LEADING ARTICLE

Dis Manage Health Outcomes 2007; 15 (1): 13-18
1173-8790/07/0001-0013/$44.95/0

Use of the Pre-Post Method to Measure Cost Savings in Disease Management

Issues and Implications

Ariel Linden[1,2]

1 Linden Consulting Group, Portland, Oregon, USA
2 School of Medicine and School of Nursing, Oregon Health & Science University, Portland, Oregon, USA

Abstract

The US disease management (DM) industry continues to endorse the use of the methodologically flawed pre-post design to evaluate financial outcomes, which regularly reports returns on investment of up to 8 : 1. This is in sharp contrast to the peer-reviewed literature and large Medicare demonstration projects that generally report little, if any, cost savings from DM. The industry defends the practice of using the pre-post evaluation design by suggesting that measuring total healthcare costs at the diseased-population level eliminates regression to the mean and accounts for indirect changes in physician behavior. The industry further argues that equivalent and concurrent control groups are not available and that instead, a cost trend of the non-diseased population should be used to provide equivalence. This article illustrates the fallacies of these arguments and demonstrates how the pre-post technique elicits financial results generally favoring the DM program. Given that the continued use and support of this methodology serves only to propagate the concerns over the financial value of DM, it is time that a collective decision be made as to whether maximizing short-term profits is worth jeopardizing the long-term viability of the entire industry. Additionally as important, other healthcare systems around the world are looking to the US DM industry for guidance as they ponder the introduction of DM in their own countries. The inability of DM to accurately measure and achieve financial savings may be inhibiting the widespread initiation of future DM programs.

The US disease management (DM) industry's continued use[1] and endorsement[2] of the methodologically flawed pre-post design[3] for evaluating return on investment (ROI) is akin to the story of the man who searches for his wallet under a lamp-post even though he lost it elsewhere. While it may be tempting to argue that the 'realities of the business world' preclude the use of experimental or quasi-experimental evaluations, the consequence is ubiquitous skepticism of DM's ability to generate ROI that have been reported to range from 2 : 1 to 8 : 1.[4-6]

In 2004, the Congressional Budget Office assessed the peer-reviewed DM literature and concluded that while DM appears to have clinical value for patients and results in high satisfaction levels, "there is insufficient evidence to conclude that disease management programs can generally reduce the overall cost of healthcare services."[7] This finding was further supported by two recent literature reviews which reached similar conclusions.[8,9] In addition, two large DM organizations have withdrawn early from tightly controlled Medicare demonstration projects intended to rigorously evaluate whether DM can improve the quality of care provided to Medicare beneficiaries with chronic conditions and reduce Medicare expenditures. In both cases, the DM organizations concluded that they would not be able to achieve the 5% reduction in costs required by Medicare.[10,11]

Given this dichotomy, the current level of skepticism about financial outcomes is not surprising. When faced with results derived from rigorously conducted studies and poorly controlled evaluations, more often than not people will place stock in the rigorously conducted studies. Thus, the question remains as to

why the US DM industry continues the use and support of the pre-post design for measuring financial outcomes.

The purpose of this paper is to illustrate how the pre-post evaluation design elicits financial results generally in favor of the DM program, to describe how the industry defends the practice of using a flawed evaluation approach, and to raise implications for the future viability of DM.

1. The Pre-Post Approach to Measuring Financial Outcomes

The pre-post DM program evaluation model was initially developed for businesspeople familiar with financial statements such as balance sheets, income statements, and cash-flows, but unfamiliar with research-based evaluation techniques that would control for threats to validity. As a result, the DM financial outcomes analysis has been designed similarly to these other reports, limited to basic comparisons of annual pre- and post-program costs for a diseased population, making it easier for a client to comprehend.

The financial evaluation (also called 'reconciliation') follows four steps. The first step is to identify individuals within the overall population (e.g. health plan, employer, or geographic location) who have the disease, based on their past medical claims history (from here on in this paper, this group will be referred to as the 'diseased population'). This identification process is performed for each measurement period (e.g. pre-program baseline and first program year).

The next step is to aggregate the medical costs accrued for the diseased population for each measurement period. Costs may include disease-specific medical costs (those medical costs directly related to the care of the chronic disease being managed) only or may be expanded to include all medical costs incurred by members identified with the disease. In either case, 'medical cost' typically includes only those types of expenses captured in medical claims; therefore, it does not include indirect health-related costs such as productivity. This model may also call for the exclusion of certain types of costs. For example, in evaluating a diabetes mellitus program, all members who also have another specifically defined disease (such as cancer or AIDS) may be removed from the analysis. Likewise, certain types of costs (such as those related to trauma or infertility treatment) may be excluded from the analysis even when incurred by an otherwise eligible member.[3]

The third step is to divide the aggregated costs derived in the step above by the sum of months of enrollment during that study period for the diseased population (referred to as 'member months') and further divide that figure by 12 months to derive a per-member-per-month (pmpm) cost.

