spending is highly concentrated in a few patients and because it avoids the differential dilution problem of the American Healthways/Johns Hopkins University method, some questions remain. First and foremost, as the authors themselves state, the approach is a diagnostic to estimate the bias introduced by the American Healthways/Johns Hopkins University method, not an independent method to estimate cost savings. Second, it is highly sensitive to the definition of the high-cost group and requires the ability to clearly differentiate those patients.

We thus recommend overcoming this problem by using a true population-based approach, which includes **all hypothetically eligible** patients or the entire population of a given client regardless of whether they have a claim for a disease in either observation period³¹. Since it is devoid of any selection into the analytic sample, this source of bias is eliminated. The main challenge to this method is the problem of power or the ability to detect an effect, because the group actually receiving the intervention represents just a small part of the analytic sample, leading to a potentially large noise to signal ratio. Thus, to successfully employ this method three conditions have to be met:

³¹ Conceptually, this approach is similar to a common industry practice to use predicted cost for patients without any of the managed conditions as a benchmark to compare changes in the managed population against. The RAND approach, however, consolidates those two steps into one estimation procedure.

1. Statistical adjustments, as described above, must be used to account for changes in case mix and severity of managed and unmanaged conditions.
2. The disease(s) under management must be prevalent enough to ensure sufficient power³².
3. Formal power calculations have to be conducted to ensure that the sample size is large enough to provide sufficient statistical power³³.

While the method does add considerably to the analytic complexity in the design, the steps can largely be automated, and it has some theoretically attractive properties:

1. As mentioned above, many sources of bias are eliminated.
2. It can accommodate health promotion programs, whose participants may have zero claims in the base period.
3. As it uses individual level data, it allows both to assess overall effects on the total population and the effects of distinguishable intervention components and effects on patient groups. For example, if an employer client uses several vendors simultaneously to manage the health of its workforce (such as health risk assessment and disease management), unique contribution of those different programs can be captured in the model³⁴. Similarly, subgroup analyses, such as diabetics only, can be conducted.
4. It is consistent with the evaluation designs that the Medicare system tends to prefer, as the evaluator of Medicare demonstrations is usually asked to assess the

³² This will be the case for the common chronic conditions that are typically managed, but not for rare disorders.

³³ This should not be a problem for health plan clients, but smaller employers may not have a sufficiently large sample.

³⁴ The following types of intervention components are commonly used: health risk assessment and management, lifestyle management, demand management/nurse line, disease management and case management.

effect on the full Medicare population regardless of eligibility or participation.

However, before adopting this method, we recommend empirically evaluating whether the gain in precision of the estimates justifies the increase in complexity. In other words, if the savings estimates derived by this method were close to the ones derived by a simple pre-post comparison, one would conclude that the magnitude of selection bias is small and could be ignored for practical purposes. If this were the result of the empirical evaluation, we would not recommend that disease management vendors and operators use such an elaborate methodology. As mentioned, any method that evaluates disease management effects without the use of a control group is prone to error and should be benchmarked against results derived from a controlled trial.

5.8 IMPACT OF HEALTH ON PRODUCTIVITY

Relevance

The cost of chronic illnesses to employers is not limited to direct medical costs, but also includes lost productivity due to days in which employees are absent (absenteeism) or working at a reduced capacity (presenteeism) because of either their own diseases or their need to take care of family members. Several studies have suggested that the cost of lost productivity may be several times greater than direct medical costs and that presenteeism generates a larger proportion of losses than absenteeism (Loeppke et al, 2003; Goetzel et al, 2004a; EHC, 1999). For example, in a meta-analysis of seven studies that estimated productivity losses from ten costly conditions³⁵ with different instruments (Goetzel et al, 2004) the overall cost of presenteeism was found to range from one fifth to three fifths of the total dollars lost to the various conditions, including costs due to absenteeism and direct medical costs. The estimates varied widely, however, depending on the

³⁵ The conditions included in the meta-analysis were allergies, arthritis, cancer, depression/sadness/mental illness, diabetes, heart disease, hypertension, migraines/headaches, and respiratory disorders.

disease and instrument used. The average presenteeism loss per employee per year across instruments and diseases was 12% of full productivity, with a low of 5.7% and a high of 17.9%. The average absenteeism loss was 4.3% of full productivity with a low of 0.8% and a high of 10.8%. Heart disease caused the smallest presenteeism loss with 6.8% and migraines or headaches the highest with 20.5%. Hypertension led to the lowest absenteeism loss with 0.4% and depression/ sadness/mental illness led to the highest with 10.7%. There was also considerable variance in the calculated loss depending on the instrument used: a range of 0.6%-14.0% for absenteeism and 10.4%-15.8% for presenteeism. The results of another study underscore the importance of the presenteeism portion of overall productivity loss: it found that days lost due to presenteeism were 7.5 times the number of days lost due to absenteeism when seventeen of the most prevalent conditions³⁶ in the workplace were considered (EHC, 1999). Thus, while there is substantial variation across diseases and studies, the available evidence underscores the relevance of health-related productivity loss for employers and the utility of such measures for a disease management report card. Ideally, measures for health-related productivity losses would capture both absenteeism and presenteeism and express those two parameters in their natural units as well as in monetary units. Substantial methodological and data availability issues, however, limit our ability to accurately measure those constructs.

Measures Availability

We conducted literature and non-literature searches to identify existing instruments to measure the impact of health on productivity, through both absence from work (absenteeism) and reduced performance

³⁶ The conditions in the EHC study, by prevalence in the workplace in 1999, were as follows: allergy, hypertension, conditions involving the neck/upper back/spine, arthritis, conditions of the lower back, sciatica, depression, peptic ulcer/acid reflux, migraine, other respiratory conditions [than asthma], diabetes, asthma, heart disease, high-risk pregnancy, hepatitis, breast cancer, prostate cancer, and colon cancer.

while at work (presenteeism). We also retrieved related material for each identified instrument, such as information on assessment of reliability and validity. In addition, we reviewed methods to derive monetary estimates from those instruments. Given the recent and rapidly developing nature of this field, we also conducted interviews with five recognized experts to help us put our findings into perspective and shed light on current research trends.

1. Measuring Absenteeism

Two different methods are used to gather data information on absenteeism: direct measurement (e.g., gathering days lost from payroll logs, etc.) and self-reporting. There are limitations and benefits to both methods. Direct measurement generates more reliable results, but tends to be hard to implement because most companies do not routinely collect data on days lost from work for each employee. Self-reported data can easily be gathered by surveys and have been found to be reliable and valid when the recall periods are short, i.e. one or two weeks vs. one month (Revicki et al., 1994). Consequently, either employer-provided or self-reported days lost from work can be included as an absenteeism measure in a disease management reporting system.

2. Measuring Presenteeism

Measuring presenteeism is obviously a more complex challenge than measuring absenteeism, as reduced performance on the job is less tangible than absence. Some attempts have been made to measure presenteeism directly, e.g., by call volume per employee in a call center (Burton et al., 2004). But generating objective data would require developing methods in partnership with each employer to suit the particular characteristics of a given firm, workplace and profession, and collecting data on a regular basis as well. Developing such methods may prove impossible for white-collar positions.

To overcome these obstacles, various self-report survey instruments have been developed that can be applied to various professions and employers (Lerner et al., 2001; Kessler et al., 2004). We identified 20

such instruments (Table 1 in Appendix D) that use three distinct approaches of measuring presenteeism:

- Perceived impairment
- Comparison of productivity/performance/efficiency with that of others and one's norm
- Estimate of unproductive time while at work

2.1. Perceived Impairment

Asking employees to rate how hindered they feel in performing common mental, physical, and interpersonal activities and meeting demands due to their illness is the most common approach found in the presenteeism instruments currently available. Tools that follow this approach include the Health and Productivity Questionnaire (HPQ), the Health and Work Questionnaire (HWQ), the Stanford Presenteeism Scale (SPS), the Work Limitations Questionnaire (WLQ), and the Work Productivity and Activity Impairment Questionnaire (WPAI). Questions about perceived impairment can be very general or very specific. An example of a very general question can be seen in the SPS: "Despite having my (health problem)*, I felt energetic enough to complete all my work." The employee receiving the SPS is invited to respond to this question using a five-point scale that goes from "Strongly disagree" to "Strongly agree." An example of a specific question is found in the WLQ, which requires a respondent to rank the difficulty he or she had in using "upper body to operate tools, equipment" on a five-point scale.

Perceived impairment questions are the most direct form of getting someone to describe their presenteeism without requiring the respondent to estimate their lost performance or lost time as a consequence of that impairment. Some studies have attempted to validate respondents' recollections against diaries (Lerner et al.), but no attempts have yet been made to verify that a respondent's perceptions correspond with fact. Thus, their reliability and validity is not fully established. In addition, it is difficult to translate someone's agreement or disagreement with statements about his or her perceived impairments into an estimate of actual productivity reduction.

2.2. Comparative productivity/performance/efficiency

Another way one can capture presenteeism is by understanding how an employee's performance differs from that of others and from some conception of his or her usual performance. Tools that take this approach include the HPQ and the HWQ. The HWQ asks a respondent to rate the overall quality and amount of work produced in the preceding week, as well as how efficiently it was done, by themselves, their supervisor, and their co-workers on a ten-point scale that goes from "worst ever" to "best possible." The HWQ additionally asks the respondent to rank their highest and lowest levels of efficiency during the week on the same ten-point scale. The HPQ works in a very similar manner. It asks the respondent to rate the performance on the job of workers in similar positions, their usual performance in "the past year or two," and their overall performance during the recall period (four weeks) using a 10-point scale that ranges from "Worst Performance" to "Best Performance." Both the HPQ and the HWQ include these comparative performance questions in addition to questions about perceived impairments.

Compared with measures of perceived impairment, measures of an employee's perceived overall performance have three main advantages when it comes to calculating presenteeism as a single meaningful number. Firstly, the attempt to "anchor" one's perceived performance with that of others, one's average, one's best, and/or one's worst allows for the idea of a standard level of performance against which loss can be measured. The questions about perceived impairment do not include any conception of what is a standard or usual level of impairment. Secondly, a 10-point performance scale can more easily be used in a monetization formula than the agreements or disagreements with statements that one sees as the norm in impairment questions, though one would imagine that perceived performance may still have to be turned into a temporal measure before being monetized and this could introduce error. Finally, attempts have been made to validate employees' self-reported performance evaluation by comparing it to their supervisor's assessments (Kessler et al, 2004). This external check of a person's perceptions lends more credence to the performance measures compared to the impairment ones.

