HEALTH POLICY

Who Gets Disease Management?

Melinda Beeuwkes Buntin, PhD¹, Arvind K. Jain, MS², Soeren Mattke, MD, ScD³,
and Nicole Lurie, MD, MSPH¹

¹RAND Corporation, Arlington, VA, USA; ²World Bank, Washington, DC, USA; ³Bain & Company, Boston, MA, USA.

BACKGROUND: Disease management (DM) has been promoted to improve health outcomes and lower costs for patients with chronic disease. Unfortunately, most of the studies that support claims of DM's success suffer from a number of biases, the most important of which is selection bias, or bias in the type of patients enrolling.

OBJECTIVE: To quantify the differences between those who do and do not enroll in DM.

DESIGN, SETTING, AND PARTICIPANTS: This was an observational study of the health care use, costs, and quality of care of 27,211 members of a large health insurer who were identified through claims as having asthma, diabetes, or congestive heart failure, were considered to be at high risk for incurring significant claims costs, and were eligible to join a disease management program involving health coaching.

MEASUREMENTS: We used health coach call records to determine which patients participated in at least one coaching call and which refused to participate. We used claims data for the 12 months *before* the start of intervention to tabulate costs and utilization metrics. In addition, we calculated HEDIS quality scores for the year prior to the start of intervention.

RESULTS: The patients who enrolled in the DM program differed significantly from those who did not on demographic, cost, utilization and quality parameters prior to enrollment. For example, compared to non-enrollees, diabetes enrollees had nine more prescriptions per year and higher HbA1c HEDIS scores (0.70 vs. 0.61, $p < 0.001$).

CONCLUSIONS: These findings illuminate the serious problem of selection into DM programs and suggest that the effectiveness levels found in prior evaluations using methodologies that don't address this may be overstated.

KEY WORDS: disease management; selection bias; HEDIS; health care utilization; evaluation.

J Gen Intern Med 24(5):649-55

DOI: 10.1007/s11606-009-0950-8

© Society of General Internal Medicine 2009

INTRODUCTION

For over a decade, disease management — a system of coordinated interventions and communications for populations with conditions in which patient self-care efforts are significant¹ — has been promoted as a promising way to improve quality of care, improve health outcomes and lower costs for patients with chronic disease²⁻⁴. The concept underlying disease management is attractive, and both private insurers and public policymakers have been quick to embrace disease management programs⁵. Ample claims have also been made for the success of these programs⁶⁻⁷. Unfortunately, there is very little evidence to support this optimistic view, in particular for the large-scale, population-based programs that are currently of most interest to purchasers⁸. Many studies that purport to support these claims are either anecdotal, involve highly selected patients in closed systems of care, or suffer from a number of biases, the most important of which is bias in the recruitment or enrollment of participants⁹⁻¹¹.

The issue of selection bias, which is the subject of this paper, is a straightforward one. Patients might be recruited into a program explicitly because they are likely to attain quality and cost benchmarks. More benignly, patients who are more likely to attain those benchmarks (e.g. because they are more engaged with their own care) might be differentially interested in signing up for disease management programs or plans that offer them¹². In either case the result is the same: in order to ascertain whether disease management programs themselves, rather than patient selection, cause improvements in quality or health outcomes or reductions in costs, studies must be able to attribute the changes in endpoints to the intervention rather than confounding factors, including selection. Short of a randomized trial, rigorous observational designs are needed to answer this question. A commonly used approach in observational studies is to use the non-participants in the program as reference group. This requires controlling for both observable and unobservable differences between the "treatment" population in disease management and the comparison population. Observable factors include factors income and co-morbidity, while unobservable factors might include how well people currently self-manage, and their motivation to change their health habits. The fact that most studies trumpet positive results without accounting for such factors speaks to the heterogeneous quality of research in the disease management field¹¹. Indeed, a recent review found only one large-scale study of a population-based disease management program that used a rigorous quasi-experimental design⁸.

This paper demonstrates that failure to address selection is a serious shortcoming in the literature: those who voluntarily enroll in a large-scale, population-based disease management

Received May 8, 2008

Revised October 20, 2008

Accepted March 2, 2009

Published online March 24, 2009

program are very different from those who choose to not enroll. Studies that evaluate the effectiveness of disease management programs must account for these differences in order to have validity. This paper quantifies the differences between the groups who do and do not enroll, and should be useful to those who are interested in evaluating the effectiveness of disease management programs.