The final step is to determine a year-over-year cost trend in the non-DM-managed population to serve as a control. As shown in the hypothetical example in figure 1, by following this simple formula the program will be deemed a financial success as long as the program year pmpm costs (inclusive of program fees) were lower than the estimated costs (based on the trend).

2. Arguments for and Responses against the Pre-Post Model

While a comprehensive list of the major threats to validity using the pre-post approach in DM have been described elsewhere,[3] the intent of this article is to review specific arguments made in support of the DM industry's standard evaluation design[2] that are problematic. This discussion may also help to explain how results of such excessive financial savings are obtained using this methodology versus those derived from more rigorous designs.

2.1 Argument 1: Measurement at the Diseased-Population Level Eliminates Regression to the Mean and Other Sources of Bias

Supporters of the pre-post design suggest that the effect of regression to the mean (RTM) is eliminated when financial outcomes are measured on the entire identified diseased-population level because the high and low outliers cancel each other out.[1] This theory is only correct if all constituents of the diseased population

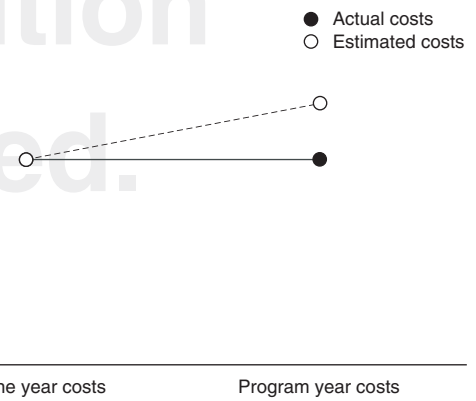

Fig. 1. Hypothetical cost outcomes derived from a pre-post evaluation compared with estimated costs based on a 'trend.'

1 An article discussing the effects of regression to the mean, entitled 'Estimating the Effect of Regression to the Mean in Health Management Programs', is published in this issue of *Disease Management & Health Outcomes* (Issue 15 Vol. 1: 7-12).

are correctly identified and there is neither attrition nor new entrants introduced into the diseased population.

However, while claims analysis can easily identify those individuals who fall into the high-cost category, individuals with no claims will never be identified. This selection bias results in over-representation of higher-cost patients in the 'identified' population who will have a natural decline in costs in the following program year. Thus, in a diseased-population-based pre-post model, the outcomes will generally favor the DM program.[12]

Advocates further justify the use of diseased-population level measurement in lieu of evaluating program participants versus controls by arguing that DM interventions impact the entire diseased population by creating a 'spill-over' effect to patients and providers not directly participating in the program. In other words, physicians will apply changes in their practice patterns (as a result of DM) across all of their patients and this will lead to cost savings across the entire diseased population. Not only is there a lack of published research to support these contentions but one could easily argue that any positive changes that occur in the treatment of chronic illness at the diseased-population level are heavily influenced by a completely independent factor: the requirements of accreditation or governmental oversight. This is supported by a 2006 National Committee for Quality Assurance study of health plans across the nation, which reports continued year-over-year improvements in all 'effectiveness-of-care' measures such as controlling high blood pressure, β-adrenoceptor antagonist use after an acute myocardial infarction, and provision of comprehensive diabetes care.[13]

These findings suggest that DM outcomes measured at the diseased-population level do not ferret out changes due to naturally occurring phenomena (e.g. RTM) or changes as a result of activities occurring outside of the intervention and, indeed, may even be a 'free rider' on improvements in usual care as a result of external influences. Moreover, the type of accreditation/regulatory measures mentioned above disproportionately focus on those cardiac, respiratory, and metabolic conditions most often targeted by DM.

2.2 Argument 2: A Comparison Group that is Both Equivalent and Concurrent May Not Be Available in Applied Settings

A second argument typically used to support the use of the pre-post methodology is that an equivalent, concurrent comparison group may not be available for the analysis. First, if the entire diseased population was not used as the unit of measure, non-participating individuals would be available for use as controls. Second, while a concurrent control group is always preferred, a historical control group can be developed to achieve equivalency. For example, biological markers of disease (i.e. blood pressure, blood glucose level, and/or cholesterol) are not time-dependent and could be used as matching variables across study groups.