2.3. Estimates of Unproductive Time

Relatively few instruments approach presenteeism in the same manner as absenteeism and ask employees to estimate lost time, but some do. One example is the Work Productivity Short Index (WPSI). The WPSI includes questions which ask the employee to estimate how many unproductive hours they spent at work during the recall period.

Although this approach would lead to the easiest monetization, validation remains an issue as no study has shown that employees can accurately transform their perceived impairments into a temporal measure. Also, unlike the measures of comparative performance, measures of estimated unproductive time do not seem to provide the respondents with any way to anchor their responses against usual or expected unproductive time and the amount of time that similar employees are perceived as being unproductive.

Assessment

We identified 20 different survey instruments for measuring absenteeism, presenteeism, or both (Table 1 in Appendix D). We excluded 6 (30%) of them because we could not identify any research that would support their reliability and/or validity, 8 (40%) because they were designed specifically for a particular disease, reducing their value for a disease management report card, and one because we could not retrieve a copy of the copyrighted survey instrument. The remaining five instruments were reviewed in detail (see Table 2 in Appendix D). There were substantial differences with respect to length (ranging from 6 to 31 items) and content (ranging from asking about physical and mental behavior changes to asking about lost hours and reduced productivity). Three of them ask for estimates of unproductive time, the Health and Productivity Questionnaire (HPQ), the Work Limitations Questionnaire (WLQ), and the Work Productivity and Activity Impairment (WPAI). The two other instruments solicit information on perceived impairment, the Health and Work Questionnaire (HWQ) and the Stanford Presenteeism Scale (SPS-6).

3. Cost Estimation Methods

In addition to the numerous problems that one encounters as one moves from straightforward measures of time lost through absenteeism to more complicated measures for quantifying presenteeism, there are many competing methodologies for monetizing lost productivity. These methodologies come in three basic flavors: (1) salary conversion methods that use survey responses and salary information to estimate productivity loss, (2) introspective methods that use survey responses as a basis for thought experiment to give businesses an idea of the magnitude of their lost productivity, and (3) firm-level methods that attempt to monetize productivity losses based on the cost of countermeasures that firms use to deal with absenteeism and presenteeism.

Salary Conversion Methods

Salary conversion methods attempt to estimate productivity losses based on self-reported lost time or decreased productivity but cannot be applied to instruments that measure perceived impairment. The simplest version is the Human Capital Approach (HCA), which expresses the loss as the product of missed workdays by daily salaries (Berger et al, 2001). Originally developed for monetizing absenteeism, the method has been extended to presenteeism losses by using self-reported unproductive hours or self-reported percentage reduction of performance instead of missed days (Lerner, 2001; Allen and Bunn, 2003a and 2003b). The obvious attraction of this method is its computational ease, its intuitive plausibility and its consistency with economic theory that, assuming perfectly competitive labor markets, wages should reflect a worker's marginal contribution to a firm's output. While its validity has not yet been assessed, there was consensus among our interviewed experts that the HCA does provide at least a lower bound for the true cost of lost productivity (Tom Parry, 2005; Sean Nicholson, 2005). One expert suggested using also salary plus the cost of fringe benefits to estimate productivity losses (Kessler, 2005). The HCA is also the typical method behind studies reporting the economic impact of health-related

productivity losses. Depending on the available data sources, authors have used actual salaries of the respondents (Stewart et al., 2003a and b), corporate average salaries (Hemp, 2004) or national median wages (Goetzel et al., 2004a) for the conversion.

An extension of the HCA is the team production model set forward by Pauly and co-workers (2002), who argued that simple salary-based conversion was appropriate for workers performing discrete tasks in isolation but failed to take into account the interdependence of job functions in the modern economy. For example, if the only surgeon in a hospital stayed home sick, the entire operating room would remain idle for the day, causing much greater losses than just the surgeon's salary. The authors proposed to operationalize this interdependence into three criteria: the replaceability of an employee, the extent to which an employee works as part of a team, and the time sensitivity of his or her work. Initial empirical work by Nicholson et al. (2004) derived a set of multipliers for 35 different job categories based on those three dimensions that can be applied to a worker's salary. Simple jobs, like a fast food cook, have a multiplier of 1.00, suggesting that the productivity loss equals the actual salary, while more demanding occupations, such as a construction engineer or a paralegal, have higher multipliers to reflect their overall impact on the firm. Different multipliers exist for short-term (3-day) and long-term (2-week) absences. Ongoing work aims at a larger set of multipliers and methods to capture the interaction between medical conditions and job characteristics (Nicholson, 2005).

Two practical challenges exist for this approach. First, a sufficiently large library of multipliers must be created and maintained for the method to be used in an operational setting. Second, the method is entirely based on individual-level characteristics and does not take firm-level factors into account. It is, for example, conceivable that the absence of a research assistant would have different implications for a consulting firm and a not-for-profit research organization. Other firm-level factors, like unionization and competitive position, may also modify the impact of a given job category.

A more fundamental challenge was posed by Koopmanschap and co-authors (1995), who argued that the HCA overestimated the absence-related productivity losses in reality, because short-term absences could be partially compensated with higher effort or overtime upon return of the sick employee or by co-workers. Longer-term absences would lead to replacement of workers with new hires. Based on those considerations, they proposed the friction cost method that aims at, in their terms, estimating only the actual lost production as opposed to the potentially lost production estimated by the HCA. They have tested their method on macroeconomic data from the Netherlands and found the estimates of lost productivity to be consistently lower than those derived by the HCA (Koopmanschap et al., 1995), but no attempt of applying the method to firm-level data could be identified in our search. Others have challenged the friction cost approach as inconsistent with economic theory, which would predict that profit-maximizing firms would not have idle reserve capacity (Johannesson and Karlsson, 1997). This discourse remains, however, largely theoretical at this point as neither of the two methods has been evaluated empirically.

Introspective Methods

Given the theoretical and practical challenges of finding a method to convert self-reported productivity reduction into monetary units, some researchers have argued that conversion should be abandoned in favor of providing guidance to firms with which those can derive their own cost estimates. Managers are provided with an analysis of the productivity survey and asked to develop scenarios such as "how much would you be willing to pay a contractor who can bring everyone's productivity to 100 percent?" or "how many FTEs could you cut if everyone worked at full productivity?" (Kessler, 2005). Another approach is to encourage managers to estimate the revenue that different staff members contribute and use this number for conversion (Parry, 2005). The aim for those thought experiments is to illustrate the magnitude of the problem rather than to derive precise estimates. But, while certainly

helpful, their validity remains untested and they have not yet been benchmarked against the HCA approach.

Firm-Level Methods

A logical extension of the introspective methods is to give up on estimation methods that are based on individual-level, self-reported data and to utilize a top-down approach that employs firm-level information to derive cost estimates for lost productivity. Managers, so the argument goes, have a fairly good sense of how their company's productivity is affected by health-related problems and use countermeasures to deal with those. For example, they may have redundant staff to compensate for absences; they may hire temporary workers or pay overtime to maintain output. Alternatively, they could forgo revenues. Economic theory suggests that a competitive firm combines these different strategies to maximize profits. Information on a firm's cost of those countermeasures can thus be used to approximate its lost productivity. The attraction of this approach is that it does not require detailed individual-level data and that the cost of many of the countermeasures is easy to quantify, such as the fees paid to temp agencies. The downside is that some of the cost may not be tangible and that foregone revenue estimation has to rely on manager's perceptions. It may also prove very difficult to elicit countermeasures to presenteeism as opposed to absenteeism, as the former is not immediately visible to a firm and may not provoke a conscientious response. Further, the correct attribution of the cost items to health-related productivity losses needs to be assured, since, for example, part of the temporary staff could also be part of a firm's usual staffing mix. As for the other methods, empirical evidence remains sparse. One study has used staffing cost to cover short-term disability absences to estimate productivity losses (Parry and Auerbach, 2001), but no attempts have been published to generalize this approach into a broader framework for measurement.

Assessment

None of the monetary conversion methods can be regarded as empirically proven at this point in time so that the effect of disease management on absenteeism and presenteeism should primarily be measured and reported in the natural units that the respective survey instrument provides.

To illustrate the economic magnitude of the problem, the available monetization methods can be used, but it needs to be emphasized that the results should be interpreted with caution. Because of its intuitive plausibility and operational ease, the HCA method is the most commonly used approach for this purpose and all interviewed experts agreed that it would provide a lower bound for the cost of lost productivity.

5.9 HEALTH-RELATED QUALITY OF LIFE

Relevance

Improving the health-related quality of life of patients with chronic conditions is one of the ultimate goals of disease management. Such measures would not just capture better health status but also the increased ability of patients to live with their disease. They are obviously of great relevance to patients and also allow purchasers to demonstrate that they procured a valuable service for their employees.

Measures Availability

Developed for the Medical Outcomes Study, a study of the impact on the differences in care on patient outcomes, the SF-36 has emerged as a universally accepted instrument for patient self-reported health status and health-related quality of life (Tarlov et al., 1989; Ware and Sherbourne, 1992). The instrument is not condition-specific and can thus be applied to various patient populations. It was extensively tested and validated in prior studies (McHorney et al, 1993). Based on the original 36-item instrument, shorter versions with 12 (SF-12) and 8 (SF-8) items have been designed and tested by the original developers. We recommend using the SF-8 in a disease management reporting system, as it is the

shortest of the tools and thus the one the least burdensome to collect for operational purposes.

Instruments to measure health-related quality of life have also been specifically developed for various chronic diseases. Well-established examples are the Kansas City Cardiomyopathy Questionnaire (KCCQ)³⁷ for CHF, the Seattle Angina Questionnaire (SAQ)³⁸ for CAD and the Chronic Respiratory Disease Questionnaire (CRQ)³⁹ for COPD. RAND has developed survey instruments for initial assessment and follow-up in patients with several chronic conditions, such as diabetes and CHF⁴⁰. While those instruments provide valuable information to determine the severity of disease and monitor treatment effects, they typically require scoring of 20 or more items, and seem therefore too burdensome for operational purposes.

5.10 PATIENT AND PROVIDER SATISFACTION

Relevance

Measures of patient and provider satisfaction are of great relevance for disease management clients. As they procure disease management services on behalf of patients, it is important for them to ascertain that the product met the acceptance of patients as the end users. Lack of satisfaction would also hint at difficulties of establishing a working relationship with program participants. Likewise, ensuring provider satisfaction is crucial for the success of a disease management program, because those programs have to be sensitive to the particular relationship between patients and their physicians and the

³⁷ Green CP, et al. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. *J Am Coll Cardiol.* 2000; 35(5): 1245-55

³⁸ Spertus JA, et al. Development and evaluation of the Seattle Angina Questionnaire: a new functional status measure for coronary artery disease. *J Am Coll Cardiol.* 1995; 25(2): 333-41.