METHODS

Study Population

The population from which the study population was identified consisted of a sample of 60,048 members of a large health plan who were identified through claims as having asthma, diabetes, or congestive heart failure, and who were considered to be at high risk for incurring high costs. Risk for high costs was assessed using a model developed by the health plan: patients in the top fifth of predicted costs based on their use patterns and diagnoses were eligible to be invited to enroll in the disease management program. (Characteristics in the models included inpatient stays, ER visits, outpatient encounters, and numbers of prescriptions.) Patients were excluded from study eligibility if they had another co-existing serious illness. Taken together these eligibility criterion were designed to produce a set of relatively homogeneous study populations.

The insurer randomly selected half of this population for enrollment in the initial phase of their disease management program (leaving approximately 30,000). We excluded 2,800 people for whom there was less than a full year of enrollment data. The resulting population included 12,236 diabetics, 5,315 patients with congestive heart failure (CHF) and 9,660 patients with asthma. We stratified our analyses of asthma patients into two groups: ages <20 years and ages 20+ years, both because of the bi-modal distribution of cases by age and because for the younger age group, the patient's parents are the ones who would actually receive the health coaching. After the patient populations were identified, the "health coaching" staff began the process of enrolling patients in the disease management program. Health plan enrollees were sent a letter inviting them to enroll and were then recruited by telephone. The insurer faced a number of barriers to enrollment including incomplete address and telephone information and patients who did not accept or return phone calls. Less than one percent of patients explicitly refused to participate in the program (92 Diabetes, 38 CHF, and 43 Asthma). However, after six months of enrollment efforts only 9.4% of the overall group targeted for enrollment had received one or more health coaching calls. We consider this the group enrolled in the disease management program.

The original intention of our study was to evaluate the effectiveness of the program by comparing the intervention and control groups. Due to the low percentage (9.4%) of intervention group members that actually enrolled, we instead focused on the differences between those who were successfully enrolled and those who were not. We explore the differences between these two groups in terms of demographics, health care costs, utilization, and quality indicators in the one-year period prior to the disease management program implementation.

Data Sources and Measures

The insurer provided us with three types of administrative data: health coach call records, health plan eligibility data and claims data. The use of these data was approved by RAND's institutional review board and protected under data use agreements. As described above, we used the call records to determine which patients received a coaching call and which refused to participate. The eligibility data included names, addresses, age, gender, region, and plan type [health maintenance organization (HMO), point-of-service (POS), preferred provider organization (PPO), and fee-for-service (FFS)]. We used the patients' last names and addresses to impute race/ethnicity and median household income via Census block group geocoding and surname analysis using methods described by Fremont et al.¹³ and Elliott et al.¹⁴. Surname analysis was first used to identify members with a high probability of being Hispanic or Asian; then a Bayesian algorithm was used to impute race/ethnicity using this information and data from the 2000 census at the block group level¹⁵. Census block groups correspond to a small neighborhood of approximately 1,000 people; median household income was also obtained at the block group level. If an address was invalid or missing, geocoding was not possible. Ultimately, we lacked income information for 9%, and lacked ethnicity information for 8% of the people eligible for health coaching. (A smaller percentage was missing ethnicity since surname analysis is used to determine ethnicity as a first pass.)

We used the claims data for the 12 months *before* the start of intervention to tabulate costs in mutually exclusive categories based on where services were incurred: office/outpatient, inpatient, emergency department, drug and other (other includes home, subacute, ambulance, lab, and other). Similarly, for utilization, we tabulated any inpatient utilization (yes/no), number of inpatient days per event, number of emergency department visits, number of office visits and number of prescriptions filled. Costs were summed and are presented on a per-person per-eligible month (PMPM) basis; utilization is presented per-person per-year (PMPY). Costs were taken from the "allowed amount" fields on the claims and include both enrollee and insurer payments.

Using the claims data we also computed HEDIS® quality scores (ratios) for the year prior to the start of the intervention¹⁶. We calculated six quality ratios, four for diabetes and one each for CHF and asthma: the numerator of each ratio was defined based on the HEDIS® criteria, and the denominator was the eligible population. Diabetes quality ratios were the proportions of individuals receiving LDL-C screening, eye exams, urine tests for microalbuminuria, and HbA1c testing. The other two quality ratios were the proportion of asthmatics with prescribed medication for long-term asthma control and the proportion of CHF patients with prescribed beta blocker medication.