In short, there is little merit to the argument that the pre-post evaluation must be used because no valid research design using a comparison group exists. On the contrary, several tutorials have been written on the application of research-based designs to DM outcomes measurement and many of them have successfully been put into practice.[12,14-23]

2.3 Argument 3: Total Healthcare Claims (Medical and Pharmacy) Costs Should Be Used as the Financial Outcomes Metric

Measurement bias is introduced when costs are used as the primary outcome. As recently demonstrated,[24] healthcare costs have been steadily rising while acute healthcare utilization for chronic illness has remained flat (see figure 2), indicating that costs are heavily influenced by unit pricing – an area beyond the control of DM. Moreover, DM programs can only realistically achieve a positive ROI by substantially reducing costly hospital admissions and, to a lesser degree, emergency department visits, given that outpatient services and pharmacy use should naturally increase as a function of better adherence to evidence-based practice guidelines.[7,24] Therefore, measuring total healthcare claims costs directly obscures the causal link between what the program is intended to impact directly (e.g. reductions in acute utilization) and what it cannot impact (e.g. unit pricing, increased costs associated with appropriate use of outpatient services and pharmacy). The industry's sole argument for using actual dollars for program evaluation calculations is that it "appears to the client as relevant and less susceptible to manipulation."[2]

Given these concerns, a more appropriate approach to address changes in costs associated with the program may be to assign a standard cost for each unit of utilization that can be impacted by the program and to track period-over-period changes in that utilization, holding unit costs constant.[24]

2.4 Argument 4: Actual Costs of the Diseased Population Should Be Compared with the Cost Trend of a Non-Chronic Population

This dimension of the DM industry's approach raises two concerns. First, as discussed in section 2.3, developing a cost-trend

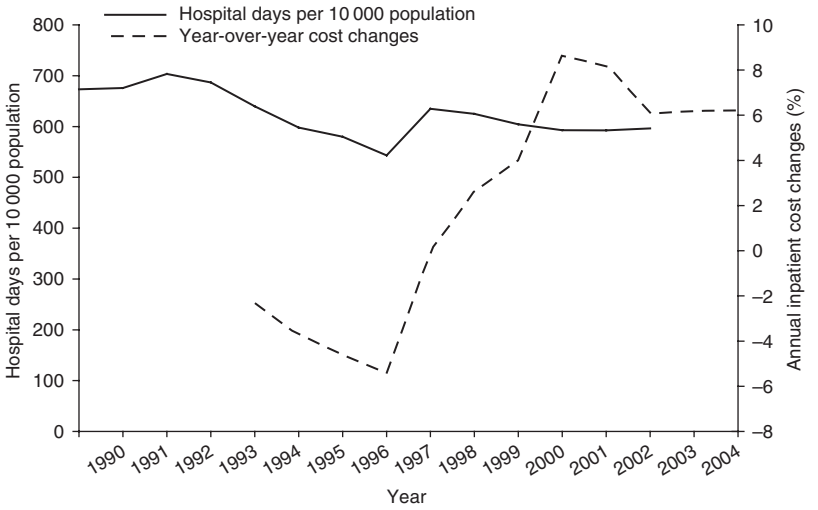

Fig. 2. Aggregate hospital days per 10 000 population compared with inpatient medical cost trend rates from 1990 to 2004.[24]

model that does not adjust for increases in unit pricing will always bias the results in favor of the DM program. More specifically, a program will appear to be financially effective as long as claims costs (driven primarily by unit pricing) are lower than the expected trend, regardless of whether or not any change actually occurred in acute utilization as a function of the program's intervention. One cannot assume that changes in unit pricing are equivalent between groups.

Second, the non-chronically ill population typically differs from the diseased population on nearly all demographic or economic variables. Using a non-comparable group to determine expected trends in cost will introduce measurement bias and limit the ability to draw accurate conclusions about the results. Only if many serial observations of cost are determined to be equivalent between the populations can some degree of confidence be achieved in using the non-chronic trend as a comparator for the chronic population.

These concerns are illustrated by the National Hospital Discharge Survey (NHDS) data[25] presented in figure 3; three major chronic disease categories (circulatory, endocrine, and respiratory) are compared with 'all other' discharges. As shown, discharges have been flat in categories in which the majority of conditions are chronic, while non-chronic discharges have been trending upward at a rate of 7.5% over the observed 5-year period. The assumption can also be made that the 'all other' category of discharges are more costly than the chronic conditions because many of the diagnoses require surgeries, (i.e. injuries, deliveries, and/or complications) as opposed to less costly medical stays. While these data do not represent chronic versus non-chronic populations *per se* and some degree of overlap is inevitable, these data do demonstrate that both the level and the trend of discharges for non-chronic conditions are significantly higher than those categories considered chronic. These findings suggest that applying a non-chronic 'trend' to the diseased population will bias the results in favor of the DM program.