³⁹ Guyatt GH, Berman LB, Townsend M, Pughley SO, Chambers LW. A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987; 42: 773-778

⁴⁰ <http://www.rand.org/health/ICICE/tools.html> accessed October 1, 2004

risk of antagonizing providers. Various satisfaction surveys exist that could be used and are being used for this purpose.

6. ISSUES IN CREATING A REPORTING FORMAT FOR THE MEASUREMENT SYSTEM

Providing such a complete performance map as outlined above will require disease management operators to invest in data collection, processing, analysis and reporting. However, implementing such a comprehensive performance system would add substantial value to disease management products. It would allow vendors and operators to demonstrate a broader value proposition to clients who have concerns beyond cost savings. It also has the distinct advantage of providing greater plausibility and credibility to end points, like total reduction in direct medical cost. For example, the reported data could illustrate how the intervention changes self-efficacy of patients early on, later leading to improvements in health-related behaviors and clinical processes, and finally to better health outcomes and reduction in spending, if all the measures are collected longitudinally on a set of patients with appropriate means of establishing attribution. It would also allow providing clients with early feedback on observable changes in some measures, while the effect on, say, direct medical cost has not yet materialized.

From a managerial standpoint, a comprehensive measurement system would help to shed light on reasons for underperforming programs or accounts or even allow proactively identifying performance problems at an early stage. Breaking down the results by operating units, such as call centers, is conceivable, but further disaggregation to groups of nurses or individual nurses would be problematic, because of sample size and attribution issues.

The complexity of the proposed measurement system, however, means that a large amount of information is being communicated to the users of the report card. This implies that considerable thought should go into the implementation and the design of a reporting format based on the finally set of measures selected. Some of the issues that need to be addressed are listed below.

6.1 MEASURE CALCULATION

- Missing data: How should missing data elements be handled? The options are to set the variable to a pre-defined value, to impute the value, to drop the indicators that require the variable for a participant or to drop each participant with any missing data.
- Stratification: Should the results be reported for all participants or by acuity level? If by acuity level, how is membership in a given group defined, as patients change between levels?
- Aggregation: Should the measures within a given category be aggregated into a summary measure? Aggregation facilitates communication of complex information but comes at the cost of loss of detail, which could be overcome with summary measures that allow drill downs.
- Weighting: Should aggregated measures be based on unit weights, i.e. each measure is weighted equally, empirical weights or expert weights?

6.2 MEASURE INTERPRETATION

- Scoring: How is performance based on an indicator or aggregate expressed? Options are change relative to previous values or relative to a benchmark, absolute change or compliance with a target or cutoff point.
- Incorporating uncertainty: As all items will be measured with error, what is the appropriate method to reflect the uncertainty that is embedded in a measure?
- Reporting: Should values or interpretative symbols (e.g., stars) be reported?

Appendix

APPENDIX

A. MEASURING SELF-EFFICACY

Self-Efficacy and Smoking Cessation

The relationship between self-efficacy and smoking cessation has been widely studied, with the results of these studies consistently showing that confidence in one's ability to abstain from smoking predicts the outcome of a smoking cessation attempt (Baer et al. 1986, Colletti et al. 1985, DiClemente 1981, DiClemente et al 1985, Godin et al 1992, Mudde et al 1995). Moreover, greater smoking abstinence self-efficacy is related to a decreased likelihood of relapse after quitting (DiClemente et al 1985, Condiotte & Lichtenstein 1988, Gulliver et al 1995, Yates and Thain 1985). The two instruments most commonly used to assess smoking abstinence self-efficacy are a 20-item scale created by DiClemente, Prochaska, and their colleagues, and a 44-item scale created by Condiotte and Lichtenstein. Both of these scales have been shown to be reliable and valid measures (Baer et al. 1986, DiClemente 1981, Condiotte & Lichtenstein 1988, Baer & Lichtenstein 1988, Prochaska et al 1985, Velicer et al 1990). A 9-item version of the scale created by Velicer and colleagues has also been validated (Velicer et al 1990); it is this instrument that we recommend for use by CorSolutions. In completing this instrument, respondents use a 5-point scale (1 = not at all confident to 5 = extremely confident) to judge how confident they are that they could avoid smoking in each of 9 tempting situations. An overall score is created by taking the average rating across the nine items.

Smoking Self-Efficacy (Short Form)

Listed below are situations that lead some people to smoke. We would like to know HOW CONFIDENT you are that you could resist smoking in each situation. Please answer the following questions using the following five-point scale.

- 1 = Not at all confident
- 2 = Not very confident
- 3 = Moderately confident
- 4 = Very confident
- 5 = Extremely confident

- 1. With friends at a party.
- 2. When I first get up in the morning.
- 3. When I am very anxious and stressed.
- 4. Over coffee while talking and relaxing.
- 5. When I feel I need a lift.
- 6. When I am very angry about something or someone.
- 7. With my spouse or close friend who is smoking.
- 8. When I realize I haven't smoked for a while.
- 9. When things are not going my way and I am frustrated.

Source: Velicer, W.F., DiClemente, C.C., Rossi, J.S., & Prochaska, J.O. (1990). "Relapse situations and self-efficacy: An integrative model", *Addictive Behaviors*, 15, 271-283.

Self-efficacy is related to the outcomes of individuals who attend weight loss treatment programs (Bernier, Avard 1986), as well as attempts to lose weight [29], weight loss maintenance (Bernier, Avard 1986, Kitsantas 2000, Westover & Lanyon 1990), and weight stability over extended periods of time (Foreyt et al. 1995). The two instruments most commonly used to evaluate self-efficacy for weight loss/maintenance are the Eating Self-Efficacy Scale (ESES) (Glynn & Ruderman 1986) and the Weight Efficacy Life-style Scale (WEL) (Clark et al 1991). Both scales are reliable and valid measures, and scores on these scales are highly correlated with one another (Clark et al 1991, Glynn & Ruderman 1986). Although both scales are relatively short (comprised of 25 and 20 items, respectively), we recommend the WEL for use by CorSolutions as it is slightly shorter. In completing this measure, respondents use a 10-point scale (0 = not confident to 9 = very confident) to rate their confidence about being able to resist the desire to eat in a variety of challenging situations (see below). A general weight management self-efficacy score is calculated by averaging across the 20 ratings.

Both the ESES and the WEL measure self-efficacy with regard to only one aspect of healthy eating - the ability to resist overeating. Because healthy eating is more than just controlling the amount of food consumed, we recommend that CorSolutions consider incorporating a few items from self-efficacy scales that pertain to other aspects of healthy eating. For example, Ling and Horwath (Ling, Horwath 1999) developed a scale to measure self-efficacy for consumption of fruits and vegetables, and Chang and colleagues (Chang et al. 2003) developed a scale to measure self-efficacy for eating low-fat diets.

Weight Efficacy Life-Style Questionnaire

Listed below are situations that lead some people to eat even when they are not hungry. We would like to know HOW CONFIDENT you are that you could resist eating in each situation. Please answer the following questions using the following ten-point scale.

0	1	2	3	4	5	6	7	8	9
Not									Very
confident									confident

Using the scale from 0-9, how confident are you that you can:

Subscale and item numbers

Negative Emotions

- 1. Resist eating when you are anxious (or nervous)?
- 6. Resist eating when you are depressed (or down)?
- 11. Resist eating when you are angry (or irritable)?
- 16. Resist eating when you have experienced failure?

Availability

- 2. Control your eating on the weekends?
- 7. Resist eating when there are many different kinds of foods available?
- 12. Resist eating even when you are at a party?

17. Resist eating even when high calorie foods are available?

Social Pressure

3. Resist eating even when you have to say "no" to others?
8. Resist eating even when you feel it's impolite to refuse a second helping?
13. Resist eating even when others are pressuring you to eat?
18. Resist eating even when you think others will be upset if you don't eat?

Physical Discomfort

4. Resist eating when you feel physically run down?
9. Resist eating even when you have a headache?
14. Resist eating when you are in pain?
19. Resist eating when you feel uncomfortable?

Positive Activities

5. Resist eating when you are watching TV?
10. Resist eating when you are reading?
15. Resist eating just before going to bed?
20. Resist eating when you are happy

Source: Clark, M.M., Abrams, D.B., Niaura, R.S., Eaton, C.A., & Rossi, J.S. (1991). Self-efficacy in weight management. *Journal of Consulting and Clinical Psychology, 59*, 739-744.

Self-Efficacy in Exercise and Physical Activity

Considerable evidence demonstrates the effects of self-efficacy on exercise behavior. Exercise self-efficacy is associated with higher levels of physical activity and frequency of exercise, greater effort exerted during exercise, and more positive feelings toward exercise (Dishman et al. 1985, Dzewaltowski et al. 1990, McAuley 1991, Petosa et al 2003, Rudoph & McAuley 1996, Yin & Boyd 2000). Several validated scales for measuring exercise self-efficacy exist (e.g., Hickey et al. 1992 and Sallis et al. 1988). We recommend that CorSolutions adopt a scale by Marcus, Selby, Niaura, and Rossi (Marcus et al. 1992), as it is the most commonly used measure of exercise self-efficacy and has demonstrated reliability and validity (Marcus et al. 1992, Buckworth et al. 2002, Marcus, Rakowski et al 1992). To complete this measure, participants use a 5-point scale to rate their confidence that they could exercise when other things get in the way (see below). The mean score for these items is calculated, and respondents are assigned a scale between 1 (not at all confident) and 5 (completely confident). Whereas the original scale developed by Marcus et al. consists 18 items, they also have validated a 6-item version of the scale, and we recommend that CorSolutions adopt this scale for measuring exercise self-efficacy.

Exercise Self-Efficacy

This scale measures how confident you are about your ability to exercise when other things get in the way. We would like to know HOW CONFIDENT you are that you could continue with your exercise plan in each situation. Please answer the following questions using the following five-point scale.

- 1 = Not at all confident
- 2 = Somewhat confident
- 3 = Moderately confident
- 4 = Very confident
- 5 = Completely confident

Subscale and items

Negative Affect

I am under a lot of stress.**

I am depressed.

I am anxious.

