

A Stepwise Approach for Measuring Health Management Program Impact

Leading the Way Toward Practical and
Scientifically Sound Reporting

A Nurtur® White Paper

June 2009

Prepared by:

Nurtur Informatics

Introduction

Nurtur is committed to long-term relationships with its clients, and the foundation for any long-term relationship is trust. When it comes to measuring the financial impact of health management programs, trust is more important than ever. Why? Because at the core of the issue, the evaluation of health improvement programs is an exercise in measuring *that which did not occur*: the avoided emergency room visit for a nebulizer treatment for a young child with asthma; the heart attack that did not happen to the middle-aged factory worker; the hospitalization that was avoided for the person with emphysema.

Such measurements are complicated at best, and the only *universally accepted* way to *estimate* the impact of most clinical interventions is to conduct a prospectively-randomized clinical trial where the results of a “treatment” group are compared to the results of a “control” group. While such trials are the norm in academic and government-sponsored

medical research, they are not viable in a commercial setting. Moreover, the various approaches to estimating the financial impact of health management programs and, ultimately, some claims to their return on investment (ROI), have been controversial and subject to much industry debate¹. Key weaknesses and criticisms of traditional measurement approaches applied in commercial settings have been: statistical “regression to the mean”; lack of an equivalent control group to which the results of the program participants can be compared; and lack of sufficient numbers of data points for statistical power.

At Nurtur, we have prided ourselves on three underlying key principles to earning our clients’ trust: honesty, integrity and transparency. We have been honest with our clients about when we believe we have contributed to results, and equally honest when the observed changes are so extraordinary that they could not possibly relate to the impact of a health management program

alone. More recently, we have applied significant resources to establishing a methodology and approach to measuring our programs in a way that builds on our legacy of trust. It is an approach that abides by rigorous standards that pass muster with measurement and analytics experts. We call this method the *Nurtur Stepwise Approach*, and we are proud that it is the basis for our participation as a beta-test site for the Population Health Impact Institute's (PHII) Methods Evaluation Process (MEP™) Accreditation Program².

The standards that the Nurtur Stepwise Approach apply to the measurement of health management programs are:

- **Data completeness.** The data must be sufficiently detailed and complete in order to measure specific events expected to be impacted by the health management program;
- **Statistical power.** The evaluations undertaken must demonstrate adequate statistical power.

- **Reasonable evidence.** There must be reasonable, observed evidence of expected impacts on behavioral and/or biometric/clinical measures.
- **Utilization.** Expected changes in medical care utilization must be observed.

If and only if all four conditions are met is it appropriate to continue the evaluation to assess financial impact, since any cost differences that are *observed* may be *reasonably attributed* to the program. If all four standards cannot be met, an alternative, customized evaluation may be performed, as determined on a case-by-case basis with each client.

¹Mattke S, Seid M, Ma S. Evidence for the effect of disease management: is \$1 billion a year a good investment? *Am J Manag Care.* 2007;13:670-676; Congressional Budget Office. An analysis of the literature on disease management programs. Washington, DC: October 13, 2004; Goetzel RZ, Ozminkowski RJ, Villagra VG, Duffy J. Return on investment in disease management: a review. *Health Care Finance Rev.* 2005;26:1-19.

²<http://www.phiiinstitute.info/>

Key Issues Regarding Health Management Program Evaluation

Challenges

Regression to the mean

This concept refers to the phenomenon that a group of people pre-selected for scoring above the mean on a clinical measure are likely to score lower on average, purely by chance, if the measurement is later repeated. For instance, consider that an individual hospitalized for a heart attack in year one is targeted for enrollment in the heart health management program in year two. A heart attack is a rare event: therefore, the likelihood that this individual will suffer from another heart attack in year two is quite low (roughly 10% or less). So, if one is measuring the heart attack hospital admission rate for year two, and the cohort in question has a disproportionate number of individuals who experienced a heart attack in year one, the results are likely to show improvement regardless of true program impact. For this reason, it is important to measure the hospital admission rate in the control group as a comparison and benchmark for true program impact.

Lack of equivalent control group

The most rigorous study design is the randomized control trial, a true experiment in which two equivalent cohorts are studied. One cohort receives care or services while the other does not. The difference in outcomes between the two cohorts are then measured and analyzed. It is, however, unrealistic to withhold benefits of health programs from a “control” population. So, most evaluations of DM programs are observational or quasi-experimental, and do not automatically establish causality in a statistically valid manner.

Lack of statistical power

There is a strong relationship between the sample size (n) and the ability to find statistical significance. The larger the sample size, the less likely it is that measurement variations are attributable to “random chance” and the more likely it is that such variations are attributable to the interventions being measured. The limited study group size in many health management program evaluations prevents the establishment of true, statistically valid causality, trends or effects.

