Sustained Economic Value of a Wellness and Disease Prevention Program: An 8-Year Longitudinal Evaluation

Steven M. Schwartz, PhD,[1,2,5] Shawn T. Mason, PhD,[1,3] Chun Wang, MS,[1] Lidia Pomana, MA,[1,5]
Maren E. Hyde-Nolan, MA,[1,4] and Eli W. Carter, BS[1]

Abstract

The objective of this study was to determine the sustained economic impact of a health promotion/disease prevention program delivered through a large regional health plan. This was a retrospective analysis of health risk, health claims, and cost using a mixed model factorial design for the years 2002–2009 that compared program participants to nonparticipants. All analyses were adjusted for age, sex, morbidity, and baseline health care costs as appropriate. The findings presented herein indicate a positive return on investment (ROI) for each program year with ratios ranging from a low of 1.16:1 to a high of 2.83:1. The average ROI collapsed across all 8 years was 2.02:1. The 2009 ROI approximated over $6 million in total savings. This study demonstrates the sustained economic value of a comprehensive health promotion program. (*Population Health Management* 2014;17:90–99)

Introduction

THE RISE OF HEALTH CARE COSTS represents a significant threat to economic competitiveness, not to mention the health and well-being of the workforce and the population at large. In recent years industry discussions have shifted from a fixation on the disease management of high-cost chronic conditions to full-spectrum programs. Addressing the entire population and health continuum is now referred to as "population health." To date, few population health-oriented programs and approaches have demonstrated the viability of meaningful impact on population-level health profiles in an economically sustainable manner over extended time periods. An evidence base is building around population health that demonstrates the effectiveness of some programs to favorably improve the risk and cost profiles of large populations,[1–13] and 3 separate reviews of this literature have concluded that such programming can produce a favorable return on investment (ROI) and positively impact the risk profile at a population level.[14–16] However, the preponderance of these studies have taken a relatively short view of the available data leaving long-term sustainability in question.[17] Recent exceptions to the relative paucity of longitudinal data include work

published by Neville, Merrill, and Kumpfer, which showed some promise with regard to favorable risk profiles over an 8-year period,[18] as well as work by Byrne and colleagues,[19] neither of which examined financial outcomes specifically.

The escalating costs associated with health are currently being assumed in large part by employers and insurers (with increasing costs shifting to the consumer and practitioners). This pressure has fostered the growth of the health and wellness industry as an increasingly ubiquitous strategy to improve the health of large populations and concurrently slow the increasing rate of associated costs. In fact, health and wellness programs have become a central business strategy for many progressive employers and health plans. Program elements range widely from community-based efforts[20] to disease-specific models,[21,22] and comprehensive continuum of health approaches,[23] all delivered through various technological and interpersonal channels.

In the current literature, the economic value of individual programs is usually ascertained retrospectively after a period of 4 years or less.[10,12,13,24] This is typically because of the difficulties of maintaining a longitudinal database on a population cohort, cohort attrition, ongoing changes in programming characteristics, and the limits of the randomized

---

[1]Wellness & Prevention Inc., a Johnson & Johnson Co., Ann Arbor, MI.
[2]University of Michigan, Ann Arbor, MI.
[3]Johns Hopkins University School of Medicine, Baltimore, MD.
[4]Wayne State University, Detroit, MI.
[5]Dr. Schwartz and Ms. Pomana have changed positions since this article was written. Dr. Schwartz is now employed at Social Wellth and Ms. Pomana is now employed at Rosetta.

trial design, including but not limited to the complexities of creating comparative groups. This begs the question—although a wide range of organizations and the general public may benefit from health and wellness programming clinically and financially[25]—of whether the benefit is sustainable and for how long.

This study reports on a longitudinal set of wellness and disease prevention program evaluations focusing on economic sustainability. The program is an established health and disease prevention program provided through a large insurer over an 8-year period. The current paper specifically addresses the longitudinal economic value of such a program with a specific emphasis on the following longitudinal questions:

- Who are program participants? How do they differ from nonparticipants?
- What are the estimated economic/cost benefits and ROI of the program over time? How does participant utilization and cost compare with that of nonparticipants?
- Is there a relationship between cost, utilization, and degree of program engagement?

Methods

The program

The health and wellness program evaluated was a comprehensive health promotion and disease prevention program. The program was provided by an independent health plan to its members as part of a variety of programs, services, and support to improve the health and well-being of its members. This health plan includes health maintenance organization, preferred provider organization, and consumer-directed health plan products. Health risk assessments (HRAs) were completed by participants, and followed by biometric screening (ie, blood pressure, total cholesterol, lipid subfractions, glucose, body composition indices of height, weight, body mass index [BMI], waist circumference, percentage of body fat) and lifestyle intervention components. Other screenings may have been recommended based on sex, age, and risk profile (eg, pap smear, bone density, sigmoidoscopy).

Participants also underwent a counseling session with a health care professional to review their biometric values, identify their particular risks, discuss wellness goals, and develop a behavior change action plan. Counseling sessions were offered in several formats including individual, group, face-to-face, and telephonic. Tailored, online digital health coaching intervention programs were provided to participants that focused on weight management, nutrition management, smoking cessation, and stress management. These digital health coaching programs were designed as behavioral health interventions to aid participants to create and reach health-promoting goals for the purpose of changing lifestyle-related behaviors. These online programs were not available to participants during the first 2 years of the program as evaluated here.