Excuse Making

I feel I don't have the time.**

I don't feel like it.

I am busy.

Must Exercise Alone

I am alone.

I have to exercise alone.**

My exercise partner decides not to exercise that day.

Inconvenient to Exercise

I don't have access to exercise equipment.**

I am traveling.

My gym is closed.

Resistance from Others

My friends don't want me to exercise.

My significant other does not want me to exercise.

I am spending time with friends or family who do not exercise.**

Bad Weather

It's raining or snowing.**

It's cold outside.

The roads or sidewalks are snowy.

Note: ** Items to be used for six-item self-efficacy short assessment.

Source: Marcus, B.H., Selby, V.C., Niaura, R.S., & Rossi, J.S. (1992). "Self-efficacy and the stages of exercise behavior change", *Research Quarterly for Exercise and Sport*, 63, 60-66.

Self-Efficacy in Treatment Adherence

Self-efficacy has been related to adherence to treatment regimens (Kaplan, Simon 1990). To our knowledge, there is not general measure of self-efficacy for treatment adherence. However, many scales exist to measure this construct in specific disease domains (e.g., Bogart et al. 2002, Catz et al. 2000, Kobau & DiIorio 2003, and Logan et al. 2003). Any of these disease-specific scales can be adapted to fit the diseases targeted by CorSolutions' interventions. We recommend the scale used by Catz, Bogart, and their colleagues because it is short, validated (Gifford et al. July 1996), and easily adapted. This scale measures patients' confidence in their ability to manage barriers to adherence and to tailor their medication regimens to fit with their daily lives. Respondents use a 10-point scale (1 = you think you cannot do it at all to 10 = you are certain that you can do it) to respond to each of the 8 items, and item responses are summed to yield a treatment adherence self-efficacy score.

Treatment Self-Efficacy

Listed below are situations that may make it difficult for you to stick with your treatment. We would like to know HOW CONFIDENT you are that you could stick with your treatment in each situation. Please answer the following questions using the following ten-point scale.

1	2	3	4	5	6	7	8	9	10
You cannot do it at all									You are certain you can do it

Using the scale from 1-10, how confident are you that you can:

1. Include your treatment in your daily routine?
2. Stick to your treatment plan even when side effects begin to interfere with daily activities?
3. Stick to your treatment schedule even when it means taking medications or doing other things in front of people who don't know you have [insert name of disease]?
4. Stick to your treatment schedule even when your daily routine changes?
5. Stick to your treatment schedule even when you're traveling?
6. Stick to your treatment schedule even when you aren't feeling well?
7. Stick to your treatment schedule even when it means changing your eating habits?

8. Stick exactly to your treatment schedule throughout the course throughout the course of your treatment?

9. Continue with your treatment even if doing so interferes with your daily activities?

10. Continue with your treatment even when you are feeling discouraged about your health?

11. Continue with your treatment even when people close to you tell you that they don't think it is doing any good?

Sources: Bogart, L. M., Gray-Bernhardt, M. L., Catz, S. L., Hartmann, B. R., & Otto-Salaj, L. L. (2002). Social and temporal comparisons made by individuals living with HIV disease: Relationships to adherence behavior. *Journal of Applied Social Psychology, 32*, 1551-1576.

Catz, S. L., Kelly, J. A., Bogart, L. M., Benotsch, E. G., & McAuliffe, T. (2000). Patterns, predictors, and barriers to medication adherence among persons prescribed new treatments for HIV disease. *Health Psychology, 19*, 124-133.

Gifford, A.L., Lorig, K., Chesney, M., Laurent, D., & Gonzalez, V. (1996, July). Patient education to improve health-related quality of life in HIV/AIDS: A pilot study. Paper presented at the 11th International Conference on AIDS (Abstract 1774). Vancouver, British Columbia, Canada.

B. CLINICAL MEASURES WITH GUIDELINES

Table B.1 Coronary Artery Disease

<u>Indicator</u>	<u>Data Source</u>	<u>CS</u>	<u>AH/JHU</u>	<u>QA Tools</u>	<u>Recommendation</u>	<u>Comment</u>	<u>Sources</u>
Prevention							
Proportion of participants who receive smoking cessation counseling	data collection	x	x	x	include		ACC/AHA 2002 (Class I)
Proportion of participants who receive screening for diabetes	claims		x		include		ACC/AHA 2002 (Class I)
Proportion of participants who receive a flu vaccination	data collection	x	x	x	include		DHHS 2005
Proportion of participants who receive pneumococcal vaccination	data collection	x	x	x	include		DHHS 2005
Proportion of participants who receive depression screening	data collection	x	x		include		ACC/AHA 2002 (Class IIb)
Diagnosis							
Proportion of participants who receive fasting lipid level	claims		x	x	include	Fasting lipid level is recommended as part of the initial workup of CAD patients.	ACC/AHA 2002 (Class I)

Proportion of participants who receive LDL screening	claims	x	x	x	include	AACE guidelines imply but do not explicitly recommend annual testing	AACE 2000
Proportion of participants who receive coronary angiograms for appropriate indications	data collection			x	drop	The decision to perform coronary angiography is not under the control of a disease management program	
Proportion of participants who receive LV function test after AMI	claims			x	include		ACC/AHA 2002 (Class I)
<i>Treatment</i>							
Proportion of participants with beta blocker usage	claims	x	x	x	include		ACC/AHA 2002 (Class I)
Proportion of participants in compliance with antiplatelet therapy	data collection	x	x	x	include	In theory, this indicator could be constructed from claims data. However, as aspirin as the most common antiplatelet drug is low-cost and OTC, few patients file a claim for it.	ACC/AHA 2002 (Class I)
Proportion of participants who receive lipid lowering therapy	claims	x	x	x	include	Rated class I for patients with LDL>130 and IIb for LDL between 100 and 129	ACC/AHA 2002 (Class I and IIa)
Proportion of participants who receive ACEI/ARB	claims		x	x	include	Rated class I for CAD patients with diabetes and/or LVSD and IIa for all patients with significant CAD and/or previous MI (ACC/AHA 2002)	ACC/AHA 2002 (Class I/IIa)
Proportion of participants who undergo elective revascularization for appropriate indications	data collection			x	drop	The decision to perform revascularization is not under the control of a Disease Management program	

Sources:

American Heart Association (AHA)/American College of Cardiology (ACC). Management of patients with chronic stable angina. 2002 guideline update.

Department of Health and Human Services (DHHS) Centers of Disease Control (CDC). Recommended adult immunization schedule. October 2004-September 2005.

American Association of Clinical Endocrinologists (AACE) medical guidelines for clinical practice for the diagnosis and treatment of dyslipidemia and prevention of atherogenesis 2000.

CAD ACC/AHA Classifications

Class I: Conditions for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective.

Class II: Conditions for which there is conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of a procedure or treatment.

Class IIa: Weight of evidence/opinion is in favor of usefulness/efficacy.

Class IIb: Usefulness/efficacy is less well established by evidence/opinion.

Class III: Conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective and in some cases may be harmful.

Table B.2 Congestive Heart Failure

<u>Indicator</u>	<u>Data Source</u>	<u>CS</u>	<u>AH/JHU</u>	<u>QA Tools</u>	<u>Recomm endation</u>	<u>Comment</u>	<u>American College of Cardiology Foundation and American Heart Association 2005 Guidelines</u>
<i>Prevention</i>							
Proportion of participants who receive flu vaccination	data collection	x	x	x	include		DHHS 2005
Proportion of participants who receive pneumococcal vaccination	data collection	x	x	x	include		DHHS 2005
Proportion of participants with atrial fibrillation and/or prior thromboembolic event who receive warfarin	claims		x	x	include	ACC/AHA 2005	ACC/AHA 2005 Class I (A)

Proportion of all participants who receive warfarin	claims		x	x	include	ACC/AHA 2005 guideline supports use of warfarin in all symptomatic CHF patients, but with weaker evidence (class III).	ACC/AHA 2005 Class III
Proportion of participants who receive depression screening	data collection	x	x		include		USPSTF 2005 Level B
Proportion of participants screened for depression and referred for follow up if at risk	data collection	x	x		include		USPSTF 2005 Level B
Treatment							
Proportion of participants with beta blocker usage	claims	x	x		include	Strictly speaking, beta-blockers are only indicated for systolic dysfunction, but identifying those patients is not possible based on claims data. One may need to add a caveat to account for this problem.	ACC/AHA 2005 Class I (A)
Proportion of participants with vasodilator usage	claims	x	x	x	include	Strictly speaking, vasodilators are only indicated for systolic dysfunction, but identifying those patients is not possible based on claims data. One may need to add a caveat to account for this problem.	ACC/AHA 2005 Class IIa (A)
Proportion of participants receiving spironolactone for severe CHF	data collection		x		drop	Identifying the subset of patients who have an indication for spironolactone treatment requires complex decision rules and is very data-intense, rendering this indicator impractical.	ACC/AHA 2005 Class I (B)
Proportion of participants with LV EF measurement	claims		x	x	include	Strictly speaking EF measurement is only required at the time of the initial diagnosis. One could look at longer time periods and restrict	ACC/AHA 2005 Class I (A)

						the denominator to patients with at least two years of claims data.	
Proportion of participants with documentation of comorbidities and other cardiac risk factors in the record	data collection			x	drop	Those include AMI, angina, other cardiac disorders, hypertension and diabetes. This process, however, refers to medical record keeping and does not seem applicable to disease management programs.	
Proportion of participants on ARB/ACEI who receive annual creatinine checks	claims			x	include		ACC/AHA 2005 Class I (B)
Proportion of participants on ARB/ACEI who receive annual potassium checks	claims			x	include		ACC/AHA 2005 Class I (B)

Sources:

American Heart Association (AHA)/American College of Cardiology (ACC). Diagnosis and management of chronic heart failure in the adult. 2005 guideline update.

Department of Health and Human Services (DHHS) Centers of Disease Control (CDC). Recommended adult immunization schedule. October 2004-September 2005.

U.S. Preventive Services Task Force (USPSTF). The guide to clinical preventive services 2005

Heart Failure

Class I Recommendations: Conditions for which there is evidence and/or general agreement that a given procedure/therapy is beneficial, useful, and/or effective.

Class II Recommendations: Conditions for which there is conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of a procedure or treatment.

Class IIa Recommendations: Weight of evidence/opinion is in favor of usefulness/efficacy.

Class IIb Recommendations: Usefulness/efficacy is less well established by evidence/opinion.

