What Can a Pilot Congestive Heart Failure Disease Management Program Tell Us about Likely Return on Investment?: A Case Study from a Program Offered to Federal Employees

CATHERINE J. vanVONNO, Ph.D.,[1] RONALD J. OZMINKOWSKI, Ph.D.,[2]
MARK W. SMITH, Ph.D.,[3] EILEEN G. THOMAS, R.N.,[4] DONIECE KELLEY,[5]
RON GOETZEL, Ph.D.,[2] GREGORY D. BERG, Ph.D.,[6] SUSHEEL K. JAIN, M.A.,[7]
and DAVID R. WALKER, Ph.D.[8]

ABSTRACT

In 1999, the Blue Cross and Blue Shield Federal Employee Program (FEP) implemented a pilot disease management program to manage congestive heart failure (CHF) among members. The purpose of this project was to estimate the financial return on investment in the pilot CHF program, prior to a full program rollout. A cohort of 457 participants from the state of Maryland was matched to a cohort of 803 nonparticipants from a neighboring state where the CHF program was not offered. Each cohort was followed for 12 months before the program began and 12 months afterward. The outcome measures of primary interest were the differences over time in medical care expenditures paid by FEP and by all payers. Independent variables included indicators of program participation, type of heart disease, comorbidity measures, and demographics. From the perspective of the funding organization (FEP), the estimated return on investment for the pilot CHF disease management program was a savings of $1.08 in medical expenditure for every dollar spent on the program. Adding savings to other payers as well, the return on investment was a savings of $1.15 in medical expenditures per dollar spent on the program. The amount of savings depended upon CHF risk levels. The value of a pilot initiative and evaluation is that lessons for larger-scale efforts can be learned prior to full-scale rollout. (Disease Management 2005;8:346-360)

INTRODUCTION

CHRONIC DISEASE is the leading cause of illness, disability and death in the United States, affecting nearly 100 million Americans.[1] Heart failure alone affects nearly 4.7 million Americans, with another 550,000 newly diagnosed cases each year.[2] Although heart failure

[1]Independent Consultant, Clifton, Virginia.
[2]Institute for Health and Productivity Studies, Cornell University, Health and Productivity Research, Thomson Medstat, Ann Arbor, Michigan.
[3]Health Economics Resource Center (HERC), VA Palo Alto Health Care System, Menlo Park, California.
[4]Population-Based Programs, BlueCross BlueShield Association, Washington, D.C.
[5]Internal Reviews, BlueCross BlueShield Association, Washington, D.C.
[6]McKesson Corporation, Broomfield, Colorado.
[7]Strategic Planning, ACS State Healthcare Solutions, Atlanta, Georgia.
[8]Health Services Research, CorSolutions, Inc., Buffalo Grove, Illinois.

can afflict people of all ages, congestive heart failure (CHF) mainly afflicts the elderly. CHF prevalence increases with age and approaches 10% of Americans over 80 years old.[3] Due to an aging population and increasing survival rates after diagnosis, the prevalence of heart failure is expected to increase two- to threefold in the coming decades.[4]

CHF directly and indirectly contributes to about 285,000 deaths per year.[2,5–8] It is the leading cause of hospitalization in the United States.[9] Despite treatment for CHF, readmissions to hospitals after discharge remain all too common, contributing to poor quality of life. As such, the immense morbidity and mortality associated with CHF increased medical costs for that condition to over $20 billion per year in the 1990s.[5,10]

Although the costs of CHF are high in terms of direct medical costs and quality of life, practical interventions exist for controlling and preventing many chronic conditions, including CHF. Reducing health risk factors is the most important part of any intervention aimed at reducing morbidity and mortality. Primarily, CHF interventions encourage smoking cessation; a low-fat, low-sodium diet; management of high cholesterol; exercise; and the use of effective medications such as angiotensin-converting enzyme (ACE) inhibitors, digoxin, and beta-blockers, all of which are currently underused in the treatment of heart failure.[11]

This paper describes the economic (financial) evaluation of a pilot CHF disease management program funded by the federal Office of Personnel Management for members enrolled in the BlueCross BlueShield Federal Employee Program (FEP). A pilot program was undertaken in one geographic area (Maryland) first, prior to full-scale rollout of the program. The value of a pilot approach is that it helped inform policy makers about likely return on investment prior to that rollout. Problems that may be ameliorated were identified early, and changes could be made to increase the likelihood that the full-scale program would perform well and lead to a substantial return on investment.

This study improves on previous evaluations of CHF programs in several important ways:

- To our knowledge, this study is the first evaluation of CHF patients at multiple risk levels; it is not limited to those at high-risk, thus having broader policy relevance.
- It estimates program impact vis-à-vis a comparison group, outside of a randomized experiment, which was not feasible here; and
- It utilizes statistical methods to control for regression to the mean and other secular trends that might otherwise impact the observed differences between the participant and control groups.

Program description

In June 1998, the BlueCross BlueShield Association contracted with McKesson Inc. and CorSolutions Inc. to administer a pilot program for the management of CHF patients enrolled as members in the FEP. McKesson managed a group of "at-risk" (ie, lower-risk) patients, those who had fewer than two hospitalizations for CHF in the previous year. CorSolutions provided its MULTIFIT Program exclusively to the remaining "high-risk" patients. Eligible FEP members in the state of Maryland were offered an opportunity to enroll as voluntary participants in the CHF program.

The pilot CHF disease management program was designed in accordance with clinical practice guidelines and administered to address risks related to diet, exercise, medication management, and smoking behaviors. The program used an individualized approach to member assessment and provided education, counseling, and monitoring of patient progress.

The CHF program aimed to reduce health risk factors among participants and was designed to provide education, monitoring and support for patients who chose to enroll. The program focused heavily on the monitoring and management of medications prescribed for the patient by his or her physician. Considerable attention also was devoted to the management of the participant's adherence to recommendations regarding appropriate self-care practices. Such practices are designed to limit complications arising from improper home management of CHF. Similar programs have been shown to reduce costs in other applications.[12]

METHODS

Study design

A cohort of CHF program participants in Maryland who were enrolled in the program for at least 3 months was matched to a cohort of non-participants from Virginia, where the CHF program was not offered. Both groups were otherwise covered by the same FEP health plan. Participants within each cohort were assessed for 12 months before the CHF program began (February 1, 1999), and 12 months afterward. The average length of enrollment in the CHF program was 11.3 months overall. Enrollment did not vary much by risk; at-risk patients were enrolled for an average of 11.2 months, and high-risk participants were enrolled for an average of 11.7 months.