3. Implications for Disease Management in the US and Internationally

Collectively, the methodological issues described herein and elsewhere[3] help explain why commercially reported financial outcomes are substantially higher using the diseased-population-based pre-post evaluation than those found in well controlled

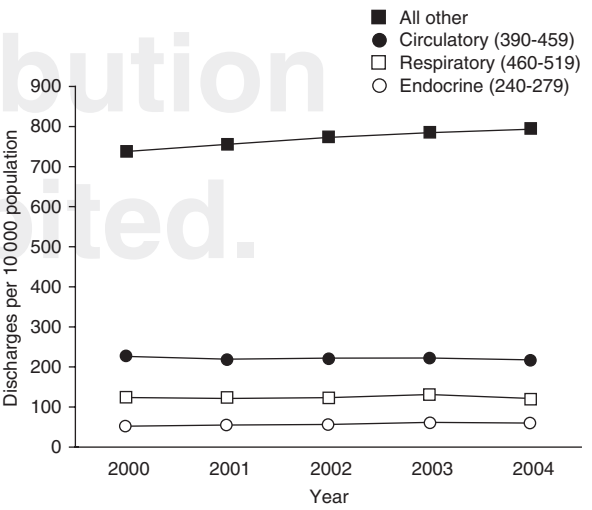

Fig. 3. Hospital discharges per 10 000 population for three major chronic disease categories and all other discharges.[25] Numbers in parentheses are International Classification of Disease (9th edition), Clinical Modification (ICD-9-CM) codes.

studies. Given that most DM programs offer purchasers a money-back guarantee if medical cost savings targets are not achieved, a DM company's risk is greatly diminished when an evaluation design is used that provides favorable outcomes.

Given that DM does not operate in a controlled environment, every attempt must be made to limit the threats to the validity of the outcomes. As indicated above, there are many quasi-experimental research designs that can be readily applied to DM in which observed sources of bias are controlled for. Moreover, it is prudent to employ at least two different evaluation techniques to ensure that the direction, and possibly the magnitude, of the outcomes are similar. Pre-post evaluations that elicit ROI values of 2 : 1 to 8 : 1[4-6] are just not credible, especially while robust evaluations do not show any program effect. Even simple plausibility tests, such as determining where supposed cost savings occurred, can indicate whether the results can stand up to scrutiny.

The US DM industry's continued use and support of the pre-post methodology serves to propagate the concerns over the financial value of DM. A collective decision must be made as to whether maximizing short-term profits is worth jeopardizing long-term viability.

Similarly, industry leaders should also consider the consequences for DM if the large, randomized controlled Medicare demonstration projects fail to achieve the cost savings they claimed based on biased evaluation designs. Additionally as important, other healthcare systems around the world are looking to the US DM industry for guidance as they ponder the introduction of DM in their own countries. The inability to accurately measure and achieve financial savings with DM may be inhibiting the widespread initiation of future DM programs.

4. Conclusion

Despite the large discrepancy in financial outcomes reported by commercial DM programs versus those reported by tightly controlled studies, the US DM industry continues to use and endorse the diseased-population-based pre-post evaluation model as the pre-eminent design. This model elicits substantially higher cost savings than those of more robust designs and has led to much concern and skepticism as to whether DM can truly impact economic measures. Contrary to statements made otherwise, there are several quasi-experimental designs that can be readily applied to DM for measuring program effectiveness, taking into account the vagaries among different people, settings, treatments, or outcomes. Given that the viability of DM is to a large extent dependent on the conviction that DM can truly achieve financial savings, a concerted move by the industry to adopt more robust program evaluation designs is warranted.

Acknowledgments

No sources of funding were used to assist in the preparation of this review. The author has no conflicts of interest that are directly relevant to the content of this review.