Analysis

We compared the demographics, health care costs, utilization, and quality scores of two groups: those who were successfully contacted and enrolled in the intervention, and those who were not. Specifically, we compared total costs, breakdowns of costs and utilization counts by type of service, and HEDIS® quality scores for the year *before* enrollment in the disease management program. We also examined the extent to which differ-

ences between the two groups can be “controlled” for using standard regression analysis and the types of demographic factors typically available to researchers evaluating disease management programs. We used a Chi-square test to compare demographic characteristics such as gender, ethnicity, region, and plan type, between those who did and did not receive the intervention. Similarly, we used the Chi-square test to compare inpatient events and HEDIS® quality ratios between the two groups. We used the Wilcoxon rank-sum test to compare continuous variables such as age, income, cost measures, and utilization measures because of the skewed distributions of these variables. When comparing ethnicity and income between the two groups, we calculated differences only for people with non-missing values. All p-values are two-sided.

To estimate the independent effects of demographic variables, plan type, and prior utilization on enrollment in the program, and to assess their predictive power, we fit logistic regression models. Co-variables included age, gender, race, income, plan type (FFS/PPO vs. HMO/POS), HEDIS® numerator(s) in the prior year (e.g., for asthma this is an indicator variable of whether they received a prescription for appropriate long-term asthma control), any inpatient event, any ER event, number of office visits, and number of prescription fills (both office visits and prescription fills are modeled as tertiles). The Hosmer-Lemeshow test and c-statistic were used to assess the model fit and the importance of observable and unobservable characteristics to the receipt of health coaching. Analyses were performed using SAS version 8.2 (SAS Institute Inc., Cary, USA).

RESULTS

Comparisons of demographic characteristics by health coaching status are provided in Table 1. Among enrollees with diabetes and adult asthma, those who received health coaching were significantly older than those who did not. Similarly, for CHF, recipients of health coaching were in older age groups than those not receiving any health coaching calls ($p=0.02$), but this difference was not found in those 75+ years of age. In contrast, for child asthma, parents of younger patients were more likely to receive health coaching. For all disease groups with the exception of child asthma, females were significantly more likely to receive health coaching. For diabetes, whites were significantly more likely to receive health coaching ($p<0.001$) whereas, blacks were significantly more likely to receive health coaching for CHF ($p<0.001$). For all three disease groups, participants in HMOs were significantly more likely to be enrolled in the disease management program (although for adult asthma the p-value is only marginally significant ($p=0.07$)).

For all three disease groups, office/outpatient, drug, and total costs were significantly higher in the baseline year prior to the commencement of the disease management program for those who later ended up enrolling in the program compared to those who did not. Cost differences ranged from a low of \$11 PMPM for child asthma to a high of \$227 PMPM for CHF (Table 2). Regarding utilization for all disease groups, the number of office visits and prescriptions filled were significantly greater ($p<0.001$) for those who later received health coaching. Diabetes and CHF coaching recipients had on average three more office visits per year in the year prior than those who did not enroll in the health coaching program (for

asthma the differential is two visits) and nine to eleven more prescriptions per year. CHF health coaching recipients had significantly higher percentages of any inpatient utilization ($p=0.04$) while adult asthma health coaching recipients also had higher frequency of any inpatient utilization, but this difference is only marginally significant ($p=0.06$).

Those who later received health coaching had significantly better quality of care in the year prior to enrollment than those who did not (Table 3). For LDL-C tests, enrollees in disease management were seven percentage points more likely to have the test than those not enrolled (0.60 vs. 0.53). Similarly, for receipt of HbA1c tests, the differential is nine percentage points. For beta-blocker medications in the CHF group, the differential is +7 percentage points and for long-term control medications in the asthma groups, the differentials are +11 and +8 percentage points for the <20 and 20+ age groups respectively.

Overall, the multivariate logistic regression models reinforced the bi-variate findings. Age and gender continued to be significantly associated with the receipt of health coaching in the multivariate framework as did race/ethnicity for diabetes and CHF. For diabetics, receipt of an HbA1c test in the prior year is strongly associated with health coaching, as is receipt of prescriptions for control medications for adult asthma. The associations between prior office visits, drug utilization, and enrollment in health coaching remain constant for all three disease groups. It should be noted, however, that while the Hosmer-Lemeshow statistics indicate acceptable fit, these models do not adequately predict enrollment well — i.e., their “r-squared” values and c-statistics are not high. Ideally the c-statistics should be above 0.70, but in our models they range from 0.63–0.65 — indicating that there are other, unobservable characteristics associated with the receipt of health coaching. We also examined this difference in a multivariable framework and found that even after controlling for age, gender, ethnicity, and plan type, those who enrolled in the disease management program were already receiving better care prior to enrollment than those who did not (results not shown).