The sample

Data presented in Table 1 show the sample sizes for all eligible plan members and program participants by year

TABLE 1. SAMPLE SIZE BY PARTICIPATION STATUS AND YEAR

		Participation	
	Total Sample Size	N	% of total
2002	189,724	15,328	8.1%
2003	200,609	13,936	6.9%
2004	212,945	14,878	7.0%
		HRA#1 N=10,215	4.8%
		HRA#2 N=4,962	2.3%
2005	224,609	16,555	7.4%
2006	325,801	18,903	5.8%
2007	329,707	16,726	5.1%
2008	340,118	15,239	4.5%
2009	336,515	13,243	3.9%
Total	457,371	57,940	12.7%
Yearly Wellness Program		1,211	0.3%

Notes:
(1) In 2004 the organization changed health risk assessments (HRAs). Differences in constructs and scales are noted where relevant. There is an overlap of 299 people between the 2 versions of the HRA.
(2) Yearly Wellness Program represents the total and percentage of individuals who participated in all 8 years evaluated within this manuscript.
(3) In some cases the constructs in the 2 different HRA instruments are incompatible or absent, which limits some comparisons from one year to another.
(4) This table reflects the total sample of the participation in the study. The number does not reconcile with most of the analyses in the manuscript because of filtrations and subpopulations that were studied in each section.

over the 8-year evaluation period. The relative trend shows a steady increase in the number of plan members from 2002 to 2009. There was a concomitant decline in the rate of participation (8.1% to 3.9%) when considered as the percent of eligible. A total of 2124 unique participants used the online digital health coaching programs from 2004 to 2009. Of those who utilized these programs, about 44% of participants used the weight management program, followed by 35% who used the nutrition management program, 18% who used the stress management program, and 3% who used the smoking cessation program. The actual gross number of participants over the 8-year term fluctuated with no clear trends. Incentives may have been used with some plan customers in some years but the details of such program elements were unavailable for consideration in any analyses.

Because there is an inherent self-selection bias with non-randomized participation in wellness programs, interpretation of program outcomes relative to nonparticipants must take this biasing into account. For all between-cohort analyses a subset of nonparticipants were matched to participants using the propensity score matching method, followed by chi-square and *t* tests, which assured no significant differences in the characteristics of participants and matched nonparticipants. The propensity score matching method is a widely used statistical matching method in observed studies. It generates the predicted probability that an individual receives the treatment of interest from 1 or many confounding variables. For each participant the procedure seeks a nonparticipant with the same or nearly the same estimated probability of inclusion in the treatment, in order to minimize the distance between matched cases on those confounding variables.[26–29]

Each year a subsample of nonparticipants was one-to-one matched to participants on sex, age, morbidity score, and total health care costs in the pre-participation year(s); no significant difference was found between participants and nonparticipants after matching ($P > .05$). These matching variables also were set as covariates in the cost comparisons to better reveal the impact of the program.

Demographic data on sex and age of participants and nonparticipants are included in Table 2. These data suggest that women were more likely to participate in the program relative to men in 2009, but this was consistent with the gender distribution in the eligible membership as a whole. This trend was consistent in years 2006–2009 but somewhat reversed from the more even gender split in years 2002–2005. This change may be related, in part, to worksite recruitment efforts or change in benefits structure initiated between 2005 and 2006. Program participants also were significantly older than their 2009 nonparticipating counterparts (a finding consistent over all 8 years of data). The mean ages for participation across all years ranged from 51.6 to 53.8 years old.

Comorbidity scores

All subjects (participants and nonparticipants) were assessed using the Johns Hopkins Adjusted Clinical Groups Case-Mix System.[30] The scale measures aggregate morbidity levels ranging from 0–5 where a score of 0 indicates no illness or morbidity and a score of 5 indicates severe illness burden or maximum level of morbidity. For every year of the program, participants received higher morbidity scores relative to nonparticipants, suggesting that the program was used by individuals with significant disease burden. For all age ranges there were 42 instances of unadjusted statistical significance, which showed that program participants had higher levels of morbidity (Table 3).

Measures

Program participants were required to complete an HRA at program entry. Biometric measures were collected by various members of the staff, and included cholesterol and related subfractions (ie, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, triglycerides), fasting blood glucose, systolic and diastolic blood pressure, BMI, and waist circumference using standard clinical protocols.

Utilization and cost measures consisted of total health care expenditures, and expenditures were categorized by inpatient costs, number of inpatient admissions, length of stay (LOS) per admission, outpatient cost/claims, medical services, and pharmacy. Data are expressed in unadjusted dollars and utilization counts except where noted.

Results

Health care claims and costs

Table 4 examines 2009 health care costs and utilization among program participants versus nonparticipants. The data indicate that 2009 program participants had significantly lower total health care costs and costs by category with the exception of medical costs relative to nonparticipants. These medical costs include professional services (eg, office visits), laboratory, radiology, surgery/anesthesia, and other facility outpatient and miscellaneous services. Outpatient costs include outpatient surgeries, emergency room, mental health partial-day services, and diagnostic and treatment services. The outpatient costs reflect the facility portion of the encounter exclusive of professional services (ie, medical costs). The proportion of program participants having made any claim in 2009 was higher in total and within the medical claim category. However, nonparticipants were more likely to have made more claims in the cost categories of inpatient and pharmacy claims relative to participants. The average number of hospital admissions, total inpatient days, and average LOS per admission also were significantly higher for nonparticipants relative to participants.

Return on investment

Net savings were calculated by determining the difference between total health care costs for the nonparticipating cohort and the participating cohort after subtracting the program expenses from the actual health care cost savings by year for participants. Program expenses included staff salaries, employee benefits, employer taxes, state general excise taxes, medical supplies, office supplies and printing, postage and freight, utilities and telephone, furniture and equipment expense, occupancy expenses, professional services, advertising and promotions, insurance, travel, depreciation and amortization, data processing and software purchases, and

TABLE 2. PROGRAM PARTICIPANT DEMOGRAPHIC DATA 2002–2009

Sample Size	2002 N=15,328	2003 N=13,936	2004 N=14,878	2005 N=16,555	2006 N=18,903	2007 N=16,726	2008 N=15,239	2009 N=13,243
Sex								
Male	43.6%	43.1%	41.9%	41.7%	36.6%	35.3%	34.3%	33.6%
Female	56.4%	56.9%	58.1%	58.3%	63.4%	64.7%	65.7%	66.4%
Age Distribution								
18–29	3.4%	3.4%	4.0%	4.9%	6.3%	6.2%	6.1%	5.4%
30–39	11.6%	10.5%	10.7%	11.4%	13.8%	14.2%	13.3%	13.1%
40–49	25.1%	24.2%	23.8%	23.1%	23.1%	22.8%	22.4%	21.3%
50–59	30.7%	30.6%	31.3%	31.3%	28.8%	28.7%	28.0%	28.0%
60–69	15.9%	17.7%	17.6%	17.6%	18.0%	18.3%	19.3%	20.4%
70+	13.4%	13.6%	12.6%	11.7%	10.0%	9.8%	11.0%	11.7%
Mean Age	53.2	53.8	53.3	52.8	51.6	51.6	52.3	52.9