Class III Recommendations: Conditions for which there is evidence and/or general agreement that a given procedure/therapy is not useful/effective and in some cases may be harmful.

Level of Evidence A: Data derived from multiple randomized clinical trials or meta-analyses.
 Level of Evidence B: Data derived from a single randomized trial, or nonrandomized studies.
 Level of Evidence C: Only consensus opinion of experts, case studies, or standard-of-care.
 USPSTF recommendations: Level A (strongly recommended), Level B (recommended)

Table B.3 Hypertension

<u>Indicator</u>	<u>Data Source</u>	<u>CS</u>	<u>AH/JHU</u>	<u>QA Tools</u>	<u>Recommendation</u>	<u>Comments</u>	<u>NIH 2003 Guidelines</u>
Prevention							
Proportion of participants who receive depression screening	data collection	x	N/A		include	Routine depression screening for hypertensive patients may not be universally recommended but many hypertensives have relevant comorbidities.	USPSTF 2005 Level B
Diagnosis							
Proportion of participants with documentation of comorbidities and other cardiac risk factors	data collection		N/A	x	drop	Those include AMI, angina, other cardiac disorders, hypertension and diabetes. This process, however, refers to medical record keeping and does not seem applicable to disease management programs.	

Proportion of participants who receive lab testing for initially diagnosed hypertension	claims		N/A	x	drop	The tests are serum potassium, glucose, creatinine, tryglicerides, cholesterol and urinalysis. Strictly speaking, they are indicator for initial evaluation of hypertension and it is not clear that we can capture those patients in the disease management context.	
Treatment							
Proportion of participants with consistent average SBP>140 or DBP>90 over 6 months who receive one of the following interventions: change in dose or regimen of antihypertensives, or repeated education regarding lifestyle modifications.	data collection		N/A	x	drop	Not relevant for disease management program that mainly deals with patients with established disease	
Proportion of participants with stage 1B, 2, 3 hypertension on pharmacotherapy.	data collection		N/A	x	drop	Not relevant for disease management program that mainly deals with patients with established disease	

Sources:
 U.S. Preventive Services Task Force (USPSTF). The guide to clinical preventive services 2005

**Table B.4 Chronic
 Obstructive Pulmonary
 Disease**

<u>Indicator</u>	<u>Data Source</u>	<u>CS</u>	<u>AH/JHU</u>	<u>QA Tools</u>	<u>Recomm</u> <u>endation</u>	<u>Comment</u>	<u>Physicians Consortium</u> <u>2004 Guidelines</u>
------------------	--------------------	-----------	---------------	-----------------	----------------------------------	----------------	--

Prevention						
Proportion of participants who receive flu vaccination	data collection	x	x	x	include	AAFP 2004/ PCPI 2004
Proportion of participants who receive pneumococcal vaccination	data collection	x	x	x	include	AAFP 2004/DHHS 2005
Proportion of participants who receive depression screening	data collection	x	x		include	USPSTF 2005 Level B
Diagnosis						
Proportion of participants who receive spirometry testing	claims		x	x	include	May be underreported as test performed with hand held devices in physician offices are not billed on an itemized basis but covered under Evaluation and Management codes. PCPI 2004
Proportion of participants who receive theophylline checks after initiation of treatment or dosage increase	claims			x	drop	While "close monitoring" of patients under theophylline treatment is recommended it is very difficult to detect dosage changes in claims data. Further, the sample size is likely to be small. AAFP 2004
Treatment						
Proportion of participants on bronchodilator	claims	x			include	While bronchodilators are recommended for all symptomatic stages of COPD, the type of treatment varies by stage. The indicator could be modified to capture the differential recommendations. Staging information would require data collection. ATS 2004
Proportion of participants with steroid inhaler use	claims	x			include	There is still some controversy over the use of steroid inhalers in COPD. It is usually only recommended after failure of bronchodilator treatment. AAFP 2004
Proportion of participants who receive oxygen therapy O2 saturation in below 88% at rest	data collection		x	x	include	This indicator requires information on O2 saturation or pO2 levels. Neither the guidelines nor the literature provide for a severity measure that could be scored from claims or self reported data. AAFP 2004/ATS 2004
Proportion of participants who receive oral steroids for exacerbation	claims		x		drop	This intervention is not necessarily under disease management control.

Proportion of participants on bronchodilators who receive ipratropium	claims		x	include		ATS 2004
Proportion of participants on inhalers who receive spacer use or proper MDI instructions	data collection		x	include	Strictly speaking, ATS recommends the use of MDI/DPI. Proper instructions to the patient are implied.	ATS 2004

Sources:

U.S. Preventive Services Task Force (USPSTF). The guide to clinical preventive services 2005
 Physician Consortium for Performance Improvement (PCPI) COPD Core Physician Performance Measurement Set 2005
 American Thoracic Society (ATS). Standards for the diagnosis and management of patients with COPD. 2004
 American Academy of Family Physicians (AAFP). Hunter HM and King DE. COPD: Management of acute exacerbations and chronic stable disease. American Family Physician 2001; 64: 603-622

Table B.5 Asthma

<u>Indicator</u>	<u>Data Source</u>	<u>CS</u>	<u>AH/JHU</u>	<u>QA Tools</u>	<u>Recomm endation</u>	<u>Comments</u>	<u>National Asthma Education and Prevention Program 2002 Guidelines</u>
<i>Prevention</i>							
Proportion of participants who receive flu vaccination	data collection	x	x	x	include		DHHS 2005

Proportion of participants who receive pneumococcal vaccination	data collection	x	x	x	include		DHHS 2005
Proportion of participants who receive depression screening	data collection	x	x		include		USPSTF 2005 Level B
Diagnosis							
Proportion of participants who receive spirometry testing	claims		x	x	include	NAEPP recommends repeat testing every 1-2 years and the use of peak-flow based treatment plans for patients with moderate to severe asthma.	NAEPP 2002 (Evidence D for routine use every 1-2 years and B for treatment plan use)
Proportion of participants on theophylline with a daily dose of >= 600mg who receive routine theophylline level checks	claims			x	include		NAEPP 2002
Proportion of participants who receive PEV or FEV1 in exacerbation	claims			x	drop	Treatment provide during acute exacerbation is not under the control of disease management	
Proportion of participants who receive theophylline level checks in exacerbation	claims			x	drop	Treatment provide during acute exacerbation is not under the control of disease management	
Treatment							
Proportion of participants with moderate to severe asthma on beta agonists or anticholinergics	claims	x			include		NAEPP 2002 (Evidence A)
Proportion of participants who receive inhalable steroids for uncontrolled asthma	data collection		x		include	The AH definition of uncontrolled asthma requires FEV1 values. The indicator could be scored for the full population, using the claims-based HEDIS criteria for moderate to severe asthma.	NAEPP 2002 (Evidence B)
Proportion of participants who receive appropriate use of long-term control medication	claims		x	x	include	Inhaled steroids are seen as first choice	NAEPP 2002 (Evidence A)
Proportion of participants who receive prescription of rescue	claims			x	include	Depending on usage, a rescue inhaler might last longer than a year.	NAEPP 2002 (page 76)

inhaler							
Proportion of participants with moderate to severe asthma in compliance with contraindication to beta-blockers	claims			x	include		Commonly accepted
Proportion of participants who receive substitution of inhaled steroids for oral ones	claims			x	drop	This indicator is conceptually appealing but hard to operationalize.	
Proportion of participants who receive proper instructions of MDI use or spacer	data collection			x	include		NAEPP 2002 (Evidence B and C)

Sources:

U.S. Preventive Services Task Force (USPSTF). The guide to clinical preventive services 2005

Department of Health and Human Services (DHHS) Centers of Disease Control (CDC). Recommended adult immunization schedule. October 2004-September 2005.

National Asthma Education and Prevention Program (NAEPP) Guidelines for the diagnosis and management of asthma. 2002 update.

Asthma

Evidence Category A: Randomized controlled trials (RCTs), rich body of data.

Evidence Category B: RCTs, limited body of data.

Evidence Category C: Nonrandomized trials and observational studies.

Evidence Category D: Panel consensus judgment.

Table B.6 Diabetes mellitus

<u>Indicator</u>	<u>Data Source</u>	<u>CS</u>	<u>AH/JHU</u>	<u>QA Tools</u>	<u>Recomm endation</u>	<u>Comment</u>	<u>American Diabetes Association Recommendations</u>
Prevention							
Proportion of participants receiving lipid testing	claims		x	x	include		ADA 2005 (E)
Proportion of participants having annual foot exam by physician	data collection	x	x	x	include		ADA 2005 (B)
Proportion of participants who receive flu vaccination	data collection	x	x		include		ADA 2005 (C)
Proportion of participants who receive pneumococcal vaccination	data collection	x	x		include		ADA 2005 (C)
Proportion of participants who receive ASA prophylaxis	data collection		x		include	ASA prophylaxis is now universally recommended, unless contraindicated, for diabetes older than 21 (ADA 2005)	ADA 2005 (A)
Proportion of participants who receive depression screening	data collection	x	x		include		ADA 2005 (E)
Diagnosis							
Proportion of participants having dilated eye exams annually	claims	x	x	x	include		ADA 2005 (B)
Proportion of participants having microalbumin testing	claims	x	x	x	include		ADA 2005 (E)
Proportion of participants receiving biannual HgbA 1c testing	claims	x	x	x	include		ADA 2005 (E)
Treatment							

Proportion of participants who receive ACEI/ARB for albuminuria	claims	x	x	include		ADA 2005 (A)
Proportion of Type 2 diabetics who receive oral hypoglycemic therapy after having failed dietary therapy	data collection		x	drop	The definition should also include newer, non-hypoglycemic drugs. The indicator may be hard to operationalize.	
Proportion of Type 2 diabetics who receive insulin after having failed oral therapy	data collection		x	drop	The indicator may be hard to operationalize.	

Sources:

American Diabetes Association (ADA). Standards of medical care in diabetes. 2005

Diabetes ADA

Evidence Category A: Clear evidence from well-conducted, generalizable, randomized controlled trials that are adequately powered.

Supportive evidence from well-conducted randomized controlled trials that are adequately powered.

Evidence Category B: Supportive evidence from well-conducted cohort studies. Supportive evidence from a well-conducted case-control study.

Evidence Category C: Supportive evidence from poorly controlled or uncontrolled studies. Conflicting evidence with the weight of evidence supporting the recommendation.

Evidence Category E: Expert consensus or clinical experience.