DISCUSSION

Our results indicate that selection is an important issue in explaining outcomes of disease management programs, and suggest that analyses of disease management programs that do not carefully control for selection are likely to present overly positive findings about their effectiveness. When comparing utilization, costs, and quality of care in the year prior to the start of a disease management program for health plan members with asthma, diabetes or congestive heart failure, we found significant differences in demographics, cost, utilization and quality parameters at baseline for those who did and did not enroll. If these differences had instead been reported a year after initiation of the program, rather than the year before, the program might have been deemed a success, particularly because the participants exhibited utilization patterns that were commonly ascribed to disease management programs. Those who enrolled later were older, had significantly higher outpatient and pharmacy utilization, and associated health care costs, as well as a higher quality level of care.

Table 1. Demographics

	Enrolled	Not Enrolled	P-value	Enrolled	Not Enrolled	P-value
	Diabetes			CHF		
N	1350	10886		549	4766	
Age*			p<0.001			p=0.02
1-19	0.30%	0.30%		0.00%	0.10%	
20-29	1.40%	2.90%		0.00%	0.50%	
30-39	4.10%	7.50%		1.60%	2.90%	
40-49	13.10%	16.90%		7.70%	9.00%	
50-59	33.50%	31.50%		21.70%	20.00%	
60-64	18.00%	16.20%		17.90%	13.70%	
65-74	20.40%	15.50%		25.90%	19.40%	
75+	9.20%	9.30%		25.30%	34.40%	
Gender			p<0.001			p=0.02
Female	51.60%	43.10%		51.70%	46.50%	
Ethnicity			p<0.001			p<0.001
White	73.80%	68.80%		70.70%	74.50%	
Black	5.30%	4.10%		10.40%	6.50%	
Latino	11.10%	14.30%		5.30%	4.80%	
Asian	2.70%	4.90%		0.50%	1.80%	
Missing†	7.10%	8.00%		13.10%	12.40%	
Income*			p=0.68			p=0.82
< \$30,000	10.50%	10.80%		12.40%	12.70%	
\$30K-\$39,999	16.60%	17.00%		19.50%	18.90%	
\$40K-\$49,999	17.90%	16.80%		15.80%	15.70%	
\$50K-\$59,999	15.30%	13.30%		11.10%	13.10%	
\$60K-\$69,999	10.70%	10.40%		10.40%	9.70%	
\$70K-\$79,999	6.30%	6.90%		7.50%	6.00%	
\$80K+	14.50%	15.40%		9.80%	11.20%	
Missing†	8.10%	9.40%		13.50%	12.70%	
Plan type			p<0.001			p<0.001
PPO	51.90%	58.30%		40.10%	40.80%	
POS	6.40%	4.50%		4.70%	4.20%	
HMO	29.90%	25.80%		33.90%	22.10%	
FFS	11.70%	11.50%		21.30%	32.80%	

	Asthma (Ages 1-19)	P-value	Asthma (Ages 20+)	P-value
Age*	260	3216	406	5778
1-19	100.00%	100.00%		
20-29			8.90%	20.20%
30-39			20.00%	26.40%
40-49			38.40%	30.00%
50-59			32.50%	23.20%
60-64			0.20%	0.20%
65-74			0.00%	0.00%
70+			0.00%	0.00%
Gender		p=0.21		p<0.001
Female	33.50%	37.30%	68.00%	58.00%
Ethnicity		p=0.84		p=0.78
White	71.50%	71.50%	80.30%	76.70%
Black	3.10%	3.20%	2.20%	3.00%
Latino	17.70%	15.70%	11.10%	11.30%
Asian	4.20%	4.90%	2.00%	2.40%
Missing†	3.50%	4.60%	4.40%	6.70%
Income		p=0.05		p=0.37
< \$30,000	5.80%	7.60%	9.40%	9.40%
\$30K-\$39,999	10.80%	11.90%	16.00%	14.60%
\$40K-\$49,999	15.80%	16.00%	17.20%	17.00%
\$50K-\$59,999	11.90%	14.90%	17.20%	15.40%
\$60K-\$69,999	11.90%	13.20%	12.10%	11.50%
\$70K-\$79,999	13.10%	8.50%	6.90%	8.10%
\$80K+	26.20%	22.30%	15.50%	16.60%
Missing†	4.60%	5.70%	5.70%	7.40%
Plan Type		p=0.01		p=0.07
PPO	50.00%	56.10%	54.40%	59.10%
POS	8.10%	5.90%	6.40%	5.60%