TABLE 3. MORBIDITY LEVELS BY AGE STRATIFICATION AND PARTICIPATION YEAR

	2002		2003		2004		2005	
	Nonparticipants	*Participants*	*Nonparticipants*	*Participants*	*Nonparticipants*	*Participants*	*Nonparticipants*	*Participants*
18–29	*1.24*	*1.46*	*1.26*	*1.57*	*1.27*	*1.57*	*1.30*	*1.57*
30–39	*1.50*	*1.69*	*1.51*	*1.67*	*1.51*	*1.74*	*1.52*	*1.70*
40–49	*1.84*	*2.00*	*1.87*	*2.03*	*1.87*	*2.01*	*1.89*	*2.07*
50–59	*2.22*	*2.35*	*2.25*	*2.37*	*2.27*	*2.41*	*2.29*	*2.45*
60–69	*2.62*	*2.76*	*2.65*	*2.79*	*2.67*	*2.81*	*2.70*	*2.82*
70+	*3.12*	*3.17*	*3.15*	*3.25*	*3.18*	*3.30*	*3.27*	*3.34*
Overall	*2.11*	*2.33*	*2.16*	*2.38*	*2.18*	*2.39*	*2.21*	*2.40*

	2006		2007		2008		2009	
	Nonparticipants	*Participants*	*Nonparticipants*	*Participants*	*Nonparticipants*	*Participants*	*Nonparticipants*	*Participants*
18–29	*1.66*	*1.80*	*1.66*	*1.86*	*1.72*	*1.91*	*1.84*	*2.05*
30–39	*1.90*	*1.97*	*1.89*	*1.98*	*1.95*	*2.05*	*2.07*	*2.18*
40–49	2.14	2.16	2.14	2.14	*2.20*	*2.25*	*2.34*	*2.40*
50–59	*2.48*	*2.54*	*2.47*	*2.53*	*2.55*	*2.59*	*2.70*	*2.76*
60–69	*2.85*	*2.93*	*2.84*	*2.90*	*2.89*	*2.98*	*3.02*	*3.10*
70+	3.45	3.44	3.49	3.47	3.51	3.49	3.57	3.56
Overall	*2.36*	*2.49*	*2.37*	*2.49*	*2.43*	*2.58*	*2.57*	*2.73*

Note: Morbidity levels are calculated for program participants and nonparticipants based on their status in the given year. Italicized text indicates a statistically significant difference between morbidity levels for each age group and overall for participants and nonparticipants for each year, $P < 0.05$.

general administrative expenses (human resources, legal, accounting, and administrative services). These expenses are displayed by year in Table 5.

Table 6 presents an ROI estimate for 2009 by analyzing the costs of wellness program members who did not use the program initially (2002–2008), but who then became partic-

ipants in 2009, relative to matched nonparticipants who never used the program during the 8 years. A subsample of nonparticipants was matched to participants on sex, age, morbidity score, and total health care costs in the previous 3 years (2006–2008) and no significant difference was found between participants and nonparticipants after matching

TABLE 4. 2009 PROGRAM PARTICIPANTS VS. NONPARTICIPANTS ON HEALTH CARE COSTS AND CLAIMS

Sample Size	2009 Participant N = 13,243	2009 Nonparticipant N = 299,096	Trend (Participant vs. Nonparticipant)
2009 health care costs			
Total costs	$2,426	$2,725	$P < 0.005$
Inpatient	$298	$509	$P < 0.005$
Outpatient	$299	$318	$P = 0.028$
Medical	$1,384	$1,306	$P < 0.005$
Rx (prescriptions)	$444	$592	$P < 0.005$
2009 health care claims			
Making any claim (%)	95.7%	88.5%	$P < 0.005$
Inpatient	2.6%	4.3%	$P < 0.005$
Outpatient	20.4%	20.8%	$P = 0.281$
Medical	94.9%	86.9%	$P < 0.005$
Rx	53.5%	54.5%	$P = 0.016$
2009 reported service counts (per 1000 members)			
Average number of service counts	36,754	35,313	$P = 0.001$
Inpatient	31	53	$P < 0.005$
Outpatient	432	428	$P = 0.134$
Medical	28,530	24,765	$P < 0.005$
Rx	7,761	10,067	$P < 0.005$
2009 average inpatient days (per 1000 members)	130	260	$P < 0.005$
2009 average Length of Stay per admission (for participants with at least one hospital admission, per 1000 members)	3,652	4,409	$P < 0.005$

Note: The comparison of 2009 total health care costs is derived from the data on those with and without claims; nonparticipants in 2009 have not participated in the wellness program in any of the 8 years.

Table 5. Program Costs per Participant

	Program Costs	Number of Participants	Costs per Participant
2002	$4,128,000	20,247	$204
2003	$4,250,000	19,442	$219
2004	$4,972,000	21,071	$236
2005	$5,004,000	23,352	$214
2006	$5,867,966	26,861	$218
2007	$5,784,864	24,494	$236
2008	$6,831,683	24,064	$284
2009	$6,671,203	23,142	$288

(P > .20). These matching variables also were set as covariates in the cost comparisons to better reveal the impact of the program. After matching and adjusting, the total costs in 2009 were estimated to be $3422 for nonparticipants and $2870 for participants, which generated a projected savings of $522 per participant. The net savings per participant was $264 after excluding the program cost. Total 2009 net savings based on this estimate was calculated at over $6 million.

Figure 1 displays the ROI by program year after statistically matching and adjusting for sex, age, baseline morbidity level, and baseline health care costs.[1] Matching variables included sex, age, morbidity score, and total health care costs in the previous 3 years prior to the index year (ie, a rolling 3-year matching window). The years 2002–2004 were matched by previous years that had available data. The figure indicates a positive ROI for each program year ranging from a low of 1.16:1 to a high of 2.83:1 with an average ROI across all 8 years of data of 2.02:1.