Study sample

Study subjects in Maryland were selected as potential program participants based on their claims data, which indicated diagnosis and treatment codes for CHF. A comparison group of non-participants who also had evidence of CHF in their claims histories was drawn from the FEP's membership in northern Virginia, where the program was not offered. Medical records were not available for either group so other evidence of CHF, such as low ejection fraction, or New York Heart Association Class III or IV measures, could not be used to find CHF patients.

To qualify for inclusion in the study, individuals had to meet the following criteria:

- *For participants, successful completion of the enrollment process by telephone.* This entailed making contact with the patient, confirming CHF status, and securing the patient's agreement to participate.
- *Availability of claims data for a 24-month period. For participants, the period was defined by the 12 months prior to enrollment in the program and the 12 months following enrollment.* For a comparison group of non-participants, a proxy enrollment date was determined (using the enrollment date for the participants to which they were matched),

and the same 24-month criterion was applied.

Enrollment and participation

A total of 3,843 plan members in Maryland were referred to the program. Of these, 68% ($n = 2,614$) were able to be contacted by telephone, after several attempts to reach them were made at different times of the day. Of those contacted, 1,429 or 55% enrolled in the program. Of those enrolled, 53% ($n = 760$) completed intake assessments for the voluntary program. Finally, a total of 457 CHF patients completed at least three months of the program (ie, 91 days) following enrollment, and met other inclusion criteria noted below. The financial performance of the FEP CHF program was assessed on the basis of the experience of this group of 457 program participants. A comparison group of non-participants also was selected, as described below.

A number of inclusion restrictions were applied to CHF program participants before the analyses were carried out. The goal was to allow more accurate estimation of the impact of the CHF management program. On the basis of advice from clinicians, claims for patients with certain conditions were excluded from analysis. The guiding principle was a desire to eliminate the impact of high-cost illnesses whose progression would not reasonably be affected by the CHF program. Thus, patients with claims related to transplants, HIV/AIDS, and end-stage renal disease were excluded from the analysis due to very high condition-related payments. (This criterion affected only one patient in the sample, however.) Patients with claims related to cancer were excluded, but only if cancer-related payments exceeded $5,000 per year. This cutoff was chosen after a review of cancer claims revealed that only a small number of individuals ($n = 9$) had cancer-related spending above the $5,000 mark. Patients whose primary payer was Medicare at any time during the 24-month study window ($n = 66$) were excluded, due to the likelihood that the FEP would not capture all claims for those individuals. Finally, we excluded four patients who died during the first year of the CHF program. A total of 80 program participants were

dropped from the study because they met one or more of these exclusion criteria.

Selected claims were dropped for other individuals as well. All claims for poisoning by substances other than food or drugs were excluded, as were trauma-related claims for eight program participants whose traumas appeared to be unrelated to CHF.

The same inclusion criteria and claims-handling processes were used for nonparticipants. (Additional information about the selection of the comparison group is provided below.) Since this study used a case-control design, applying the same inclusion and exclusion criteria to both groups of patients helped to avoid bias when return on investment (ROI) was estimated for the CHF program.

Choosing a comparison group

The Diagnostic Cost Grouper (DCG) risk-adjustment software package was used to categorize all program participants and potential comparison group members into one or more of 30 possible Aggregated Condition Categories (ACCs).[13] ACC assignment was based on patients' ICD-9-CM diagnosis codes noted during the 12-month pre-intervention period. Treatment and comparison group members with the same combination of the eight most frequently found ACCs and gender were matched. Using this method, 480 non-participants were matched to participants.

Several of the 457 program participants could not be matched to any nonparticipant using the ACC and gender method. Thus, a second matching process was used. We identified the set of unique ICD-9-CM diagnoses assigned to each CHF program participant during the 12-month pre-intervention period. The diagnosis codes were ranked according to frequency and the five most common were chosen. We then linked participants with nonparticipants who had the same combination of ICD-9-CM codes and gender. A total of 323 nonparticipants were matched to program participants in this manner. The final sample size of comparison group members was 803.

After carrying out these procedures, there were still 136 CHF program participants who were not matched to anyone in the other group.

The 136 participants were left in the main analysis, however, even though they did not match any nonparticipants, in order to maintain a large enough sample for meaningful analysis. Additional clinical, demographic, and other variables were used in the multivariate analyses described below to account for remaining measurable differences between participants and non-participants. Later, we conducted a sensitivity analysis to see how results would change if these 136 patients were dropped.

Analysis

Descriptive analyses. As described above, the matching process used to link participants and nonparticipants was not perfect, and some differences between CHF program participants and nonparticipants were expected. To assess remaining differences between the two groups, chi-square tests of independence were conducted to compare the 457 treatment group members to the 803 comparison group members on the basis of demographics (age and gender), type of heart disease observed in the preprogram period, comorbidities, medications taken during that period, and the types of healthcare services used before the CHF program began.

The list of variables selected for comparison was chosen after consultation with physicians to determine which factors were likely to influence utilization trends for CHF patients. The heart disease measures included indicators for whether patients had rheumatic heart disease, hypertensive heart disease, ischemic heart disease, or pulmonary heart disease.

To address comorbidity differences between groups, we counted the number of unique ICD-9-CM diagnosis codes observed during the year prior to enrollment in the program. The number and its square were included in our statistical models. (Including the squared value allowed us to control for any non-linear relationships between the number of illnesses and expenditures.) We also included an indicator for endocrine disorders because the DCG software indicated these to be a problem whose prevalence differed substantially between program participants and nonparticipants.