(continued on next page)

Table 1. (continued)

	Asthma (Ages 1–19)		P-value	Asthma (Ages 20+)		P-value
HMO	41.20%	37.00%		38.90%	34.20%	
FFS	0.80%	1.00%		0.20%	1.10%	

*Because age and income are continuous measures, we used the Wilcoxon rank-sum test to compare the distributions between those who did and did not receive health coaching. Otherwise the Chi-square test (or Fisher's exact test) is used for all other statistical comparisons

†Those people missing ethnicity (or income) are not included in the comparison between those who did and did not receive health coaching

‡Wilcoxon rank-sum p-value is $p < 0.001$. The median age for those who received health coaching is 9 where as the median age for those that did not receive health coaching is 11

Table 2. Costs and Utilization

		Enrolled		Not Enrolled		Enrolled		Not Enrolled	
		Diabetes		P-value§		CHF		P-value§	
N		1350		10886		549		4766	
Costs*									
Office/Outpatient	mean	141.6	113.4	$p < 0.001$		183.8	164.8	$p < 0.001$	
	median	32.3	12.6			60.2	35.2		
Inpatient	mean	270	240.2	$p = 0.12$		806.2	658.1	$p = 0.12$	
	median	0	0			0.2	0		
ER	mean	5	4.9	$p = 0.04$		11.4	6.9	$p = 0.14$	
	median	0	0			0	0		
Drug	mean	156	126.9	$p < 0.001$		140.2	109.2	$p < 0.001$	
	median	92.2	68.1			69.1	26.8		
All Other	mean	112.2	98.4	$p = 0.22$		141.6	116.8	$p = 0.07$	
	median	8.8	8.2			9.4	6		
Total	mean	684.8	583.9	$p < 0.001$		1283.3	1055.8	$p < 0.001$	
	median	271.5	211.4			365.9	236.3		
Utilization†									
IP Event (Any)	percent	21.00%	20.30%	$p = 0.54$		42.30%	37.80%	$p = 0.04$	
IP Days Per Event‡	median	4.2	4.2	$p = 0.55$		5.3	5.9	$p = 0.01$	
	mean	3	3			4	4		
ER Visits (Any)	percent	21.70%	20.00%	$p = 0.15$		36.30%	34.90%	$p = 0.52$	
Office Visits	mean	9.4	7	$p < 0.001$		14.1	11.1	$p < 0.001$	
	median	7	3			12	9		
Rx filled	mean	45.4	36.2	$p < 0.001$		46.1	34.6	$p < 0.001$	
	median	40	29			42	22		

		Asthma		P-value§		Asthma		P-value§	
		(Ages 1–19)				(Ages 20+)			
		260		3216		406		5778	
Costs*									
Office/Outpatient	mean	57.6	42.7	$p = 0.06$		90.1	78.6	$p = 0.01$	
	median	7.3	2.6			7.6	1.6		
Inpatient	mean	39.3	44	$p = 0.74$		139.4	80.4	$p = 0.06$	
	median	0	0			0	0		
ER	mean	1.9	3.1	$p = 0.60$		4.2	4	$p = 0.83$	
	median	0	0			0	0		
Drug	mean	57.3	44.4	$p = 0.003$		111.4	79.1	$p < 0.001$	
	median	25.5	15.5			54.3	35		
All Other	mean	23.3	34.3	$p = 0.99$		71.8	51.6	$p = 0.11$	
	median	0	0			1	0		
Total	mean	179.4	168.5	$p = 0.02$		416.9	293.8	$p < 0.001$	
	median	70.7	55.7			137.5	94.4		
Utilization†									
IP Event (Any)	percent	8.10%	8.30%	$p = 0.89$		12.80%	9.90%	$p = 0.06$	
IP Days Per Event‡	median	2.5	2.6	$p = 0.59$		3.6	3.4	$p = 0.39$	
	mean	2	2			3	2.3		

(continued on next page)