Cost and participation

The research team also examined the association between program participation and health care costs before and after utilizing the program within the same cohort. This was done by analyzing the claims trend for program members who initially did not use the program services in 2002 through 2007 (a 6-year cost "run-in" period) and then began participating in the index year of 2008 relative to a propensity score

Table 6. 2009 Estimated Costs and ROI from Trend Analysis

	Health Care Costs	HP Program Costs	Total Costs
2009			
Nonparticipants (n = 2161)	$3,422		$3,422
Participants (n = 2161)	$2,870	$288	$3,158
		Net Savings/ Participant	$264
	X	Participants #	23,142
	=	Total Net Savings	$6,109,488

2009 ROI = $Savings/$Costs = $552/$288 = $1.92:$1. The estimated net savings in 2002–2008 years are not presented here but are available from the author by request. HP, health plan; ROI, return on investment.

matched cohort of nonparticipants (matched one-to-one on sex, age, morbidity in 2006, and total health care cost in 2006) who never participated in the program from 2002 to 2009. The matching strategy yielded successful matches (ie, nonsignificant group differences for these variables, P > .40). Sex, age, and morbidity in 2006 also were used as covariates in the model to better equate the 2 groups. The trajectory depicted in Figure 2 suggests participants maintained an essentially flat cost trend (Zero Trend)[31] following program participation in 2008 while nonparticipants showed a significant and escalating increase on total costs year over year.

In order to further examine the program's cost impact based on level of engagement, the following analyses were based on the consistency of program participation operationalized as the total number of years a plan member participated in the program relative to total 2009 costs. Figures 3 and 4 present these data for all 2002–2009 members and for members who made 1 or more claims in 2009, respectively. In both cases the analyses demonstrate a strong linear association between the number of program participation years and lower total health care costs in 2009.

Finally, the research team decided to examine the relative value of engagement (ie, number of years of program participation) relative to morbidity values as measured by the Johns Hopkins Adjusted Clinical Groups Case-Mix System. Figure 5 plots average 2009 costs by morbidity values. The graph indicates greater spread with increased morbidity suggesting greater direct ROI as an interaction between baseline morbidity and degree of long-term engagement. The actual dollar values are presented in Table 7.

Discussion

The traditional medical establishment, with its effective emphasis on acute care, cannot be expected to solve the health care crisis alone. Nor will the evidence base for population health programming be built through the conduct of traditional randomized clinical trials.[32] However, an evidence base must be built nonetheless and such evidence certainly should include real-world demonstration of the sustainability of population health programming. As part of a logic map (Figure 6) that guides the development of a leading and lagging indicator framework, such evidence should address participation, behavior change, risk modification and economic outcomes. This study sought to contribute to building that evidence base by examining 8 years of economic and utilization data in a cohort of program participants and a matched set of nonparticipants. The findings here do not and are not intended to provide a comprehensive evaluation of each proof point in the map but rather to begin to fill in puzzle pieces of and support for portions of this logic map.

This study posed several questions of these data. *Who are program participants? How do they differ from nonparticipants?* Participants tended to be female; a finding consistent with a robust literature that supports the notion that women are more active consumers of health information.[33] Importantly, there are reasons to believe that gender-specific messaging may help with the participation of both sexes. Participants also tended to be older. However, given the power of the test in large samples such as those used here and the small mean arc differences, it is unlikely that this finding has meaningful
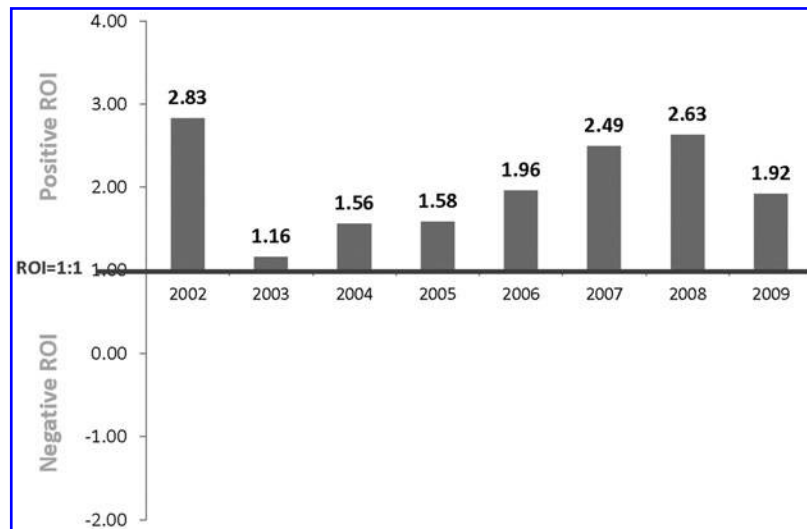
FIG. 1. Return on investment (ROI) by program year.

implications for how one might develop and/or implement program elements.

For every evaluation year of the program, participants received higher morbidity scores relative to nonparticipants, suggesting that the program was used by individuals with significant need. This runs contrary to the common argument that the "worried well" are overrepresented as participants in wellness programs. However, this finding is consistent with what is known about illness behavior[34] and the health belief model.[35] In other words, signs, symptoms, and reduced functionality often serve as motivators for seeking information, skills, treatment, and, ultimately, relief from the consequences of reduced health.

The research team also wanted to ask a couple of related questions such as *What are the estimated economic/cost benefits and ROI of the program over time? How does participant utilization and cost compare with that of nonparticipants?*

The findings presented here indicate a positive ROI for each program year with ratios ranging from a low of 1.16:1 to a high of 2.83:1. The average ROI collapsed across all 8 years was 2.02:1. The 2009 ROI approximated over $6 million in total savings. Further, data indicated that for a secondary cohort there was a strong positive change in health care expenditures (ie, bending the trend)[30] associated with the index year of participation relative to a "never participated" matched comparison cohort. The pattern of utilization also was different for the 2 cohorts. Participants were more likely to make a claim in any given year but nonparticipants had a disproportionate pattern of utilization in the cost categories of hospital admissions, average LOS, total annual inpatient days, and pharmacy. When one considers the overhead program costs that were included in the program cost adjustments (ie, salaries, benefits, employer taxes, state excise taxes, medical supplies, office supplies, printing, postage and freight, utilities and telephone, furniture and equipment expense, occupancy expenses, professional services, advertising/promotion, insurance, travel, depreciation and amortization, data processing, software purchases, general