The list of medications we considered in our

analyses included those commonly prescribed for CHF plus additional medications in the same categories listed in the 1999 Drug Topics Red Book.[14] These included ACE inhibitors, digoxin, diuretics, cardiac glycosides, calcium channel blockers, hypotensives, vasodilators, and other cardiac agents (eg, beta-blockers).

The list of healthcare services we considered during the preprogram period included the use of any nursing home care, any home care, any CHF-related inpatient care, and any CHF-related outpatient office visits. Using these services in the preprogram period was expected to be associated with higher treatment costs.

Finally, descriptive analyses were used to help interpret the ROI findings that were obtained from the multivariate analyses described below. For example, if the multivariate analyses showed a beneficial change in medical expenditures due to program participation, descriptive studies could be advanced to help find the source of those savings. To address this issue, we also conducted t-tests to see whether changes in payments for inpatient, outpatient, and drug services may have occurred at different rates for program participants and nonparticipants.

Multivariate analyses. The demographic, heart disease, comorbidity, medication use, and service use indicators mentioned above also were used as independent variables in multiple regression analyses to account for their impact on healthcare expenditures before the impact of the CHF program was estimated. We also controlled for the impact of early disenrollment from the program (ie, dropping out after three months of participation, but before 12 months of participation). Using the regression analyses helped to adjust for any imperfections in the matching process that were related to demographics, type of heart disease, comorbidities, medication patterns, and previous service use.

Separate regression analyses were conducted for high-risk and low-risk patients, using the same control group for each. This was done because a Chow test showed that the demographics, heart disease, comorbidity, medication use, and service use variables had a different impact on expenditures for each risk group ($p = 0.0001$). Thus, combining low- and high-risk patients into the same analysis would have distorted the program impact estimates, leading to biased estimates of return on investment.

The dependent variable for each regression analysis was defined as healthcare expenditures during the 12-month period after the CHF program began, minus expenditures during the 12-month period before it began. For the first set of analyses, the dependent variable was defined in terms of payments for these expenditures that were made by the FEP. For a second set of analyses, the dependent variable considered all expenditures, regardless of payer.

The major independent variable for each regression analysis was an indicator denoting whether patients participated in the at-risk or high-risk CHF programs administered by McKesson and CorSolutions. This indicator was included in addition to the demographic, heart disease, service use, medication, and comorbidity measures noted above. The regression coefficients for the indicator of program participation showed whether trends in expenditures over time were higher or lower for program participants as compared to the matched group of nonparticipants, after controlling for the potentially confounding factors mentioned above.

Calculating return on investment. All payments for medical care were cast in year 2001 dollars to adjust for inflation before the regression analyses were completed. An index based upon the Gross Domestic Product Implicit Price Deflator was used for the inflation adjustment, as suggested by Getzen.[15]

After the regression analyses were conducted, we calculated the ROI and the net present value (NPV) for the pilot CHF program. The ROI figure is the estimated number of dollars saved by the program, cast as a ratio per dollar paid to the program vendors. The NPV figure included the total program savings minus those vendor payments. If the CHF program was cost beneficial (ie, if benefits exceeded costs), the NPV would be greater than zero and the ROI ratio would be greater than 1.0.

Because all program participants and nonparticipants were followed for only 1 year af-

ter the CHF program began, the benefits and costs of participation were not discounted to adjust for the changing value of a dollar over time. Only the inflation adjustment noted above was made. The time horizon was short enough to preclude the need to discount program costs and benefits.

RESULTS

Sample characteristics

Table 1 compares baseline characteristics of CHF program participants and comparison group members across several dimensions. We found differences in age group, primarily because the at-risk group varied from the comparison group in the age distribution. Some differences in type of heart disease and pre-period drug use were observed as well, with these more likely to be due to differences between high-risk participants versus comparison group

members. The high-risk participants also were more likely to use home care or to have CHF-related inpatient care in the preprogram period.

Multivariate analyses

Multiple regression analyses were used to control for demographic and other potential confounding variables that could not easily be controlled by the matching process. By controlling for these potentially confounding factors, regression analyses provided a more accurate estimate of program impact, accounting for the differences found in Table 1 that remained after matching.

Separate regressions were estimated for at-risk and high-risk participants and the comparison group members. There were 339 at-risk participants and 118 high-risk participants. Tables 2 and 3 report regression results, noting the change over time in payments for these groups relative to the comparison group mem-

TABLE 1.   SAMPLE CHARACTERISTICS DURING YEAR BEFORE CONGESTIVE HEART FAILURE (CHF) PROGRAM BEGAN

Variable	Control group mean or percent (n = 803)	High-risk treatment group mean or percent (n = 118)	Chi-square test for comparison to control group p-*value*	Low-risk treatment group mean or percent (n = 339)	Chi-square test for comparison to control group p-*value*	Combined treatment group mean or percent (n = 457)	Chi-square test for comparison to control group p-*value*
Age <30	1.50%	0.00%	0.22	0.00%	0.04	0.00%	<0.01
Age 30–49	13.20%	9.32%	0.24	5.90%	<0.01	6.78%	<0.01
Age 50–69	62.39%	65.25%	0.55	82.60%	<0.01	79.00%	<0.01
Age >69	22.91%	23.73%	0.84	10.91%	<0.01	14.22%	<0.01
Female	45.85%	59.32%	0.01	48.08%	0.49	50.98%	0.08
Has rheumatic heart disease	7.35%	14.41%	0.01	7.08%	0.87	8.97%	0.31
Has hypertensive heart disease	56.66%	77.97%	<0.01	69.32%	<0.01	71.55%	<0.01
Has ischemic heart disease	30.76%	75.42%	<0.01	63.72%	<0.01	66.74%	<0.01
Has pulmonary heart disease	4.23%	11.86%	<0.01	3.54%	0.59	5.69%	0.25
Has endocrine disorder	16.81%	22.88%	0.11	15.34%	0.54	17.29%	0.83
Used ACE inhibitor	25.53%	41.53%	<0.01	31.27%	0.05	33.82%	<0.01
Used digoxin	19.18%	31.36%	<0.01	25.96%	0.01	27.35%	<0.01
Used diuretics	28.77%	64.41%	<0.01	43.66%	<0.01	49.02%	<0.01
Used cardiac glycosides, calcium channel blockers, hypotensives, vasodilators, and other cardiac agents	13.70%	23.73%	<0.01	13.27%	0.85	15.97%	0.27
Used nursing home	4.98%	9.32%	0.05	0.88%	<0.01	3.06%	0.11
Used home care	20.05%	40.68%	<0.01	15.63%	0.08	22.10%	0.39
Had CHF-related inpatient stay	25.90%	57.63%	<0.01	23.01%	0.30	31.95%	0.02
Had CHF-related outpatient visit	73.85%	89.83%	<0.01	91.74%	<0.01	91.25%	<0.01
No. of diagnosis codes	16.91	21.40	<0.01	15.07	<0.01	16.71	0.74

1998–2000 FEP Data.