Table 2. (continued)

		Asthma		P-value§	Asthma		P-value§
		(Ages 1–19)			(Ages 20+)		
ER Visits (Any)	percent	18.90%	19.90%	p=0.67	18.20%	17.40%	p=0.68
Office Visits	mean	6.5	4.5	p<0.001	7	5.1	p<0.001
	median	4	2		3	2	
Rx filled	mean	14.5	12.9	p=0.002	34.5	26.2	p<0.001
	median	13	10		27	20	

*Costs are presented per eligible month

†Utilization measures that are continuous are presented per eligible year

‡The summary statistics and statistical comparison for the number of inpatient days per event are only for the subset of individuals that had at least one inpatient event

§All p-values are from the Wilcoxon rank-sum test except for inpatient/ER event (Any) where the Chi-square test is used for statistical comparisons

The fact that patients were more likely to satisfy HEDIS® criteria *before* they were enrolled in the program suggests that favorable selection of patients into disease management programs poses a methodological issue that must be addressed in the evaluation of disease management programs. In addition, it is important to note that this effect remained even after controlling for demographic characteristics, including some that are not traditionally captured in administrative data (e.g., income, race/ethnicity) and thus not included when evaluating programs.

We offer several alternative explanations for these findings. First, it is possible that those who enrolled in disease management were sicker, as reflected in higher utilization and drug costs, and that their greater utilization afforded them more opportunities to receive recommended elements of care. Second, it is possible that those who could be contacted for potential enrollment were not only older but in more stable life situations, providing another explanation for more medication refills and better quality of care. Alternatively, it is possible that those who enrolled already took better care of themselves, with more follow-up visits and better medication adherence. This unobservable attribute, which we might call motivation, may explain why some patients were more willing to accept the offer to participate in a disease management program. In addition, the low uptake rate itself may have contributed to the differences. We would have liked to compare the enrollment response rates achieved by this insurer to those of others. However, outside of small, carefully controlled experiments testing disease management strategies, most studies do not report these rates. Finally, the intervention lacked clear patient and provider involvement in decision-making, which the literature suggests are important elements in successful

disease management programs, and this may have contributed to the low uptake of the program¹⁷. The data do not necessarily favor one explanation over another. However, they clearly suggest that those who enrolled systematically differed from their counterparts who did not.

A common evaluation approach that disease management vendors, and the companies that purchase their services, have relied on is to compare costs, utilization, and quality of care for program enrollees to those who do not enroll and/or were not eligible to enroll. As we have demonstrated, this strategy could result in misleading conclusions and calls into question whether the positive outcomes attributed to these programs have been overstated. Fortunately, there are approaches to the evaluation of disease management programs that can, at least partially, avoid these potential biases. While conducting randomized, controlled trials of disease management programs utilizing an intent-to-treat perspective is the best evaluation method, this is rarely feasible in the business world. However, other statistical methods may be used to control for observable and unobservable differences between the intervention and the control groups^{11,17–19}. An approach that does not rely on statistical modeling considers all program eligible members as the treatment group, regardless of whether or not they actually enroll, and use the non-eligible members as a comparison group²⁰.

There are a few potential limitations to our study. First, the insurer had significant difficulty reaching its members because of incorrect contact information. It is possible that with more complete and accurate contact information, some of the differences in unobserved variables might be lessened. However, based on discussions with other health plans and disease management vendors, we do not believe this problem to be unique to the health plan in this study. Second, this study used a definition of enrollment in disease management that may differ from some in the field. While we required a call with a health coach to define enrollment, other practices may simply consider all of those that received a mailing to be enrolled. It is also possible that the use of geocoding and surname analysis, which we used to estimate race/ethnicity and income, introduced biases and weakened our ability to control for those factors. However, this seems unlikely given reports that indirect measures of race/ethnicity and income, as calculated in this analysis, perform approximately as well as direct measures in assessing quality¹³.

Our findings should be of use to those who purchase, participate in, and evaluate disease management programs. They point to the importance of rigorous evaluation of such

Table 3. HEDIS Quality Measures in Year Prior to Enrollment

	Enrolled	Not Enrolled	P-value*
Diabetes			
Lipid	0.60	0.53	p<0.001
Eye exam	0.95	0.91	p<0.001
Microalbumin	0.30	0.26	p=0.003
HbA1c	0.70	0.61	p<0.001
CHF			
Beta blockers	0.43	0.36	p<0.001
Asthma			
Meds (Ages 1–19)	0.67	0.56	p<0.001
Meds (Ages 20+)	0.60	0.51	p=0.002

programs: any results that do not account for these factors may be misstated. We hope that these findings promote more discussion of the value of disease management and encourage those who purchase disease management services to require rigorous evaluations.