References

1. American Healthways and John Hopkins Consensus Conference. Standard outcome metrics and evaluation methodology for disease management programs. Dis Manag 2003; (6) 3: 121-38
2. Disease Management Association of America. Outcomes guidelines report. Washington, DC: DMAA, 2006
3. Linden A, Adams J, Roberts N. An assessment of the total population approach for evaluating disease management program effectiveness. Dis Manag 2003; 6 (2): 93-102
4. Johnson A. Disease management: the programs and the promise. Milliman USA research report May 2003 [online]. Available from URL: http://www.milliman.com/tokyo/publications/hrr14.pdf [Accessed 2006 Dec 9]
5. Shutan B. The DM Rx: disease management programs producing fast and meaningful outcomes, impressive ROI. Employee Benefit News 2004; 18 (13) [online]. Available from URL: http://www.matria.com/resources/articles/dm/EBNPetit.pdf [Accessed 2006 Dec 9]
6. Health Information Designs, Inc. HID provides a disease management program to address the specialized needs of high risk patients [online]. Available from URL: http://www.hidinc.com/WhatWeDo/whatwedo_utilizingexperiences.htm [Accessed 2006 Dec 9]
7. Congressional Budget Office. An analysis of the literature on disease management programs [online]. Available from URL: http://www.cbo.gov/showdoc.cfm?.index=5909&sequence=0 [Accessed 2006 Dec 9]
8. Ofman JJ, Badamgarav E, Henning JM, et al. Does disease management improve clinical and economic outcomes in patients with chronic diseases? A systematic review. Am J Med 2004; 117 (3): 182-92
9. Goetzel RZ, Ozminkowski RJ, Villagra VG, et al. Return on investment on disease management: a review. Health Care Financ Rev 2005; 26: 1-19
10. Benko LB. Disease management strikes out. Mod Healthc 2006 Jan 23; 36 (4): 8-10
11. E-CareManagement News. Commentary: lifemasters pulls the plug on Oklahoma Medicare Health Support project [online]. Available from URL: http://www.bhtinfo.com/10_25_06.htm [Accessed 2006 Dec 9]
12. Linden A. Estimating the effect of regression to the mean in health management programs. Dis Manage Health Outcomes 2007; 15 (1): 7-12
13. National Committee for Quality Assurance. State of health care quality 2006 [online]. Available from URL: http://www.ncqa.org/Communications/SOHC2006/SOHC_2006.pdf [Accessed 2006 Dec 16]
14. Linden A, Adams J, Roberts N. Evaluating disease management program effectiveness: an introduction to time series analysis. Dis Manag 2003; 6 (4): 243-55
15. Linden A, Adams J, Roberts N. Evaluating disease management program effectiveness: an introduction to survival analysis. Dis Manag 2004; 7 (3): 180-90
16. Linden A, Adams J, Roberts N. Evaluating disease management program effectiveness adjusting for enrollment (tenure) and seasonality. Res Healt Fin Manage 2004; 9 (1): 57-68
17. Linden A, Adams J, Roberts N. Using propensity scores to construct comparable control groups for disease management program evaluation. Dis Manage Health Outcomes 2005; 13 (2): 107-27
18. Linden A, Adams J. Evaluating disease management program effectiveness: an introduction to instrumental variables. J Eval Clin Pract 2006; 12 (2): 148-54
19. Linden A, Adams J, Roberts N. Strengthening the case for disease management effectiveness: unhiding the hidden bias. J Eval Clin Pract 2006; 12 (2): 140-7

20. Linden A. Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis. J Eval Clin Pract 2006; 12 (2): 132-9
21. Linden A, Adams J, Roberts N. Evaluating disease management program effectiveness: an introduction to the regression-discontinuity design. J Eval Clin Pract 2006; 12 (2): 124-31
22. Linden A, Trochim WMK, Adams J. Evaluating program effectiveness using the regression point displacement design. Eval Health Prof 2006; 29 (4): 1-17
23. Linden A, Adams J. Determining if disease management saves money: an introduction to meta-analysis. J Eval Clin Pract. In press
24. Linden A. What will it take for disease management to demonstrate a return on investment? New perspectives on an old theme. Am J Manag Care 2006; 12 (4): 217-22
25. National Center for Health Statistics. Centers for Disease Control and Prevention. National Hospital Discharge Survey 1990-2003 [online]. Available from URL: www.cdc.gov/nchs/datawh/ftpserv/ftpdata/ftpdata.htm-nhds. [Accessed 2006 Dec 14]

About the Author: Dr Linden is a health services researcher whose focus is on evaluating the effectiveness of disease management and health promotion initiatives. In this area alone, he has recently published more than 40 peer-reviewed articles, book chapters, editorials, and a position paper for the Disease Management Association of America. He is President of the Linden Consulting Group and holds adjunct positions in both the School of Medicine and the School of Nursing at Oregon Health and Science University.

Correspondence: Dr *Ariel Linden*, Linden Consulting Group, 6208 NE Chestnut Street, Hillsboro, OR 7124, USA.
E-mail: alinden@lindenconsulting.org