TABLE 2.   REGRESSION ANALYSIS OF DIFFERENCE (POST-PRE) IN TOTAL PAYMENTS
FOR AT-RISK PARTICIPANTS AND COMPARISON GROUP MEMBERS

	All payers			FEP only		
Independent variable	Coefficient	SE	p-value	Coefficient	SE	p-value
Treatment programs						
AHG program participant	−1112.16	1408.50	0.43	−999.34	1294.89	0.44
Early disenrollment: number of months[a]	1368.67	462.76	0.01	1428.18	425.43	<0.01
Demographics						
Age 30–49	−1085.65	1888.02	0.57	−1132.67	1735.73	0.51
Age 70 or more	1189.11	1541.55	0.44	640.64	1417.21	0.65
Female	−1784.05	1198.80	0.14	−1798.09	1102.10	0.10
Pre-program diagnoses						
Rheumatic heart disease	−3991.17	2239.12	0.07	−3550.76	2058.51	0.08
Hypertensive heart disease	−1101.48	1320.74	0.40	−831.41	1214.21	0.49
Ischemic heart disease	−3731.95	1336.71	0.01	−3697.01	1228.89	<0.01
Pulmonary heart disease	−6667.53	3021.65	0.03	−6113.80	2777.92	0.03
Endocrine disorders	−4011.96	1592.93	0.01	−3684.70	1464.44	0.01
Number of unique 3-digit ICD-9 codes	167.54	166.01	0.31	117.12	152.62	0.44
Square of number of unique ICD-9 codes	20.60	3.35	<0.01	19.84	3.08	<0.01
Pre-program medication use						
Any ACE inhibitor	−491.76	1343.93	0.71	−479.50	1235.53	0.70
Any digoxin	−2516.86	1536.43	0.10	−2137.62	1412.50	0.13
Any diuretic	−823.25	1328.84	0.54	−950.28	1221.65	0.44
Pre-program CHF-related service use						
Any nursing home care	−7115.09	3179.57	0.03	−4453.13	2923.11	0.13
Any home care	−9405.46	1604.54	<0.01	−8076.48	1475.12	<0.01
Any CHF-related inpatient admission	−1652.62	150.34	<0.01	−1480.56	138.21	<0.01
Any CHF-related outpatient office visit	−3643.31	1633.07	0.03	−3232.00	1501.35	0.03
Intercept	5195.31	2050.53	0.01	5058.27	1885.14	<0.01
R squared	0.38			0.38		

[a]Equal to (12 − [no. of months enrolled]); range 0–9.
1998–2000 FEP Data.

bers. Table 2 shows that a typical at-risk participant enrolled in the CHF program saved the FEP about $999.34 per year, but this result was not statistically significant ($p = 0.44$). Once all other payers were considered, savings increased slightly to $1,112.16 per patient per year, but the results still were not statistically significant ($p = 0.43$).

Table 3 shows that a typical high-risk participant enrolled in the CHF program saved the FEP health plan $4,581.22 ($p = 0.04$). Results for all payers suggest a savings of $4,754 per patient ($p = 0.05$).

Regardless of payer, a number of other variables were significant in the expenditure regression models. Trends in cost over time were significantly lower for at-risk patients who had diagnoses for rheumatic, ischemic, or pulmonary heart disease in the pre periods or endocrine disorders during that period, compared to others without these diseases (Table 2).

Cost increases also were lower for high-risk patients who had rheumatic, ischemic, and pulmonary heart disease in the pre period as compared to patients without these disorders (Table 3).

Finally, we found greater savings associated with patients who had received more intensive care prior to enrollment, as evidenced by claims for nursing homes (in three of the four regression models), home care, or a CHF-related hospitalization. The use of major heart medications (ACE inhibitors, dioxins, or di-

TABLE 3. REGRESSION ANALYSIS OF DIFFERENCE (POST-PRE) IN TOTAL PAYMENTS FOR HIGH-RISK PARTICIPANTS AND COMPARISON GROUP MEMBERS						
	All payers			FEP only		
Independent variable	Coefficient	SE	p-value	Coefficient	SE	p-value
Treatment programs						
CorSolutions program participant	−4754.50	2426.22	0.05	−4581.22	2250.42	0.04
Early disenrollment: number of months[a]	3047.34	911.56	<0.01	2849.84	845.51	<0.01
Demographics						
Age 30–49	−1729.90	2212.04	0.43	−1677.81	2051.76	0.41
Age 70 or more	1707.40	1803.24	0.34	1355.29	1672.58	0.42
Female	−2030.50	1503.49	0.18	−2108.05	1394.54	0.13
Pre-program diagnoses						
Rheumatic heart disease	−8934.26	2678.35	0.00	−8053.02	2484.28	
Hypertensive heart disease	−1778.14	1668.01	0.29	−1732.76	1547.15	0.26
Ischemic heart disease	−4806.97	1732.55	0.01	−4446.46	1607.01	
Pulmonary heart disease	−7109.98	3397.01	0.04	−6361.15	3150.87	0.04
Endocrine disorders	−2447.50	1950.42	0.21	−2118.52	1809.10	0.24
Number of unique 3-digit ICD-9 codes	344.74	201.04	0.09	288.44	186.48	0.12
Square of number of unique ICD-9 codes	15.91	3.85	<0.01	15.40	3.57	<0.01
Pre-program medication use						
Any ACE inhibitor	−876.60	1710.61	0.61	−697.86	1586.66	0.66
Any digoxin	−2713.74	1946.12	0.16	−2347.75	1805.11	0.19
Any diuretic	−2175.49	1711.53	0.20	−2151.32	1587.52	0.18
Pre-program CHF-related service use						
Any nursing home care	−7869.81	3303.19	0.02	−6966.49	3063.84	0.02
Any home care	−8994.17	1891.39	<0.01	−7907.92	1754.34	<0.01
Any CHF-related inpatient admission	−1712.38	182.59	<0.01	−1552.01	169.36	<0.01
Any CHF-related outpatient office visit	−1651.17	1960.22	0.40	−1404.81	1818.18	0.44
Intercept	4160.34	2490.69	0.10	4188.23	2310.21	0.07
R squared	0.39			0.38		