FIG. 2. Total costs comparison between those who never participated in the program and those who started the program in 2008 and continued it in 2009. Note: Total health care costs include inpatient, outpatient, medical, and pharmacy costs. Members who had conflicting sex in multiple data sets, who reported being pregnant, or more than 365 total inpatient days in any 1 year were excluded. Participants and nonparticipants were matched on sex, age, morbidity in 2006, and total health care cost in 2006. Savings also were adjusted for sex, age, and morbidity in 2006. Cost differences between groups were nonsignificant in 2006–2007, $P > .50$; cost differences between groups in 2008 and 2009 were significant, $P < .02$, as tested by general linear model.
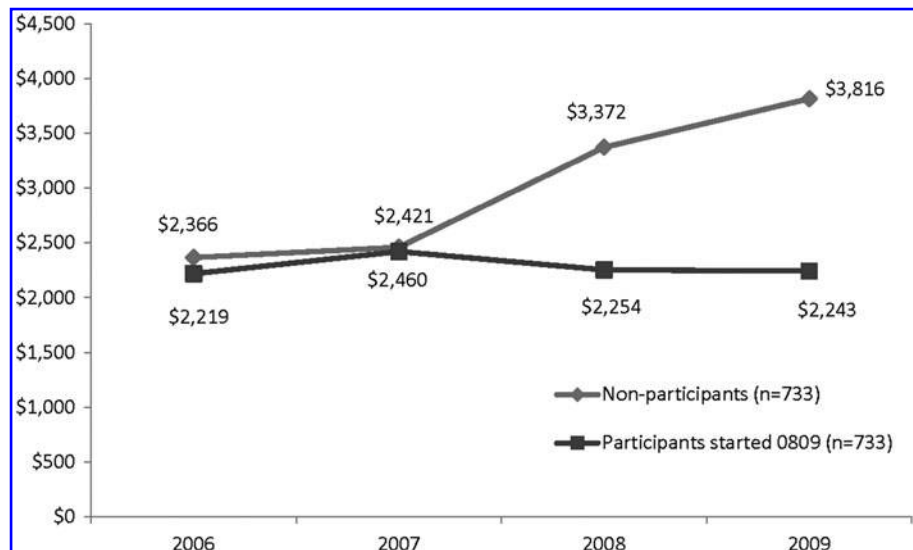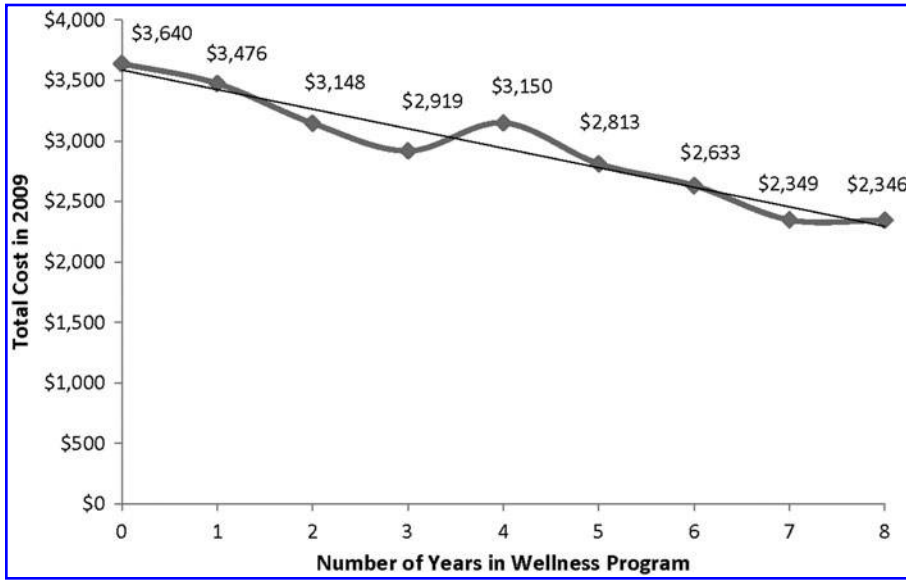
FIG. 3. 2009 total member care costs by number of years using the program. Note: Health care costs include inpatient, outpatient, medical, and pharmacy costs. Members who had conflicting sex in multiple data sets, who reported being pregnant, or more than 365 total inpatient days in any 1 year were excluded. Savings were adjusted for sex, age, and morbidity in 2002, and total health care cost in 2002. The 2009 health care costs were significantly different by number of participation year, $F(8, 172903) = 8.92$, $P < .0001$.

administrative expenses that include human resources, legal, accounting, and administrative services) this constitutes a fully burdened (conservative) ROI analysis.

In order to test whether or not participants dropped out when they became sick, 2 cohorts with 5 years of claims data were compared. The first 3 years were used as baseline data for matching purposes. Year 4 was the participation year. Year 5 was the year that differentiated the 2 groups into participants who dropped out after taking the program, and participants who were still in the program in Year 5.

Propensity score matching was used to match the 2 groups on sex, age, morbidity in the previous 3 years (Year 1–Year 3), and total cost in the previous 3 years (Year 1–Year 3). Sex, age, and morbidity in the previous 3 years also were used as covariates in the model. When 2005 was the participation year, there was not a significant difference in medical costs between the 2 groups in 2006 ($P = .367$). This same nonsignificant difference was found when the participation year

was 2006 ($P = .807$), 2007 ($P = .260$), and 2008 ($P = .930$). These findings suggest there was no significant difference between participants who dropped out versus those who remained in the wellness program.

In addition to health care claim costs, health-related productivity cost savings is well established[36–37] as an important component of ROI analyses. Although information on productivity was obtained from program participants, this information was not available for nonparticipants and comparative analyses could not be performed on health-related productivity cost savings. The health plan-focused nature of this project directed attention more closely to claims-related costs.