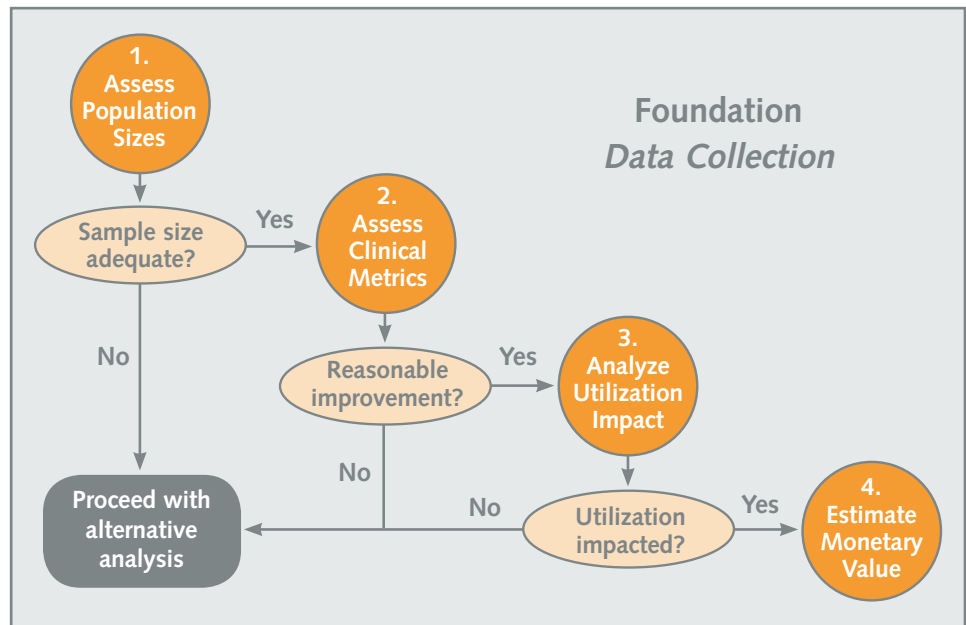
Description of Approach

The foundation of the Nurtur approach is accurate, timely and complete data collected from sources such as medical and pharmacy claims, biometric screenings and health risk assessments (HRAs), and activity and clinical data collected as part of the health coaching process. At the onset of the program evaluation process, we assess whether we have received all data and whether the data are complete. The data fields we need include claimant identification, diagnosis code(s) and Centers for Medicare & Medicaid Services (CMS) revenue codes, quantity, cost details and provider information. Examples of claims data codes that may be used in the evaluation are ICD-9-CM, CPT-4, HCPCS, and NDC.³ If it appears that any data received by Nurtur are incomplete or inaccurate, Nurtur will collaborate with the client and/or the claims data provider to ensure that all necessary data elements are properly collected.

³ICD-9-CM (International Statistical Classifications of Diseases) codes are alphanumeric designations given to every diagnosis, description of symptoms and cause of death attributed to human beings. CPT-4 (Current Procedural Terminology) codes are numbers assigned to every task and service a medical practitioner may provide to a patient including medical, surgical and diagnostic services. HCPCS (Healthcare Common Procedure Coding System) codes are numbers assigned to every task and service a medical practitioner may provide to a Medicare patient including medical, surgical and diagnostic services. NDC Codes, found in the National Drug Code Directory, are unique, three-segment numbers for each prescription or insulin drug.

Once it is determined that the data are sufficiently complete and accurate, our program methodology proceeds according to the four steps outlined in **Figure 1**. In order to proceed from one step to the next as prescribed in the Stepwise Approach, the previous step must be satisfactorily completed. For example, once the expected utilization impact is detected, it is reasonable to continue the evaluation to assess financial impact, since any cost differences that are “observed” may be reasonably attributed to the program. In situations where the requirements of a specific step cannot be met, an alternative type of evaluation may be performed, as determined on a case-by-case basis.

Figure 1: Stepwise Approach to Health Management Program Evaluation



Step 1: Assess the defined population sizes.

Question: Are there, *at minimum*, 700 eligible lives with the target conditions?

The presence of random variation requires that the target population is large enough to make the analysis meaningful. The resources required for the analysis are not well spent if a meaningful result is not possible due to lack of statistical power. We have determined that there needs to be at least 700 eligible lives with the target conditions in order to evaluate a health management program. The actual number, however, will vary based on specific data, the number of data observations and metrics in any given evaluation. The goal is to have enough statistical power to be able to reasonably attribute any positive outcomes to program participation.

Step 2: Assess available clinical and behavioral measures for evidence of program effectiveness.

Question: Between the initial baseline and subsequent outcome measurements, were there improvements in the self-reported clinical (example: LDL cholesterol) and behavioral (example: minutes of exercise per week) outcome metrics?

The objective of this step is to establish a reasonable “link” between the interventions and the intended results. Nurtur assesses the data to determine whether sufficient evidence exists to form the basis for reasonably assuming causation:

- The strength of the association (the stronger the statistical relationship between a supposed cause and an effect, the greater the chance that a causal relationship exists).
- The temporal relationship (health improvements occur after program participation).
- Biologic plausibility (the association should make sense from the perspective of biology (example: increased exercise leads to reduced weight)).
- Analogy (the participant attributes positive changes to the program).

In addition to the administrative data, Nurtur collects participant-reported outcomes and participant satisfaction. This data include participant performance on key measures and leading indicators of long-term cost and healthcare quality improvement, such as condition-specific risk-related behavior, biometric values, treatment compliance, disease severity and/or functional status.