[a]Equal to (12 − [no. of months enrolled]); range 0–9.
1998–2000 FEP Data.

uretics) in the pre period was not associated with differences in payments over time for either patient group.

Return on investment estimates

The regression coefficients from Tables 2 and 3 were multiplied by the associated number of at-risk and high-risk program participant-years to estimate the total dollars saved by the CHF program. These savings were then compared to total program expenditures to estimate ROI and NPV. The results are reported in Table 4.

Overall, Table 4 shows that the CHF program saved money. The FEP invested about $1,706 for each of the 457 patients enrolled in the program and saved $1,841 per enrollee. In total, the FEP saved about $61,728 on the pilot CHF program, resulting in an ROI of $1.08 in medical care expenditure savings for every dollar invested in the program. Accounting for all payments regardless of payer (ie, including copayments, deductibles, and coordination-of-benefit payments from other insurers, thereby invoking a broader perspective), the ROI estimate increased to $1.15 in savings per dollar spent on the program.

We also found that the ROI varied according to patient risk level (Table 4). Payments for each of the 118 patients enrolled in the high-risk program were $4,697 whereas savings were $4,581, a lower amount. Thus, this program led to a modest loss of about $27,746 for the FEP. The ROI was $0.95, a loss of about $0.05 per dollar spent on the program. The loss was about $7,834 when all payers were considered, re-

TABLE 4.   NET PRESENT VALUE AND RETURN ON INVESTMENT IN CONGESTIVE
HEART FAILURE (CHF) PROGRAM, IN YEAR 2001 DOLLARS

	Payments by FEP		Payments by all payers	
	Net present value	Return on $1.00 investment	Net present value	Return on $1.00 investment
A. All participants (*n* = 457)				
Average savings in all payments (pre − post)	$841,499.92		$896,979.75	
Total program expenses	$779,771.88		$779,771.88	
Total program benefit (net present value)	$61,728.04	$1.08	$117,207.87	$1.15
B. High-risk participants (*n* = 118)				
Average savings in all payments (pre − post)	$526,458.73		$546,371.18	
Total program expenses	$554,205.10		$554,205.10	
Total program benefit (net present value)	−$27,746.37	$0.95	−$7,833.92	$0.99
C. Low-risk participants (*n* = 339)				
Average savings in all payments (pre − post)	$315,041.19		$350,608.57	
Total program expenses	$225,566.78		$225,566.78	
Total program benefit (net present value)	$89,474.41	$1.40	$125,041.79	$1.55

Savings estimates are based upon multivariate analyses, adjusted for length of program participation.
FEP, Federal Employee Program.

sulting in an ROI estimate of $0.99 (ie, a loss of a penny per dollar spent on the program).

For the 339 at-risk patients, program spending was $665 and program savings were $999–$1,112 per patient, thus producing an ROI of between $1.40 and $1.55 per dollar spent on the program.

Changes in service use and medications

The regression analyses presented in Tables 2 and 3 and the NPV and ROI figures presented in Table 4 show that the CHF program seemed to be cost beneficial for at-risk patients, and was roughly a breakeven endeavor for high-risk patients. However, these tables do not explain why that may be the case. Data presented in Table 5 provide some clues as to the possible sources of program savings. Table 5 shows differences in inpatient, outpatient, and pharmaceutical payments over time for the year before and the year after the CHF program began, comparing program participants to nonparticipants. These differences were not adjusted via regression analysis, and are therefore more speculative, but still may provide a rough perspective on why the ROI values noted above were obtained. The top three rows in the table show that av-

erage medical (inpatient and outpatient) payments decreased over time for program participants, while these payments increased over time for nonparticipants. These unadjusted expenditure trends were significantly different for participants vs. nonparticipants (t-test p-values < 0.05).

The next several rows of the table show unadjusted changes over time in payments for various types of pharmaceuticals. Only a few of these were significant and, in dollar terms, they tended to show lower pharmaceutical expenditures for nonparticipants. For example, total CHF-related drug payments increased by about $208 on average for program participants, but increased only about $155 on average for nonparticipants. This is due primarily to a larger increase in payments for ACE inhibitors for program participants. This may reflect improved quality of care, however, because ACE inhibitors are recommended for all CHF patients. There was also a slightly higher increase in drug payments due to the use of vasodilators among program participants (ie, expenditures for vasodilators increased by $29.56 for participants on average, but only by $19.83 on average for nonparticipants).