Acknowledgements: *We would like to acknowledge funding from the insurer whose experience is described in this paper and assistance from our colleagues Sarah Zakowski and Erin Murphy.*

Conflict of Interest: *Soeren Mattke has done research and consulting projects for operators and purchasers of disease and care management programs. None of the other three authors have any conflicts of interest to declare.*

Corresponding Author: *Nicole Lurie, MD, MSPH; RAND Corporation, 1200 South Hayes Street, Arlington, VA 22202-5050, USA (e-mail: lurie@rand.org).*

REFERENCES

1. Disease Management Association of America (DMAA). 2004. Definition of disease Management. Available at http://www.dmaa.org/dm_definition.asp. Accessed February 12, 2009.
2. Todd W, Nash D, eds. Disease Management: A Systems Approach to Improving Patient Outcomes. San Francisco: Jossey Bass; 2001.
3. Ellrodt G, Cook DJ, Lee J, Cho M, Hunt D, Weingarten S. Evidence-based disease management. JAMA. 1997;278:1687-92.
4. Bodenheimer T. Disease management in the American market. BMJ. 2000;320:563-6.
5. Casalino LP. Disease management and the organization of physician practice. JAMA. 2005;293:485-8.
6. American Association of Health Plans/Health Insurance Association of America. The Cost Savings of Disease Management Programs: Report on a Study of Health Plans. Washington DC: AAHP/HIAA.; 2003. Available at: http://www.aahp.org/Content/ContentGroups/Homepage_News/DM_Short_Report.doc. Accessed February 12, 2009.
7. Villagra VG. Remarks before the US House of Representatives Ways and Means Subcommittee on Health. 7 May 2003. Available at: http://www.dmaa.org/news_releases/2003/PressRelease05072003.asp. Accessed February 12, 2009.
8. Mattke S, Seid M, Ma S. Evidence for the impact of disease management: is \$1 billion a year a good investment? Am J Managed Care. 2007;13(12):670-8.
9. Congressional Budget Office. An analysis of the literature on disease management programs. 2004. Available at: <http://www.cbo.gov/showdoc.cfm?index=5909&sequence=0>. Accessed February 12, 2009.
10. Linden A, Roberts N. A user's guide to the disease management literature: recommendations for reporting and assessing program outcomes. Am J Manag Care. 2005;11:113-20.
11. Linden A, Adams J. Evaluating disease management programme effectiveness: an introduction to instrumental variables. J Eval Clin Pract. 2006;12:148-54.
12. Beaulieu N, Cutler D, Ho K, et al. The business case for diabetes disease management for managed care organizations. Forum for Health Economics and Policy. 2006;9(1).
13. Fremont AM, Bierman A, Wickstrom SL, et al. Use of geocoding in managed care settings to identify quality disparities. Health Aff. 2005;24:516-26.
14. Elliott MN, Fremont A, Morrison PA, Pantoja P, Abrahamse A, Lurie N. A new method for estimating racial/ethnic disparities where administrative records lack self-reported race/ethnicity. Health Serv Res. 2008;43(5p1):1722-36.
15. Abrahamse AP, Morrison A, Bolton NM. Surname analysis for estimating local concentration of Hispanics and Asians. Popul Res Policy Rev. 1994;13:383-98.
16. National Committee on Quality Assurance (NCQA). HEDIS® 2004 Final NDC Lists. Available at: <http://www.ncqa.org/tabid/346/Default.aspx>. Accessed February 12, 2009.
17. Von Korff M, Gruman J, Schaefer J, Curry S, Wagner E. Collaborative management of chronic illness. Ann Intern Med. 1997;127(12):1097-102.
18. Linden A, Adams JL, Roberts N. Strengthening the case for disease management effectiveness: un-hiding the hidden bias. J Eval Clin Pract. 2006b;12:140-7.
19. Linden A, Adams JL, Roberts N. Using propensity scores to construct comparable control groups for disease management program evaluation. Dis Manag Health Outcomes. 2006c;13:107-15.
20. DMAA. Outcomes Guidelines Report Volume II. Washington, DC; 2007.