C. ESTIMATION PROCEDURE FOR RISK ADJUSTMENT

This method adjusts the estimates for changes in cost and utilization under disease management program as the difference between observed and predicted cost in the intervention period. The prediction is based on a statistical model, which incorporates patients' demographic characteristics and diagnostic variables. As the first step, the cost or utilization parameter of interest is regressed on vectors of patient level diagnostic and demographic variables using pre-intervention data (from time t).

$$y_t = \alpha_t + X_t \beta_t + \theta_t \gamma_t + \varepsilon_t \quad (\text{Equation 1})$$

The estimated coefficients from Equation 1 are saved and then applied to data from the intervention period (time $t+1$) to derive predicted or estimated spending/utilization:

$$y_{t+1} = \alpha_{t+1} + X_{t+1} \beta_t + \theta_{t+1} \gamma_t \quad (\text{Equation 2})$$

This method allows for adjustments to be made for changes in demographic mix, due to such factors as aging and movement in and out of the intervention group, and disease severity over time. It is also operationally appealing because many disease management firms already use commercial software for risk stratification that can be used to derive the variables needed for the model from administrative data. It cannot, however, account for factors that influence cost and utilization, which change over time, such as technology change, changes in benefits and change in contractual arrangements between purchasers

An important operational issue is how to construct the diagnostic variables to capture patient risk. Given the complexity of designing algorithms to extract such information from claims data, we would recommend using commercially available grouping software. A recent study compared the predictive accuracy of the models based on several such groupers and found that they can predict about 15-25% of future medical cost (Cumming et al. 2002). The authors found that incorporating more

information in a grouper improves model performance, i.e. groupers that combine medical and pharmacy claims have the greatest predictive power.

Another issue is excluding patients with extremely high cost from the analysis based on the rationale that disease management cannot influence the high cost of some conditions, like organ transplantation or severe burns. Also, the predictive models perform much worse for such outlier events. It is common practice in the industry to exclude patients with certain high-cost conditions, such as organ transplantation and chronic dialysis, from the analysis and to truncate or exclude claims exceeding a certain threshold, usually above \$100,000 to \$300,000. We propose to assess the effect of such restrictions empirically and make a recommendation based on the findings.

D. SURVEY INSTRUMENTS FOR HEALTH-RELATED PRODUCTIVITY

Table D.1 Summary of Worker Productivity Measurement Instruments*

Instrument	Year	Productivity Metric	Construct Validity	Internal consistency reliability	Test-retest reliability	Responsiveness	Administrator/respondent burden	Generalisability	Applied for economic valuation	Tested Diseases	References
American Productivity Audit	2001	B	Unk	Unk	Unk	Unk	Moderate	Yes	Yes	O	J, K
Angina-Related Limitations at Work Questionnaire	1998	B	Established	Established	Established	Unk	Low/Moderate	NA	No	O	F
Endicott Work Productivity Scale	1997	B	Established	Established	Established	Unk	Low	Yes	No	O	F, H
Health and Labor Questionnaire	1995	B	Unk	Unk	Unk	Unk	Low/Moderate	Yes	Yes	O	F, H
Health and Productivity Questionnaire (HPQ) **	2003	B	Unk	Established	Established	Unk	Low/Moderate	Yes	Yes	O	L
Health and Work Questionnaire (HWQ)	2001	B	Established	Established	Unk	Unk	Low/Moderate	Yes	Unk	O	F, H, I
Health-Related Productivity Questionnaire Diary	2003	B	Established	Unk	Unk	Unk	High	Yes	No	O	C
Migraine Disability Assessment Questionnaire	1999	B	Established	Established	Established	Unk	Low	NA	Yes	O	H
Migraine Work and Productivity Loss Quest.	1999	B	Established	Established	Unk	Unk	Moderate	NA	Yes	O	F, H
Osterhaus Technique	1992	B	Unk	Unk	Unk	Unk	Low	Unk	Yes	O	D, F
Stanford Presenteeism Scale	2002	P	Established	Established	UNK	Unk	Low	High	No	Unk	F, M
Unnamed Hepatitis Instrument	2001	B	Unk	Unk	Unk	Unk	Unk	Unk	Unk	O	F
Work Limitations Questionnaire	2001	B	Established	Established	Established	Unk	Low	Yes	Yes	A, O	A, F, G, H, E

Work Productivity and Activity Impairment Questionnaire (WPAI) - General	1993	B	Established	NA	Established	Unk	Low	High	Yes	A, CHF, H, D, O	B, F, H
WPAI-Allergic Rhinitis	1993	B	Established	NA	Established	Established	Low	High		O	H
WPAI-Chronic Hand Dermatitis	1993	B	Established	NA	Established	Established	Low	High	Yes	O	H
WPAI-Gastro-Esophageal Reflux Disease	1993	B	Established	NA	Unk	Unk	Low	High	Yes	O	H
WPAI-Specific Health Problem	1993	B	Established	NA	Established	Unk	Low	High	Yes	G	H
Worker Productivity Index	1999	B	Unk	NA	Unk	Unk	Unk	Unk	Unk	D, H, O	F
Work Productivity Short Inventory (WPSI)	2003	B	Established	Unk	Unk	Unk	Unk	High	Yes	O	G, N

* This report has been adapted and expanded from similar tables located in references F and H.

** This instrument was previously referred to as the Health and Work Performance Questionnaire.

Productivity Metric

Condition List

Absenteeism	A	Asthma	A
Presenteeism	P	Chronic Obstructive Pulmonary Disease	COPD
Both	B	Congestive Heart Failure	CHF
		Coronary Artery Disease	CAD
Insufficient Info. Available	UNK	Diabetes	D
Not Applicable	NA	Hypertension	H
		General Health	G
		Other	O

References

- A Burton, WN et al. The association of medical conditions and presenteeism. J Occup Environ Medicine, 2004 Jun; 46(6): Suppl: S36-45.
- B Burton, WN et al. The role of health risk factors and disease on worker productivity. J Occup Environ Med. 1999 Oct; 41(10): 863-77.
- C Kumar, RN et al. Validation of the HRPQ-D on a sample of patients with infectious mononucleosis. J Occup Environ Med, 2003 Aug; 45(8): 899-907.
- D Lavigne JE et al. Reduction in individual work productivity associated with type 2 diabetes mellitus. Pharmacoeconomics. 2003; 21(15): 1123-34.
- E Lerner, D et al. The clinical and occupational correlates of work productivity loss among employed patients with depression. J Occupational and Environmental Medicine. 2004 Jun; 46(6) Suppl: S46-55.
- M Loepke, Ronald, et al. Health-Related Workplace Productivity Measurement. JOEM, April 2003; 45(4): 349-359.
- F Lofland, Jennifer, et al. A Review of Health Related Workplace Productivity Loss Instruments. Pharmacoeconomics, 2004; 22(3): 165-184.
- G Ozminkowski, RJ. Et al. The application of two health and productivity instruments at a large employer. J Occup Environ Medicine, 2004 Jul; 46(7): 635-48.
- H Prasad, Manishi, et al. A Review of Self-Report Instruments Measuring Health-Related Work Productivity. Pharmacoeconomics, 2004; 22(4): 225-244.
- I Shikiar R et al. Development of the Health and Work Questionnaire. Work, 2004; 22(3): 219-29.
- J Stewart WF et al. Cost of lost productive work among US workers with depression. JAMA. 2003 Jun 18; 289(23): 3135-44.

K Stewart WF et al. Lost Productive work time costs from health conditions in the US. J Occup Environ Med. 2003 Dec; 45(12): 1234-46.

L Wang PS et al. Chronic medical conditions and work performance in the health and work performance questionnaire calibration surveys. J Occup Environ Med. 2003 Dec; 45(12): 1303-11.

N Goetzel R et al. Development and Reliability Analysis of the Work Productivity Short Inventory (WPSI) Instrument Measuring Employee Health and Productivity. JOEM. July 2003; 45(7): 743-753.

Table D.2 Detailed Properties of Worker Productivity Instruments

Instrument	HPQ	HWQ	SPS	WLQ	WPAI
Performance Scales? (Yes/No)	Yes	Yes	Yes	Yes	Yes
Scale Gradation	1-5/7/10	1-10	1-5	5 point	1-10
Scale Anchoring? (Yes/No)	Yes	Yes	No	No	No
Time Frame (weeks)	1 and 4	1	4	2	1
Estimates of Time Lost? (Yes/No)	Yes	No	No	Yes	Yes
Separation of Time Lost (Due To Vacation and Health)?	Yes	N/A	N/A	No	Yes
Estimation of Time Worked?	Yes	N/A	N/A	Yes	Yes
Time Units	Days and hours	N/A	N/A	Percentage	Hours
Monetary Conversion Possible? (Yes/No)	Yes	No	No	Yes	Yes
Questions on Type of Employment? (Yes/No)	Yes	No	No	NS	No
System for Monetary Conversion	NS	N/A	N/A	NS	NS
Questions on Salary? (Yes/No)	Yes	No	No	Yes	No
Demographic Questions?	Yes	No	No	Yes	No
Questions on Medical Conditions?	Yes (employer version)	No	No	No	No
Questions on Treatment?	Yes (employer version)	No	No	No	No
Number of Questions	31 (employer version)	24	6	25	6
Sample Available?	Yes	Yes	Yes	Yes	Yes
Fee for Use *	Yes	Unk	Yes	Yes	No

*Ozminkowski, RJ et al. The Application of Two Health and Productivity Instruments at a Large Employer. JOEM. 46(7): 635-648.

Table D.3 Content of Worker Productivity Instruments

Instrument	Type of Questions
Endicott	Assessing frequency of productivity behavior
HPQ	Questions on overall health, effect of conditions, frequency of depressive feelings, frequency of low/high performance relative to other workers, amount of insufficient quality, concentration and hindered work due to health, job performance of most workers, respondent's usual performance, and performance in the 7- or 28-day time period
HWQ	Questions on relationship to job, work environment, co-workers, family, and friends; Questions on efficiency, quality and amount of work as rated by self, supervisor, and co-workers; Questions on frequency of concentration, impatience, exhaustion, etc.
SPS	Questions on finishing tasks, handling stress, achieving goals, having energy, etc. due to health problem
WLQ	Questions on time, physical, mental, interpersonal and output demands
WPAI	Questions on productivity and regular activities: health problems had no effect/completely prevented working or daily activities

BIBLIOGRAPHY

Standard Outcome Metrics and Evaluation Methodology for Disease Management Programs, 2nd Disease Management Outcomes Summit presented by American Healthways and Johns Hopkins, November 7-10, 2002, Palm Desert, CA.