Finally, one last question posed was *whether there was a relationship between cost, utilization, and degree of program engagement.* There was a strong linear association between the number of program participation years and total health care costs, regardless of whether total costs across all members or total costs for a subcohort who had incurred 2009 costs were
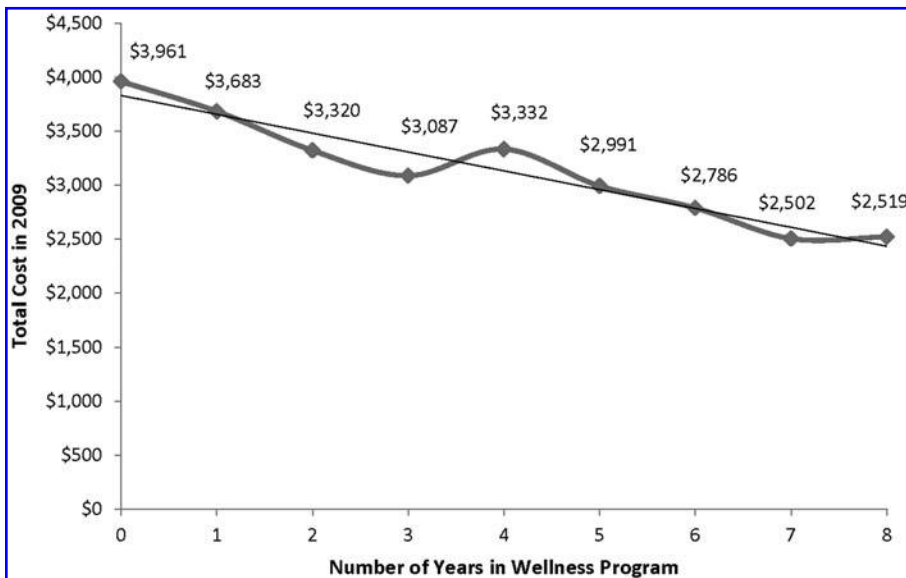

FIG. 4. 2009 total member care costs by number of years using the program, among members with 2009 costs > $0. Note: Health care costs include inpatient, outpatient, medical, and pharmacy costs. Data were derived from members with 2009 total health care costs > 0. Members who had conflicting sex in multiple data sets, who reported being pregnant, or more than 365 total inpatient days in any 1 year were excluded. Savings were adjusted for sex, age, morbidity in 2002, and total health care cost in 2002. The 2009 health care costs were significantly different by number of participation year, $F(8, 159993) = 11.52$, $P < .0001$.
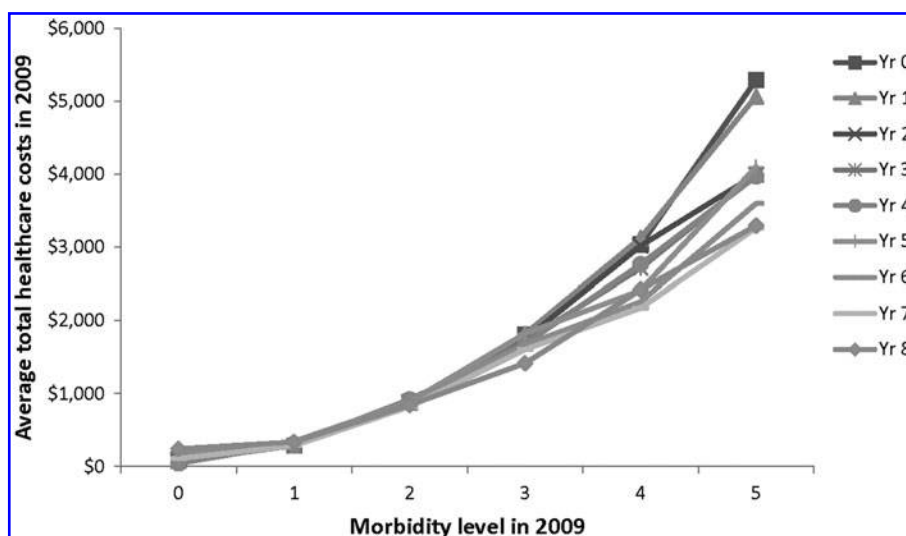
FIG. 5. 2009 health care costs (among members with 2009 costs > $0) on 2009 morbidity and number of years of wellness program participation. Note: Health care costs include inpatient, outpatient, medical, and pharmacy costs. Data were derived from members with 2009 total health care costs > 0. Members who had conflicting sex in multiple data sets, who reported being pregnant, or more than 365 total inpatient days in any 1 year were excluded. Savings were adjusted for sex, age, morbidity in 2002, and total health care cost in 2002. The 2009 health care costs were significantly different by number of participation year (F(8, 159912) = 2.67, P = .006), by 2009 morbidity levels (F(5, 159912) = 1411.36, P < .0001), and by the interaction of number of participation year and 2009 morbidity levels (F(40, 159912) = 3.41, P < .0001).

used in the analyses. The ROI value becomes more pronounced with higher levels of baseline morbidity.

This study, like all studies, has some notable limitations. First, the study was a retrospective cohort analysis without the benefit of random assignment to the treatment/comparison groups or even an a priori set of questions or hypotheses from the outset of data collection. As such, selection bias remains a primary potential threat to internal validity. Participants and nonparticipants did differ significantly on several key demographic variables that are worthy of methodological control, but these same constructs also may have implications for future program development and recruitment. However, in order to better equate groups, the research team matched cohorts statistically using propensity score matching[25–28] and further controlled group differences with covariate statistical adjustment. Because this analysis is taking into account the inherent self-selection bias in non-randomized participation in wellness programs, the research team believes that it represents a rigorous, albeit imperfect test of ROI given the limitations of this retrospective methodology.

Another limitation is a lack of specific health plan product data for each individual throughout the 8-year time period.

Therefore, analyses regarding the impact of shifting among plan designs, differences in plan payments, and members' out-of-pocket payments for the different health plan types could not be performed. There is also a relative paucity of non-claims-related data (ie, behavioral, psychographic) for the nonparticipants. In particular, the lack of psychographic and behavioral data limits the ability to detail how the 2 samples may have differed in terms of psychological, attitudinal, or behavioral constructs that are known to mediate health behavior, risk profiles, utilization, and claims patterns. In addition, biometric data were not collected from nonparticipants, so it was not possible to analyze the impact of program participation on potential biometric changes.