In order for the analysis to proceed to the subsequent step, these measures should show statistically significant improvement. Not every measure needs to improve for every participant for every condition, but there needs to be a clear trend in the appropriate direction with some measures showing statistical significance. These measures can be aggregated across conditions; for instance, into rates of participants with any favorable change, participants meeting a certain threshold

of favorable change, participants with two favorable changes, etc., to help achieve statistical power. Evidence of positive results that can be plausibly attributed to the program is needed in order to assign program effect.

If there is no evidence of program effectiveness, there is no reason to continue the analysis since utilization differences cannot and should not be “reasonably” attributed to the program. Nurtur identifies the reason a program effect could not be detected (e.g., a lack of statistical power or inadequate participant time in program) before concluding that the intervention is ineffective. It is also important to understand that the failure to find a program effect may not mean that there was no effect, only that it could not be heard above the normal, random “noise” that exists in healthcare data.

Step 3: Analyze the utilization of services for evidence of program effectiveness.

Question: Using administrative claims data, is there a difference in utilization rates between the participant population and the matched control? (Propensity score-matched cohort analysis).

Propensity score-matched cohort analysis: Using this approach, the participant population is compared to a control group. The control group is generated by matching each health management program participant with a comparable non-participant determined by a propensity score. The propensity score is calculated using a logistic regression model with participation as the dependent variable and with candidate program, presence of comorbidities, age and gender as the independent variables. Logistic regression is a statistical tool that is used extensively in medical and social sciences to predict the probability of an occurrence of an event.

Question: Using administrative claims data, is there a difference in utilization rates between individuals in the historical control period vs. the study period? (Target population utilization trend analysis).

Target population utilization trend analysis: The utilization rates in the historical control period are compared to the utilization rates during the study period (the period since program implementation). It is imperative that this analysis measures the utilization of condition-specific services, as opposed to costs. This analysis needs to account for changes in population make-up as well as any other factors that show significant change over time that could affect the use of services (such as benefit levels or therapeutic developments).

Step 4: Estimate monetary value from the program.

Question: What is the cost impact of the changes in utilization identified in Step 3?

Question: Taking into account the fees paid to Nurtur, what is the overall ROI?

The basic calculations are as follows:

$$\begin{aligned} \text{Expected Costs} - \text{Actual Costs} &= \text{Savings Estimate} \\ \text{Savings Estimate} / \text{Fees} &= \text{ROI} \end{aligned}$$

There are a variety of reasonable methodologies for assigning monetary value to program outcomes. Nurtur's preferred approach is to measure the condition-relevant cost savings in each of the condition populations. Depending on the group of conditions being evaluated, the analysis may require that the savings or cost impact for each condition be calculated separately. Of course, this analysis is most often conducted with a focus on medical costs. True ROI takes into account other key items of value to the client, which may include return-to-work (for instance for a Medicaid population) or productivity/absenteeism (for an employed population). Nurtur works with each client to assess the client's ability to provide additional data that can be included in ROI calculations.

Summary

The Nurtur Stepwise Approach allows for the proper prerequisites to be met before proceeding to the next level of analysis, thereby avoiding inappropriate conclusions about the impact of programs on costs. Additionally, this evaluation approach reflects ongoing efforts at Nurtur regarding the most appropriate methodologies for reasonably and responsibly measuring the impact of health management programs, and, ultimately, assessing their impact on claim costs.

In order to ensure that our methodology is consistent with the highest industry standards for fair and unbiased analysis of health management program outcomes, we have engaged as a beta test site in the Methods Evaluation Process (MEP™) Accreditation developed by the Population Health Impact Institute (PHII). The Population Health Impact Institute is a 501(c) 3 non-profit organization that has developed standards to promote high-quality reporting in healthcare. The MEP Accreditation offered through PHII creates a set of standards for reporting and identifies the issues of equivalence in the referent group, transparency of methods and confounding factors as essential to quality outcomes reporting. As a beta-test site, Nurtur has submitted its five health management programs for accreditation. Clients can be assured that Nurtur will provide a high quality, valid analysis of our interventions and outcomes.

Nurtur continues to dedicate time and effort to the scientific improvement, rigor, integrity and transparency of outcome measurement reporting efforts for our clients as well as for the health population management and the overall health and wellness industry.

About Nurtur

Nurtur is the life, health and wellness company dedicated to helping people transform their lives with support, encouragement and motivation. The company offers an inclusive people-centered approach designed to help people at all life stages address the life issues that get in the way of health as well as the health issues that complicate living. Nurtur provides life and health management programs, concierge, EAP, training and consulting services to multi-market segments including employers, unions, plan sponsors, Third Party Administrators, commercial health plans and government agencies to encourage healthy behaviors, promote healthier workplaces, improve productivity and reduce healthcare costs.

To learn more about the Nurtur Stepwise Approach and how we can help your organization, please contact us at:

Corporate Headquarters
20 Batterson Park Road
Farmington, CT 06032

National Operations Center
4000 McEwen Road
Dallas, TX 75244

800 293 0056 toll free
www.nurturhealth.com