The competitive edge: Employee Health and Productivity, Employers Health Coalition, Tampa FL, 1999.

Allen, Harris M., William B. Bunn III. (2003a). Validating Self-Reported Measures of Productivity at Work: A Case for Their Credibility in a Heavy Manufacturing Setting. *Journal of Occupational and Environmental Medicine*, 45, 926-940.

Allen, Harris M., William B. Bunn III. (2003b). Using Self-Report and Adverse Event Measures to Track Health's Impact on Productivity in Known Groups. *Journal of Occupational and Environmental Medicine*, 45, 973-983.

Baer, J. S., & Lichtenstein, E. (1988). "Classification and prediction of smoking relapse episodes: An exploration of individual differences", *Journal of Consulting and Clinical Psychology*, 56, 104-110.

Baer, J. S., Holt, C. S., & Lichtenstein, E. (1986). "Self-efficacy and smoking reexamined: Construct validity and clinical utility", *Journal of Consulting and Clinical Psychology*, 54, 846-852.

Bandura, A. *Self-efficacy: The exercise of control*. New York: Freeman, 1997.

Bandura, A. "Self-efficacy mechanism in physiological activation and health-promoting behavior", in Neurobiology of learning, emotion and affect, J. Madden (Ed.), New York: Raven Press, 1991, p. 229-269.

Bandura, A. Social foundations of thought and action: A social cognitive theory, Englewood Cliffs, NJ: Prentice Hall, 1986.

Bandura, A. "Self-efficacy: Toward a unifying theory of behavior change", *Psychological Review*, Vol. 84, 1977, p. 191-215.

Berger, Marc L., James F. Murray, Judy Xu, Mark Pauly, "Alternative Valuations of Work Loss and Productivity," *JOEM*, January 2001.

Bernier, M., & Avard, J. (1986). Self-efficacy, outcome, and attrition in a weight-reduction program. *Cognitive Therapy and Research*, 10, 319-338.

Bodenheimer T. "Disease management in the American market", *BMJ*, Vol. 320, 2000, p. 563-6.

Bodenheimer T. 1999. Disease management- promises and pitfalls. *New England Journal of Medicine*; 340: 1202-1205

Bogart, L. M., Gray-Bernhardt, M. L., Catz, S. L., Hartmann, B. R., & Otto-Salaj, L. L. (2002). Social and temporal comparisons made by individuals living with HIV disease: Relationships to adherence behavior. *Journal of Applied Social Psychology*, 32, 1551-1576.

Buckworth, J., Granollo, D. H., & Belmore, J. (2002). Incorporating personality assessment into counseling to help college students adopt and maintain exercise behaviors. *Journal of College Counseling*, 5, 15-25.

Burns H. 1996. Disease management and the drug industry: carve out or carve up? *Lancet*; 347: 1021-1023

Burton Wayne N., Glenn Pransky, Daniel J. Conti, Chin-Yu Chen, Dee W. Edington. (2004). The Association of Medical Conditions and Presenteeism. *Journal of Occupational and Environmental Medicine*, 46. S38-S45.

Callahan, CM. "Quality improvement research on late life depression in primary care", *Medical Care*, Vol. 39, 2001, p. 772-84.

Catz, S. L., Kelly, J. A., Bogart, L. M., Benotsch, E. G., & McAuliffe, T. L. (2000). Patterns, correlates, and barriers to medication adherence among persons prescribed new treatments for HIV disease. *Health Psychology*, 19, 124-133.

Celli BR, 2000. "The Importance of Spirometry in COPD and Asthma. Effect on Approach to Management." *Chest*, Vol. 117, p. 15S-19S

Centers for Medicare and Medicaid Services, *National Health Expenditures and Selected Economic Indicators, Levels and Average Annual Percent Change: Selected Calendar Years 1980-2011*, 2002, Baltimore.

Chassin, M.R., "Assessing strategies for quality improvement", *Health Affairs*, 1997, Vol. 16, No. 3, p. 151-61.

Chang, M., Nitzke, S., Brown, R. L., Baumann, L. C., & Oakley, L. (2003). Development and validation of a self-efficacy measure for fat intake behaviors of low-income women. *Journal of Nutrition Education & Behavior* 35, 302-307.

Clark, M. M., Abrams, D. B., Niaura, R. S., Eaton, C. A., & Rossi, J. S. (1991). Self-efficacy in weight management. *Journal of Consulting and Clinical Psychology*, 59, 739-744.

Clark, C.M., J.E. Fradkin, R.G. Hiss, R.A. Lorenz, F. Vinicor, and E. Warren-Boulton, "Promoting early diagnosis and treatment of type 2

diabetes: The National Diabetes Education Program", *JAMA*, 2000, Vol. 284, No. 3, p. 363-5.

Colletti, G., Supnick, J. A., & Payne, T. J. (1985). "The Smoking Self-Efficacy Questionnaire (SSEQ): Preliminary scale development and validation", *Behavioral Assessment*, 7, 249-260.

Condiotte, M. M., & Lichtenstein, E. (1981). "Self-efficacy and relapse in smoking cessation programs", *Journal of Consulting and Clinical Psychology*, Vol. 49, p. 648-658.

Dzewaltowski, D. A., Noble, J. M., & Shaw, J. M. (1990). Physical activity participation: Social cognitive theory versus the theories of reasoned action and planned behavior. *Journal of Sport and Exercise Psychology*, 12, 388-405.

DiClemente, C. C., Prochaska, J. O., & Gilbertini, M. (1985). Self-efficacy and the stages of self-change of smoking. *Cognitive Therapy and Research*, 9, 181-200.

DiClemente, C. C. (1981). Self-efficacy and smoking cessation maintenance: A preliminary report. *Cognitive Therapy and Research*, 5, 175-187.

Dishman, R. K., Sallis, J., & Orenstein, D. (1985). The determinants of physical activity and exercise. *Public Health Reports* 100, 158-171.

Employers Health Coalition. *The Hidden Competitive Edge: Employee Health and Productivity*. Newton, MA: Managed Care Communications Inc., 2000.

Enright PL, 2003. "The Six-Minute Walk Test." *Respiratory Care*, Vol. 48, No. 8, p. 783-785.

Foreyt, J. P., Bruner, R. L., Goodrick, G. K., & Cutter, G. (1995). Psychological correlates of weight fluctuation. *International Journal of Eating Disorders*, 17, 263-275.

Glynn, S. M., & Ruderman, A. J. (1986). "The development and validation of an eating self-efficacy scale", *Cognitive Therapy & Research*, Vol. 10, p. 403-420.

Godin, G., Valois, P., LePage, L., & Desharnais, R. (1992). "Predictors of smoking behavior: An application of Ajzen's theory of planned behavior", *British Journal of Addiction*, 87, 1335-1343.

Goetzel R.Z., Stacey R. Long, Ronald J. Ozminkowski, Kevin Hawkins, Shaohung Wang, Wendy Lynch. (2004a). Health, Absence, Disability, and Presenteeism Cost Estimates of Certain Physical and Mental Health Conditions Affecting U.S. Employers. *Journal of Occupational and Environmental Medicine*, 46(4), 1-15.

Goetzel RZ, Long SR, Ozminkowski RJ, Chang S. (2004b) "The application of two health and productivity", *Journal of Occupational and Environmental Medicine*, Vol. 46, Iss. 7, July 2004, p. 635-648.

Gulliver, S. B., Hughes, J. R., Solomon, L. J., & Dey, A. N. "Self-efficacy and relapse to smoking in self-quitters", *Addiction*, Vol. 90, 1995, p. 767-772.

Hemp, P. (2004). Presenteeism: At Work-But Out of It. *Harvard Business Review*. 82(10): 49-58.

Hickey, M. L., Owen, S. V., & Froman, R. D. (1992). Instrument development: Cardiac diet and exercise self-efficacy. *Nursing Research*, 41, 347-351.

Hofstetter, C. R., Sallis, J. F., & Hovell, M. F. (1990). Some health dimensions of self-efficacy: Analysis of theoretical specificity. *Social Science and Medicine*, 31, 1051-1056.

Hoppe MJ, Graham L, Wilsdon A, Wells EA, Nahom D, Morrison DM, 2004. "Teens Speak Out About HIV/AIDS: Focus Group Discussions about Risk and Decision-Making" *J Adolescent Health*, Vol. 35, p.345.e27-345.e35

IOM (J. M. Corrigan, A. Greiner, and S. M. Erickson, eds., 2003a), *Fostering Rapid Advances in Health Care: Learning from System Demonstrations.*, Washington, D.C.: National Academy Press.

IOM, *Priority Areas for National Action: Transforming Health Care Quality*, K. Adams and J. M. Corrigan, eds., 2003b, Washington, D.C.: National Academy Press.

Johannesson, Magnus and Goran Karlsson. (1997). The friction cost method: A comment. *Journal of Health Economics*, 16(2), 249-255.

Kaplan, R. M. & Simon, H. J. (1990). Compliance in medical care: Reconsideration of self-predictions. *Annals of Behavioral Medicine*, 12, 66-71.

Kessler RC, Ames M, Hymel PA, Loeppke R, McKenas DK, Richling DE, Stang PE, Ustun TB. "Using the World Health Organization Health and Work Performance Questionnaire (HPQ) to evaluate the indirect workplace costs of illness", *J Occup Environ Med (Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine.)* Jun 2004, Vol. 46, Iss. 6, Suppl: S23-37.

Kitsantas, A. (2000). The role of self-regulation strategies and self-efficacy perceptions in successful weight loss maintenance. *Psychology and Health*, 15, 811-820.

Kessler, Ronald C., Minnie Ames, Pamela A. Hymel, Ronald Loeppke, David K. McKenas, Dennis E. Richling, Paul E. Stang, T. Bedirhan Ustun. (2004). Using the World Health Organization Health and Work Performance Questionnaire (HPQ) to Evaluate the Indirect Workplace Costs of Illness. *Journal of Occupational and Environmental Medicine*, 46(6), S23-S37.

Kessler, Ronald, telephone communication with the researcher, February 8, 2005.

Kobau, R., & DiIorio, C. (2003). Epilepsy self-management: A comparison of self-efficacy and outcome expectancy for medication adherence and lifestyle behaviors among people with epilepsy. *Epilepsy & Behavior*, 4, 217-225.

Koopmanschap, M. A., F. F. H. Rutten, B. M. van Ineveld et al. (1995). The friction cost method of measuring the indirect costs of disease. *Journal of Health Economics*, 14, 171-89.