Taken together, the results noted in Table 5 suggest that declines in inpatient and outpa-

TABLE 5. UNADJUSTED HEALTHCARE PAYMENTS, BY TREATMENT STATUS AND PERIOD, IN YEAR 2001 DOLLARS

| | CHF program group (n = 457) | | | | | All matched nonparticipants (n = 803) | | | | |
| | Pre-program | | Post program | | | Pre-program | | Post program | | |
	Mean	SD	Mean	SD	Difference	Mean	SD	Mean	SD	Difference
Healthcare services										
Total CHF-related	7,172.19	15,167.39	4,220.20	10,535.80	−2951.99	4,970.71	11,435.04	6,072.92	14,160.85	1,102.21[c]
Total inpatient	5,859.36	13,808.61	3,292.75	9,443.31	−2,566.60	4,466.18	10,262.99	5,471.27	14,586.73	1,005.09[c]
Total outpatient	4,734.29	6,622.34	3,729.51	5,055.52	−1,004.78	5,613.96	18,871.78	5,912.68	18,804.77	298.73[c]
Medications										
Total	2,041.63	1,746.42	3,329.83	3,057.67	1,288.20	1,035.22	1,472.68	2,412.83	2,900.36	1,377.62
Total non-CHF	1,870.91	1,714.82	2,950.77	2,989.55	1,079.86	882.77	1,356.34	2,105.05	2,796.15	1,222.28
Total CHF-related	170.72	258.64	379.06	400.06	208.34	152.44	260.23	307.78	381.64	155.34[c]
ACE-inhibitors	63.39	124.77	154.39	204.90	91.01	57.65	143.50	117.60	214.84	59.95[c]
Cardiac glycosides	10.47	20.32	14.30	27.47	3.83	7.64	25.16	12.63	36.83	4.98
Calcium-channel blockers	1.17	19.44	4.47	41.21	3.30	1.73	29.14	6.06	54.67	4.33
Hypotensives	15.17	84.67	36.84	143.21	21.67	19.94	96.48	33.01	134.02	13.07
Vasodilators	26.93	87.50	56.49	142.91	29.56	15.46	69.47	35.28	126.67	19.83[b]
Diuretics	26.31	61.50	49.87	111.69	23.56	24.54	69.84	44.29	97.50	19.74
Other cardiac agents	27.30	101.45	62.70	157.68	35.40	25.48	99.38	58.90	157.83	33.43
Total payments										
CHF-related	7,342.90	15,189.20	4,599.26	10,582.35	−2,743.65	5,123.16	11,448.44	6,380.71	14,167.62	1,257.55[c]
All	12,635.27	18,927.88	10,352.09	13,688.10	−2,283.19	11,115.35	23,291.29	13,796.79	27,215.52	2,681.44[c]

[a]Matched to CHF program participants by ACC/gender or ICD-9/gender.
[b]Significantly different from the treatment group, $p < 0.10$.
[c]Significantly different from the treatment group, $p < 0.05$.
The values in this table have not been adjusted via multiple regression, so they provide a crude estimate of changes over time.
CHF, congestive heart failure.

tient expenditures that may be due to program participation were not offset by increased drug use; these changes may have led to the positive ROI and NPV figures reported above.

Sensitivity analyses

We conducted sensitivity analyses to learn how the program impact estimates would change if we dropped the 136 CHF patients who could not be matched to comparison group members. Excluding these 136 program participants had a mixed impact on the results.

After dropping 87 unmatched participants who were in the at-risk group, the estimated savings to the FEP increased to $1,011.29 per person per year ($p = 0.49$); this compares to the estimated savings of $999.34 ($p = 0.44$) that was noted in the main analysis that did not exclude these 87 patients. The roughly $11 difference in these two impact estimates would have had a very minor impact on ROI for the at-risk group. In contrast, dropping these 87 unmatched participants from the analysis of expenditures by all payers (not just FEP) led to a substantially greater savings estimate of $1,235.96 per person per year, compared to the $1,112 figure obtained prior to dropping these participants. However, the results still were not statistically significant ($p = 0.44$).

Dropping the high-risk participants who could not be matched to any comparison group members had a large impact on the results as well. After dropping those 49 patients, the estimated savings to the FEP decreased from $4,581.22 ($p = 0.04$) to $4,191.73 per person per year ($p = 0.15$). This reduced the ROI for the high-risk group by roughly 8.5%. When payments by all payers were considered, the new savings estimate was $4,509.86 per person per year ($p = 0.15$), compared to the $4,754.50 estimate obtained prior to dropping the unmatched participants. Thus, the ROI estimate dropped by about 5.1%.

In summary, dropping the unmatched high-risk participants reduced savings and led to statistically insignificant findings. The lack of significance should not be surprising, however, because only 69 high-risk participants were left after dropping the 49 who could not be matched to comparison group members.

DISCUSSION

Medical claims data were analyzed to find trends in medical expenditures for CHF patients who participated in a pilot disease management program offered to federal employees. Most program participants were matched to a comparison group of patients in a neighboring state who had identical health insurance plan features, but who were not offered enrollment in the CHF program. Regression analyses were used to control for the remaining measurable differences between program participants and nonparticipants.

Overall, we found the CHF disease management program to be cost beneficial, saving the health plan about $1.08 in medical expenditures for every dollar invested in the program. Adding benefits to patients (in the form of fewer copayments or deductibles resulting from lower service use) and other payers, the program saved $1.15 in medical expenditures for every dollar spent on it.

ROI estimates varied according to level of risk, however, and others may want to take note and design their evaluations accordingly. In our case, ROI estimates were higher for at-risk (ie, lower risk) patients. Thus, a pilot ROI analysis may suggest that program sponsors or vendors pay more attention to refinements that can lead to higher ROI for particular subsets of patients as programs are expanded beyond the pilot stage.

The cost savings reported here are similar to estimates derived from studies of other CHF disease management programs.[16] A review of 16 studies found that CHF programs often resulted in decreased hospitalization.[17] While most disease management programs for CHF seem to work by reducing hospital length of stay or readmission rates,[12,18] the impact is not always statistically significant[19] and the success of these programs may be a function of the types of patients enrolled and the severity of their illnesses.[9,20] In our analysis, we stratified program participants by risk status and controlled for differences in comorbidities. The impact of disease management on medical expenditures was greater and statistically significant among the high-risk patients we studied, but program costs were higher for them too.

Because of the higher program costs, the NPV and ROI figures were more favorable for the at-risk (ie, lower risk) patients.

Overall, we found a positive ROI for the CHF disease management program. Cost savings were realized even though spending for prescription drugs appeared to be higher for program participants than for nonparticipants. Thus, a positive ROI may have been due, in part, to improvements in medication use (eg, higher increases in the use of ACE inhibitors and digoxin, which often are recommended for CHF patients[21]) and lower inpatient and outpatient service use for participants as compared to nonparticipants.