Finally, this study suffers from the "black box" problem in that one cannot attribute change to any specific program element, nor can one assure that the program was consistently applied within or across the years represented in this article. There are ways to address these problems in prospective studies where methodological approaches such as dismantling or statistical approaches such as structural equation modeling can tease apart the relative strength of the program's "active ingredients." However, these methods likely would provide little information on sustainability.

TABLE 7. 2009 HEALTH CARE COSTS BY MORBIDITY LEVEL AND DEGREE OF PRIOR PROGRAM ENGAGEMENT

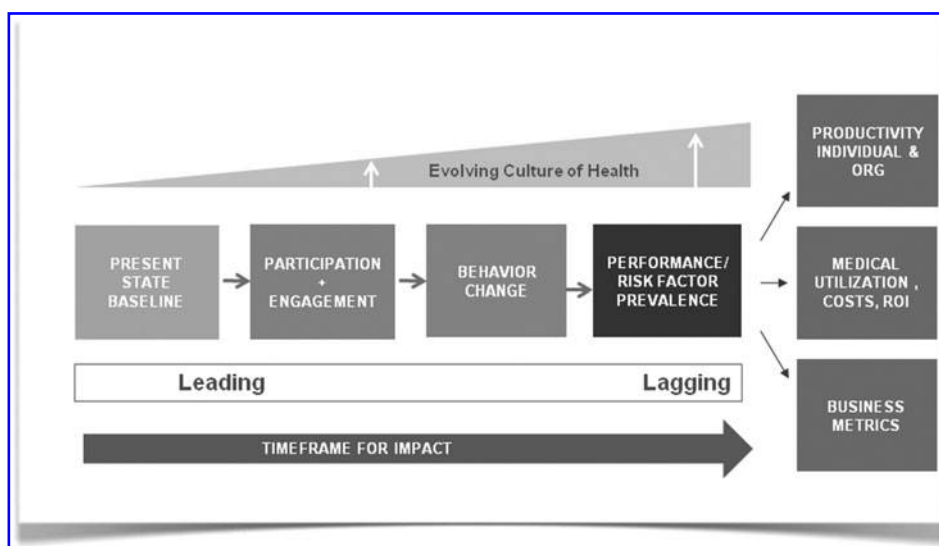| 2009 Morbidity	Number of Years Using Wellness Program								
	Yr 0	Yr 1	Yr 2	Yr 3	Yr 4	Yr 5	Yr 6	Yr 7	Yr 8
0	$12	$12	$13	$14	$9	$8	$32	$122	$57
1	$129	$166	$185	$172	$197	$220	$201	$220	$288
2	$714	$767	$802	$784	$871	$814	$831	$753	$805
3	$1666	$1705	$1641	$1656	$1471	$1758	$1559	$1529	$1311
4	$2750	$2902	$2778	$2477	$2486	$2265	$2109	$1962	$2178
5	$4659	$4577	$3569	$3605	$3418	$3417	$3285	$3025	$2966

FIG. 6.   Logic map.

Traditional randomized controlled trial methods are expensive and impractical for comprehensive outcomes study. Program evaluation methods[38] that start with functional logic maps, such as the one used here, coupled with sophisticated model building analytical tools are perhaps better suited to this task.

Despite these challenges and the documented variability in methodologies and analytical approaches in the literature, an evidence base continues to build with the general trend suggesting that when properly promoted and implemented, such population health-focused programming can result in both clinical and economic value. However, several significant and related challenges face population health management programming. The first is the magnitude of the problem. Health care faces an aging population, followed by a youth increasingly at risk for morbid obesity and all that entails, both biologically and behaviorally. Second is the relevance of behavior and behavior change technologies for optimizing health and well-being. Much of the developing risk is driven by modifiable psychological and behavioral habits that center largely on activity, diet, and varying forms of psychological distress. Unfortunately, traditional approaches to providing intervention to modify behavior and lifestyle do not scale sufficiently (there is a paucity of available expertise and resources at all levels of care to meet the demand). The challenge of scalability is the delivery of high-quality services to address the range of behavioral habits, problems, and attitudes in the service of health and well-being. Face-to-face and telephonic services will always have a place in the continuum of care but, given the growing demand, new and novel methods of service delivery must be developed that are economically sustainable and that address the needs of the most people possible.

High technologies in the form of Web interventions, telemedicine, mobile messaging, online health information, and mobile telemetry among others, have been cited as potential solutions to the challenges of both scale and sustainability.[39] They are occasionally offered as freestanding interventions but more often coupled with other services (eg, face-to-face counseling, home visits, telephonic support) and are part of a continuum of care approach. Such technology, thoughtfully applied, can foster greater participation in self-care across the health spectrum as the industry moves toward a patient-centered model of care. What remains to be seen is how best to use these mediums to deliver content that represents a true intervention as opposed to simply providing, via a novel medium, the same old health information and content that has been demonstrated to be necessary but insufficient to change health behavior.

Author Disclosure Statement

Drs. Schwartz and Mason, Ms. Wang, Ms. Pomana, Ms. Hyde-Nolan, and Mr. Carter are/were employees of Wellness & Prevention Inc. during the preparation of manuscript. They declared no conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Anderson DR, Whitmer RW, Goetzel RZ, et al. The relationship between modifiable health risks and group-level health care expenditures. Am J Health Promot 2000;15:45–52.
2. Goetzel RZ, Anderson DR, Whitmer RW, et al. The relationship between modifiable health risks and health care expenditures. An analysis of the multi-employer HERO health risk and cost database. J Occup Environ Med 1998; 40:843–854.
3. Goetzel RZ, Jacobson BH, Adldana SG, Vardell K, Yee L. Health care costs of worksite health promotion participants and non-participants. J Occup Environ Med 1998;40:341–346.
4. Hochart C, Lang M. Impact of a comprehensive worksite wellness program on health risk, utilization, and health care costs. Popul Health Manag 2011;14:111–116.
5. Jee SH, O'Donnell MP, Suh I, Kim IS. The relationship between modifiable health risks and future medical care expenditures: The Korea Medical Insurance Corporation (KMIC) Study. Am J Health Promot 2001;15:244–255.
6. Leigh JP, Hubert HB, Romano PS. Lifestyle risk factors predict healthcare costs in an aging cohort. Am J Prev Med 2005;29:379–387.