Legorreta, A.P., X. Liu, C.A. Zaher, and D.E. Jatulis, "Variation in managing asthma: Experience at the medical group level in California", *American Journal of Managed Care*, 2000, Vol. 6, No. 4, p. 445-53.

Lerner, Debra, Benjamin C. Amick III, William H. Rogers, Susan Malspeis, Kathleen Bungay, Diane Cynn. (2001). The Work Limitations Questionnaire. *Medical Care*, 39(1), 72-88.

Lewis, A, "Return on Investment and Savings Methodology: This is Our Final Answer", Disease Management Purchasing Consortium, 2003.

Linden A, Adams JL, Roberts N, 2003. "An Assessment of the Total Population Approach for Evaluating Disease Management Program Effectiveness" *Disease Management*, Vol. 6, p. 93-102

Ling, A. M. C., & Horwath, C. (1999). Self-efficacy and consumption of fruits and vegetables: Validation of a summated scale. *American Journal of Health Promotion*, 13, 290-298.

Loeppke, R., Pamela A. Hymel, Jennifer H. Lofland, Laura T. Pizzi, Doris L. Konicki, George W. Anstadt, Catherine Baase, Joseph Fortuna, and Ted Scharf. (2003). Health-Related Workplace Productivity Measurement: General and Migraine-Specific Recommendations from the ACOEM Expert Panel. *Journal of Occupational and Environmental Medicine*, 45(4), 349-359.

Logan, D., Zelikovsky, N., Labay, L., & Spergel, J. (2003). The illness management survey: Identifying adolescents' perceptions of barriers of adherence. *Journal of Pediatric Psychology*, 28, 383-392.

Marcus, B. H., Rakowski, W., & Rossi, J. S. (1992). Assessing motivation readiness and decision making for exercise. *Health Psychology*, 11, 257-261.

Marcus, B. H., Selby, V. C., Niaura, R. S., and Rossi, J. S. (1992). Self-efficacy and the stages of exercise behavior change. *Research Quarterly for Exercise and Sport*, 63, 60-66.

McAlister FA, Lawson FM, Teo KK, Armstrong PW, 2001. A systematic review of randomized trials of disease management programs in heart failure. *American Journal of Medicine*;110:378-84.

McAuley, E. (1991). Efficacy, attributional, and affective responses to exercise participation. *Journal of Sport and Exercise Psychology*, 13, 382-393.

McBride, P., H.G. Schrott, M.B. Plane, G. Underbakke, and R.L. Brown. 1998. *Primary care practice adherence to National Cholesterol Education Program guidelines for patients with coronary heart disease*. *Archives of Internal Medicine* 158 (11): 1238-44.

McCulloch DK, Price MJ, Hindmarsh M, Wagner EH, 2000. "Improvement in diabetes care using an integrated population-based approach in a primary care setting", *Disease Management*; 3:73-80.

McCulloch DK, Price MJ, Hindmarsh M, Wagner EH, 1998. A population-based approach to diabetes management in a primary care setting: early results and lessons learned. *Effective Clinical Practice*;1:12-22.

McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, & Kerr EA, 2003. *The Quality of Health Care Delivered to Adults in the United States*. *New England Journal of Medicine*, 348(26) pp. 2635-2645.

McHorney CA, Ware JE Jr., Raczek AE. "The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs", *Med Care (Medical care.)*, 1993 Mar; 31(3): 247-63.

McKusick, L., Wiley, J., Coates, T. J., & Morin, S. F. (1986). Predictors of AIDS behavior risk reduction: The AIDS Behavioral Research Project. Paper presented at the New Zealand AIDS Foundation Prevention Workshop, Auckland, New Zealand.

Mudde, A. N., Kok, G. J., & Strecher, V. J. "Self-efficacy as a predictor for the cessation of smoking: Methodological issues and implications for smoking cessation programs", *Psychology and Health*, Vol. 10, 1995, p. 353-367.

Murphy, D. A., Stein, J. A., Schlenger, W., Maibach, E., et al. (2001). Conceptualizing the multidimensional nature of self-efficacy: Assessment of situational context and level of behavioral challenge to maintain safer sex. *Health Psychology*, 20, 281-290.

Murphy, D. A., Multhauf, K. E., & Kalichman, S. C. (1995). Development and validation of a graded, safe-sex self-efficacy scale. *The Behavior Therapist*, January, 8-10.

NSBA, 2003. *Employer health insurance costs up 13.9 percent*. National Small Business Association Advocate (Washington, D.C.) September 9.

Nicholson, Sean, Mark Pauly, Daniel Polsky, Claire Sharda, Helena Szrek, Mark Berger. "Measuring the Effects of Workloss on Productivity with Team Production," April 26, 2004.

Nicholson, Sean, telephone communication with the researcher, February 22, 2005.

Norris SL, Engelgau MM, Narayan KM. "Effectiveness of self-management training in type 2 diabetes: a systematic review of randomized controlled trials", *Diabetes Care*, Vol. 24, 2001, p. 561-87.

O'Leary, A. (1985). Self-efficacy and health. *Behavioral Research Therapy*, 23, 437-451.

Ornstein, SM, Jenkins RG. "Quality of Care for Chronic Illness in Primary Care: Opportunity for Improvement in Process and Outcome Measures", *American Journal of Managed Care*. Vol. 5, 1999, p. 621-627

Parry T, Auerbach R. (2001). Linking Medical Care to Productivity. Integrated Benefits Institute, San Francisco, CA.

Parry, Thomas, telephone communication with the researcher, February 15, 2005.

Pauly, Mark V., Sean Nicholson, Judy Xu, Dan Polsky, Patricia Danzon, James F. Murray, and Marc L. Berger. (2002). A New General Model of the Impact of Absenteeism on Employers and Employees. *Journal of Health Economics*, 11(3), 221-231.

Petosa, R. L., Suminski, R., & Hertz, B. (2003). Predicting vigorous physical activity using social cognitive theory. *American Journal of Health Behavior*, 27, 301-310.

Prochaska, J. O., DiClemente, C. C., Velicer, W. F., Gimpil, S. A., & Norcross, J. C. "Predicting change in smoking status for self-changers", *Addictive Behaviors*, Vol. 10, 1985, p. 395-406.

Reilly MC, Zbrozek AS, Dukes EM. "The validity and reproducibility of a work productivity and activity impairment instrument", *Pharmacoeconomics*, Vol. 4, Iss. 5, November 1993, p. 353-65.

Renders CM, Valk GD, Griffin SJ, Wagner EH, Eijk Van JT, Assendelft WJ. 2001. Interventions to improve the management of diabetes in primary care, outpatient, and community settings: a systematic review. *Diabetes Care*;24:1821-33.

Revicki DA, Irwin D, Reblando J, Simon GE. "The accuracy of self-reported disability days", *Med Care (Medical care.)*, 1994 Apr; 32(4): 401-4.

Rudolph, D. L., & McAuley, E. (1996). Self-efficacy and perceptions of effort: A reciprocal relationship. *Journal of Sport and Exercise Psychology*, 18, 216-223.

Sallis, J. F., Pinski, R. B., Grossman, R. M., Patterson, T. L., & Nader, P. R. (1988). The development of self-efficacy scales for health-related diet and exercise behaviors. *Health Education Research*, 3, 283-292.

Samsa GP, Matchar DB, Goldstein LB, Bonito AJ, Lux LJ, Witter DM, Bain J. 2000, *Quality of anticoagulation management among patients with atrial fibrillation: Results of a review of medical records from 2 communities*. *Archives of Internal Medicine* 160 (7): 967-73.

Schaefer EJ, Augustin LJ, Schaefer MM, et al., 2000. "Lack of Efficacy of a Food-Frequency Questionnaire in Assessing Dietary Macronutrient Intakes in Subjects Consuming Diets of Known Composition." *Am J Clin Nutr* Vol. 71, p. 746-751

Schwarzer, R. (1992). Self-efficacy in the adoption and maintenance of health behaviors: Theoretical approaches and a new model. In R. Schwarzer (Ed.), *Self-efficacy: Thought control of action* (pp. 217-243). Washington, DC: Hemisphere.

Stewart WF, Ricci JA, Chee E, Morganstein D, Lipton R, 2003a. "Lost Productive Time and Cost Due to Common Pain Conditions in the US Workforce." *JAMA*, Vol. 290, No. 18, p.2443-2454

Stewart WF, Ricci JA, Chee E, Morganstein D, Lipton R, (2003b). Lost Productive Work Time Costs From Health Conditions in the United States: Results from the American Productivity Audit. *Journal of Occupational and Environmental Medicine*, 45(12), 1-13.

Stone, R.E, " Playing Disease Management Numbers Games", *Disease Management and Health Outcomes*, December 6, 1999, Vol. 6, p. 343-348.

Velicer, W. F., DiClemente, C.C., Rossi, J. S., & Prochaska, J. O. (1990). Relapse situations and self-efficacy: An integrative model. *Addictive Behaviors*, 15, 271-283.

Von Korff M, Katon W, Unützer J, Wells K, Wagner EH. 2001 Improving depression care: barriers, solutions, and research needs. [PMID: 11401751] *Journal of Family Practice*;50:1-1.

Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, & Bonomi A. 2001. Improving chronic illness care: translating evidence into action. *Health Affairs (Millwood)*; 20:64-78.

Wagner EH, 1998a. "Chronic disease management: what will it take to improve care for chronic illness?" *Effective Clinical Practice*; 1: 2-4.

Wagner EH, 1998b. "More than a case manager", *Annals of Internal Medicine*, Vol. 129, p. 654-6.

Walsh JC, Mandalia S, Gazzard BG, 2002. "Responses to a 1 month self-report on adherence to antiretroviral therapy are consistent with electronic data and virological treatment outcome." *AIDS*, Vol. 16, No. 2, p.269-277.

Westover, S. A., & Lanyon, R. I. (1990). The maintenance of weight loss after behavioral treatment: A review. *Behavior Modification*, 14, 123-137.

Yates, A. J., & Thain, J. (1985). Self-efficacy as a predictor of relapse following voluntary cessation of smoking. *Addictive Behaviors*, 10, 291-298.

Yin, Z., & Boyd, M. P. (2000). Behavioral and cognitive correlates of exercise self-schemata. *Journal of Psychology*, 134, 269-282.

Young, A.S., R. Klap, C.D. Sherbourne, and K.B. Wells. 2001. The quality of care for depressive and anxiety disorders in the United States. Arch Gen Psychiatry 58 (1): 55-61.