Limitations

The analyses we conducted were limited by several factors. These included (1) a small sample size, (2) the imperfect matching process, (3) the voluntary nature of program participation, (4) the inability to randomize patients to treatment or control group status, (5) a relatively short outcome period of interest (1 year), and (6) a primary focus on monetary outcomes that excluded description of the program's impact on quality of life. Each of these is addressed below.

First, the pilot nature of our study resulted in a small sample of program participants (although the 457 participants included here are still greater in number than those reported in some other studies[20,22,23]). Apparently this was more problematic for low-risk beneficiaries, causing the impact estimate for this group to be statistically insignificant (ie, $p > 0.05$; Table 2).

Expanding upon this notion for a moment, it is worth some discussion of why the statistically insignificant impact estimate for the low-risk group was used in the ROI and NPV formulas. If the impact estimate cannot be differentiated from zero dollars statistically, does this mean that zero is a better estimate of program impact? The answer to this question is "no." The ordinary least squares (OLS) regression procedure used for the analysis provides what is commonly known as the "best linear unbiased estimate" of program impact. This means that, in the class of linear models, there is no other estimate which is more accurate and more efficient than the one produced by OLS regres-

sion.[24] The fact that the estimate has a confidence interval that includes zero dollars should lead readers to note its inherent variability with a sample this size, but not its accuracy. Zero is neither a more accurate nor a more likely answer to the question of the impact of the CHF program for low-risk patients, so zero dollars should not be used in the ROI and NPV formulas.

The second limitation of the evaluation relates to the imperfect matching process that was used. We noted that 136 of the CHF program participants could not be matched on the basis of ACC category or diagnosis. Other matching algorithms, such as propensity score matching, could have been tried. However, no matching process can ever be perfect, and even propensity score matches (which are designed to match on the basis of several variables at once) often require supplemental regression analyses to control for variables that are still significantly different between treatment and comparison group members after matching is completed.[25]

We addressed the imperfect matching by using regression analyses to control for several demographic and case mix-related variables that were significantly different between participants and nonparticipants. We also conducted sensitivity analyses to see how results for the primary payer (FEP) would change after dropping the 136 CHF program participants who could not be matched to nonparticipants. Dropping these participants tended to raise the ROI estimates for at-risk patients and reduce the ROI estimates for high-risk patients. The reductions among the high-risk patients were associated with the loss of 41% of the high-risk participants, though, casting doubt on the utility of dropping these patients.

The third limitation of the evaluation deals with the voluntary nature of the CHF disease management program. Because it was voluntary, a substantial number of patients refused to enroll or dropped out within 90 days of enrollment. Although the results might have been different if enrollment had been higher, two factors help lessen this concern. First, decisions regarding inclusion and exclusion criteria were made before analyses were begun. Second, the research organization performing the evalua-

tion was paid by the health plan and not the program vendors, in order to ensure objectivity. We therefore expect internal validity of this study was high, noting that a matched comparison group was used and other adjustments were carried out in the multivariate analyses to limit threats to internal validity.

Considered together, the first two limitations of our study (ie, the small sample size and voluntary nature of the program) might have resulted in a sample that was highly motivated to change their behavior. Without a direct measure of motivation for those who participated and those who did not, it is not possible to separate the impact of motivational factors from the impact of program participation. However, it seems reasonable to expect that the associated reduction in expenditures may not have taken place as quickly without a CHF program to serve as a stimulus for that change.

Another limitation of our study was the inability to test the CHF program as part of a randomized controlled trial. As a result, there may be unmeasurable differences between program participants and nonparticipants that influenced our findings. We attempted to avoid bias by selecting a comparison group from a neighboring state subject to the same insurance program and the same inclusion criteria, and by using regression analyses to control for measurable differences between program participants and nonparticipants. However, some unmeasured differences may still remain and these differences may have influenced findings to an unknown degree. Selection bias due to unmeasurable differences in treatment and control group members would exist regardless of which matching strategy is chosen in nonrandomized studies.

Next, the focus of this study was on ROI after only 1 year of program implementation. In many disease management programs, it may take more than 1 year to realize gains from changes in behavior that are induced by the program. In addition, program providers may learn to be more efficient as time goes on, and may reap benefits in later years that cannot be experienced in the first program year. Thus, the ROI for subsequent years may be higher than the ROI estimated for the first year.

Finally, this study focused primarily on financial outcomes. Even though economic studies can yield rational decisions about the usefulness of CHF disease management programs,[26] they are not complete. Nonmonetary outcomes also should be considered before a complete understanding of the impact of the program is obtained.

Lessons from the CHF application

Health plans may wish to pilot test their programs to find out what the likely impact of their programs may be before rolling out to a larger audience, as the sponsoring health plan did here. Lessons from the pilot test can be gained via empirical evaluation. Some lessons gained from the CHF application described here are as follows:

1. Pilot tests typically involve small samples. These small samples may have substantial variability in medical expenditures. The appropriate sample size for the pilot can be informed by conducting power analyses to estimate the size of the pilot needed to be able to detect a significant program effect.
2. It can be cumbersome and difficult to match treatment and comparison group members. More systematic methods, such as the use of propensity score techniques, are gaining popularity and should be considered.
3. Regardless of the matching strategy, tests for differences between treatment and comparison group members should be conducted and statistical analyses should be used to adjust for any remaining differences.
4. Despite the intricacies involved in adjusting for differences, a comparison group should be used. Otherwise it will not be possible to infer whether observed changes are really due to program participation.
5. The results of the ROI estimation process may vary by level of patient risk. Thus, policy implications also may vary for different patient subgroups.
6. No vendor is perfect, no program is perfect, no sponsor is perfect, and no evaluation is perfect. Mistakes can be made inadvertently at many levels. Thus "bad" or unexpected results should be viewed as opportunities for discussions about how to better organize

programs, how to run them more efficiently, and how to better work with patients and providers. A key benefit of conducting a pilot study is that these discussions can occur prior to full-scale rollout, thereby benefiting patients and health plans tremendously, once the larger rollout of the program is implemented.