7. Long DA, Reed R, Lehman G. The cost of lifestyle health risks: obesity. J Occup Environ Med 2006;48:244–251.

8. Musich S, Hook D, Barnett T, Edington DW. The association between risk status and health care costs among membership of an Australian health plan. Health Promot Int 2003; 18:57–65.

9. Naydeck BL, Pearson JA, Ozminkowski RJ, Day BT, Goetzel RZ. The impact of the Highmark employee wellness programs on 4-year healthcare costs. J Occup Environ Med 2008;50:146–156.

10. Rust G, Strothers H, Miller WJ, McLaren S, Moore B, Sambamoorthi U. Economic impact of a Medicaid population health management program. Popul Health Manag 2011;14:215–222.

11. Schwartz SM, Day B, Wildenhaus K, Silberman A, Wang C, Silberman J. The impact of an online disease management program on medical costs among health plan members. Am J Health Promot 2012;25:126–133.

12. Schwartz SM, Ireland C, Strecher V, Nakao D, Wang C, Juarez J. The economic value of a wellness and disease prevention program. Popul Health Manag 2010;13:309–317.

13. Aldana SG. Financial impact of health promotion programs: a comprehensive review of the literature. Am J Health Promot 2001;15:296–320.

14. Grossmeier J, Terry PE, Anderson DR, Wright S. Financial impact of population health management programs: re-evaluating the literature. Popul Health Manag 2012;15:129–134.

15. Thorpe KE. The rise in health care spending and what to do about it. Health Aff (Millwood) 2005;24:1436–1445.

16. Pelletier KR. A review and analysis of the clinical and cost-effectiveness studies of comprehensive health promotion and disease management programs at the worksite: Update VI 2000–2004. J Occup Environ Med 2005;47:1051–1058.

17. Robertson J, Walkhom EJ, Henry DA. Health systems and sustainability: doctors and consumers differ on threats and solutions. PLoS One 2011;6(4):e19222.

18. Neville BH, Merrill RM, Kumpfer KL. Longitudinal outcomes of a comprehensive incentivized worksite wellness program. Eval Health Prof 2011;34:103–123.

19. Byrne DW, Goetzel RZ, McGown PW, et al. Seven-year trends in employee health habits from a comprehensive workplace health promotion program at Vanderbilt University. J Occup Environ Med 2011;53:1372–1381.

20. Mayer C, Williams B, Wagner EH, LoGerfo JP, Cheadle A, Phelan EA. Health care costs and participation in a community-based health promotion program for older adults. Prev Chronic Dis 2010;7(2):A38.

21. Dall TM, Roary M, Yang W, et al. Healthcare use and costs for participants in a diabetes disease management program, United States, 2007–2008. Prev Chronic Dis 2011;8(3):A53.

22. Greving JP, Visseren FL, de Wit GA, Algra A. Statin treatment for primary prevention of vascular disease: whom to treat? Cost-effectiveness analysis. BMJ 2001;342:d1672.

23. Lundy T. A paradigm to guide health promotion into the 21st century: the integral idea whose time has come. Glob Health Promot 2010;17:44–53.

24. Williams LC, Day BT. Medical cost savings for web-based wellness program participants from employers engaged in health promotion activities. Am J Health Promot 2011; 25:272–280.

25. Phillips JF. Using an ounce of prevention: does it reduce health care expenditures and reap pounds of profits? A study of the financial impact of wellness and health risk screening programs. J Health Care Finance 2009;36(2):1–12.

26. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41–55.

27. Heckman JJ, Ichimura H, Todd PE. Matching as an econometric evaluation estimator: evidence from evaluating a job training program. Rev Econ Stud 1997;64:605–654.

28. Heckman JJ, Ichimura H, Todd PE. Matching as an econometric evaluation estimator. Rev Econ Stud 1998;65:261–294.

29. Foster EM. Propensity score matching: an illustrative analysis of dose response. Med Care 2003;41:1183–1192.

30. Clark DO, Von Korff M, Saunders K, Baluch WM, Simon GE. A chronic disease score with empirically derived weights. Med Care 1995;33:783–795.

31. Edington DW. *Zero Trends: Health as a Serious Economic Strategy*. Ann Arbor, MI: Health Management Research Center, 2009.

32. Trask PC, Schwartz SM, Deaner SL, et al. Behavioral medicine: the challenge of integrating psychological and behavioral approaches into primary care. Eff Clin Pract 2002;5:75–83.

33. Cohen RA, Adams PF. *Use of the Internet for Health Information: United States, 2009*. NCHS data brief, no 66. Hyattsville, MD: National Center for Health Statistics, 2011.

34. Mechanic D. Social psychologic factors affecting the presentation of bodily complaints. N Engl J Med. 1972;286: 1132–1139.

35. Janz NK, Becker MH. The health belief model: a decade later. Health Educ Q 1984;11(1):1–47.

36. Loeppke R, Taitel M, Haufle V, Parry T, Kessler RC, Jinnett K. Health and productivity as a business strategy: a multiemployer study. J Occup Environ Med 2009;51:411–428.

37. Schultz AB, Chen CY, Edington DW. The cost and impact of health conditions on presenteeism to employers: a review of the literature. Pharmacoeconomics 2009;27:365–378.

38. McDavid JC, Hawthorn LR. Program Evaluation and Performance Measurement: An Introduction to Practice. Thousand Oaks, CA: Sage Publishing, 2006.

39. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of web-based vs. non-web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2006;6(4):e40.

Address correspondence to:
Shawn T. Mason, PhD
130 South 1st Street
Ann Arbor, MI 48104

E-mail: smason5@its.jnj.com

This article has been cited by:

1. Rebecca Masters, Elspeth Anwar, Brendan Collins, Richard Cookson, Simon Capewell. 2017. Return on investment of public health interventions: a systematic review. *Journal of Epidemiology and Community Health* **71**:8, 827-834. [Crossref]
2. Wolfgang Ahrens, K.-H. Jöckel. 2015. Der Nutzen großer Kohortenstudien für die Gesundheitsforschung am Beispiel der Nationalen Kohorte. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* **58**:8, 813-821. [Crossref]