CONCLUSIONS

The pilot CHF disease management program offered to FEP enrollees seemed to be cost beneficial overall during its initial year of operation. For high-risk patients, it was roughly a financial breakeven program. For at-risk patients, the return was higher but less certain. Savings from the program probably resulted from decreases in inpatient and outpatient expenditures that were not offset by an increase in pharmaceutical spending. These results were obtained for a sample of CHF patients that included low-risk as well as high-risk patients, potentially broadening the application and benefits of a CHF disease management program to a much larger pool of patients than previously thought possible.

It has been noted that many CHF disease management program evaluations have been conducted with little rigor and that their reported results often rely on before-and-after study designs that exclude comparison or control groups.[27,28] This may make their conclusions suspect. In our case, a comparison group was included. Randomization was not feasible, so we used matching techniques and multiple regression analyses to adjust for differences between program participants and comparison group members. The matching methods we used were not perfect, and several of the treatment group members found no corresponding match in the control group based upon gender and comorbidities. We relied on multiple regression analyses to control for remaining differences, but it is still possible that unobserved differences between the two groups influenced our results. Other health plans are likely to have the same problem when their evaluations are conducted, because matching involves art as well as science. Nevertheless, researchers are advised to look for ways to balance the samples of treatment and comparison group members when analyses are conducted. Matching and regression-based methods may limit regression to the mean and other threats to validity. Thus, they generally are preferred to the absence of a comparison group or to less rigorous attempts to control for differences between program participants and nonparticipants.[28]

REFERENCES

1. Robert Wood Johnson Foundation (RWJF). Trends monitoring in chronic care, United States. Princeton, NJ: Robert Wood Johnson Foundation, 1997.
2. American Heart Association. 2001 Heart and stroke statistical update. Dallas, TX: American Heart Association, 2000.
3. Ho KKL, Pinsky JL, Kannel WB, Levy D. The epidemiology of heart failure: The Framingham Study. J Am Coll Cardiol 1993;22:6A–13A.
4. Eriksson H. Heart failure: A growing public health problem. J Intern Med 1995;237:135–141.
5. Massie BM, Shah NB. Evolving trends in the epidemiologic factors of heart failure: Rationale for preventive strategies and comprehensive disease management. Am Heart J 1997;133:703–712.
6. Ghali JK, Cooper R, Ford E. Trends in hospitalization rates for heart failure in the United States, 1973–1986: Evidence for increasing population prevalence. Arch Intern Med 1990;150:769–773.
7. Graves EJ. National hospital discharge survey: annual summary, 1993. Vital and Health Statistics Series 13: Data from National Health Survey. Washington, DC: US Government Printing Office, 1995;121:1–63.
8. Centers for Disease Control and Prevention. Cerebrovascular disease mortality and Medicare hospitalization—United States, 1980–1990. MMWR 1992;41:447–480.
9. Simons WR, Haim M, Rizzo J, Zannad F. Effect of improved disease management strategies on hospital length of stay in the treatment of congestive heart failure. Clin Ther 1996;18:726–746.
10. O'Connell JB, Bristow MR. Economic impact of heart failure in the United States: Time for a different approach. J Heart Lung Transplant 1994;13:S107–S112.

11. Centers for Disease Control and Prevention. Facts about heart failure in older adults [On-line]. Available: ⟨http://www.cdc.gov/od/oc/media/fact/f990305.htm. 1999⟩, accessed December 17, 2002.

12. Knox D, Mischke L. Implementing a congestive heart failure disease management program to decrease length of stay and cost. J Cardiovasc Nurs 1999;14:55–74.

13. DxCG Incorporated. DxCG® risk adjustment software release 3.0. For additional information, see ⟨http://www.dxcg.com⟩.

14. Drug topics red book. Montvale, NJ: Medical Economics, Inc., 1999.

15. Getzen TE. Medical care price indexes: Theory, construction, and empirical analysis of the U.S. series 1927–1990. Adv Health Econ Health Serv Res 1992;13:83–128.

16. Goetzel RZ, Ozminkowski RJ, Villagra V, Duffy J. Return on investment in disease management: A review. Health Care Fin Rev 2005;26:1–19.

17. Rich MW. Heart failure disease management: A critical review. J Card Fail. 1999;5:64–75.

18. McAlister FA, Lawson FM, Teo KK, Armstrong PW. A systematic review of randomized trials of disease management programs in heart failure. Am J Med 2001;110:378–384.

19. Philbin EF, Rocco TA, Lindenmuth NW, Ulrich K, McCall M, Jenkins PL. The results of a randomized trial of a quality improvement intervention in the care of patients with heart failure. Am J Med 2000;109:501–503.

20. Riegel B, Carlson B, Glaser D, Hoagland P. Which patients with heart failure respond best to multidisciplinary disease management? J Card Fail 2000;6:290–299.

21. Mark DB. Economics of treating heart failure. Am J Cardiol 1997;80:33H–38H.

22. Whellan DJ, Gaulden L, Gattis WA, et al. The benefit of implementing a heart failure disease management program. Arch Intern Med 2001;161:2223–2228.

23. Hershberger RE, Ni H, Nauman DJ, et al. Prospective evaluation of an outpatient heart failure management program. J Cardiac Fail 2001;7:64–74.

24. Kmenta J. Elements of econometrics, 2nd ed. New York: Macmillan, 1986.

25. McCaffrey D, Ridgeway G, Morral A. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. Psychol Methods 2004;9:403–425.

26. Weintraub WS, Cole J, Tooley JF. Cost and cost-effectiveness studies in heart failure research. Am Heart J 2001;43:565–576.

27. Moser DK. Heart failure management: Optimal health care delivery programs. Annu Rev Nurs Res 2000;18:91–126.

28. Walker DR, McKinney BK, Cannon-Wagner M, Vance RP. Evaluating disease management programs. Dis Manage Health Outcomes 2002;10:613–619.

Address reprint requests to:
Ronald J. Ozminkowski, Ph.D.
Thomson Medstat
777 East Eisenhower Parkway, 804B
Ann Arbor, MI 48108

E-mail: ron.ozminkowski@thomson.com