The Role of Disease Management in Pay-for-Performance Programs for Improving the Care of Chronically Ill Patients

By: Jeff Beich, Dennis P. Scanlon, Jan Ulbrecht, Eric W. Ford, Ibrahim A. Ibrahim

*** Note: Figures may be missing from this format of the document

Abstract:
To date, pay-for-performance programs targeting the care of persons with chronic conditions have primarily been directed at physicians and provide an alternative to health plan-sponsored chronic disease management (DM) programs. Both approaches require similar infrastructure, and each has its own advantages and disadvantages for program implementation. Pay-for-performance programs use incentives based on patient outcomes; however, an alternative system might incorporate measures of structure and process. Using a conceptual framework, the authors explore the variation in 50 diabetes DM programs using data from the 2002 National Business Coalition on Health's eValue8 Request for Information (RFI). The authors raise issues relevant to the assignment of accountability for patient outcomes to either health plans or physicians. They analyze the association between RFI scores measuring structures and processes, and HEDIS diabetes intermediate outcome measures. Finally, the strengths and weaknesses of using the RFI scores as an alternative metric for pay-for-performance programs are discussed.
Keywords: disease management; diabetes; pay for performance; managed care; chronic conditions

Article:
Despite substantial dollars spent on health care, shortcomings in quality and patient safety documented in the Institute of Medicine's reports Crossing the Quality Chasm: A New Health System for the 21st Century (Institute of Medicine 2001) and To Err Is Human: Building a Safer Health System (Institute of Medicine 2000) suggest that purchasers are not receiving optimal returns for their investment. In response to this concern, pay-for-performance (P4P), which rewards providers for delivering high-quality care, is becoming a widely discussed strategy for addressing these deficiencies (Ferman 2004; Rosenthal et al. 2004). While P4P strategies to date have targeted the treatment of both acute and chronic conditions, the prevalence and impact of the latter make a compelling argument for application to these diseases. Based on 1998 Medical Expenditure Panel data, it is estimated that while 45 percent of the population are afflicted with one or more chronic conditions, they account for 78 percent of health care spending (Anderson and Horvath 2004). Despite the substantial amount of resources expended on these diseases, McGlynn et al. (2003) found that adults receive, on average, only 56.1 percent of the recommended care for their conditions.

Prior to the emergence of P4P initiatives, health plans have been actively engaged in improvement efforts targeted at chronic care through the provision of disease management (DM) programs (Villagra 2004). DM is a population-based approach to the treatment of chronic illness using evidence-based clinical guidelines, multidisciplinary treatment approaches, and information systems to achieve good clinical outcomes at acceptable costs (Couch 1998). DM emphasizes the use of primary and secondary screening procedures, as well as prevention activities, to minimize the occurrence of costly and debilitating complications. Health plan–sponsored DM programs and many P4P programs targeted at physicians such as the Bridges to Excellence (BTE 2005) Diabetes Care Performance Assessment Program share a similar goal, namely, better patient outcomes. Current indicators of patient outcomes for chronic conditions in DM and P4P are composed of patient care process and intermediate outcome indicators, such as the HEDIS Comprehensive Diabetes Care measures (National Committee for Quality Assurance 2003). Achievement of good HEDIS scores requires a

combination of guideline-driven medical management by practitioners and self-management of the condition by the person with the disease (Wagner et al. 2001). A well-integrated DM program addresses both of these components.

While physician P4P and health plan DM programs share similar objectives, the emergence of these two strategies raises interesting issues in terms of accountability, the flow of dollars, and the composition of the measurement systems used for assigning rewards. If P4P is going to be used as a strategy for motivating better chronic care, does it make more sense to target the incentives at health plans, physician groups, or a combination of the two? Since the management of a chronic condition requires significant effort on the part of individuals with the disease, should patients also be eligible for incentives? Current chronic care P4P programs use patient-level measures of intermediate outcomes and generally do not take into consideration the actual systems and practice changes undertaken by clinicians to achieve these results. It has been suggested that a broad-based measure of the structure and process of DM programs may be a better metric on which to base provider payments.

One example of a broad-based measure of DM is the National Business Coalition on Health's eValue8 RFI, which scores health plans on the degree to which they have implemented successful DM programs. The eValue8 measurement system incorporates a combination of outcome measures (e.g., HEDIS results) and measures of DM processes undertaken by the plans. However, there have been no formal attempts to validate the utility of the eValue8 instrument for distinguishing among health plan–sponsored DM programs or for helping physicians or payers to understand the components of DM programs that can help them to achieve the objectives of P4P programs. Thus, it is important to understand the link between P4P and DM from multiple perspectives and to examine empirical evidence regarding this relationship.

NEW CONTRIBUTION
This article makes three significant contributions. First, it discusses a conceptual framework for understanding the link between DM and P4P in the management of chronically ill patients. The model is examined from the perspective of stakeholders in the P4P movement, including physicians and other health care providers (e.g., nurse case managers, dieticians), private (e.g., employers) and public sector health care purchasers, and risk-assuming health plans or non-risk-assuming third party administrators. Second, the article uses a unique data source to empirically assess the variation in managed care organization–sponsored DM programs, providing the first baseline information about the content and structure of DM programs across multiple organizations. For purposes of this article, we focus our analysis on the management of diabetes, although one could apply the same framework to other chronic illnesses. Third, the relationship between a broad-based measure of DM programs (e.g., the eValue8 RFI) and the currently used measures of patient "outcomes" (i.e., the HEDIS Comprehensive Diabetes Care indicators) is examined empirically. This analysis is useful for considering whether broader or more narrowly focused measures of chronic illness care are desirable in P4P programs.

CONCEPTUAL FRAMEWORK
To illustrate the relationship between DM and P4P in the context of the treatment of chronic conditions, we borrow from a model of a managed care organization–sponsored DM program (with modifications) developed by Beich (2005) depicted in Figure 1. Several sources of information were used to develop this model, including the Chronic Care Model (Wagner et al. 2001), the Disease Management Association of America (2002), DM accreditation standards from the National Committee for Quality Assurance (2001), URAC (2002) and the Joint Commission on Accreditation of Healthcare Organizations (2002), and the published and peer-reviewed DM literature.

FIGURE 1 Elements of a Disease Management Program

As Figure 1 illustrates, starting on the far right-hand side, the key objective of the DM process is to improve patient outcomes for persons with chronic conditions. As one moves left in the figure from the outcomes, the model identifies the key components and processes required to achieve these desired outcomes. Details are described elsewhere in Beich (2005), but patients with the condition (e.g., diabetes) must be identified, stratified based on illness severity, and tracked and managed appropriately. The primary objective of implementing the infrastructure depicted on the left side of this figure is to enable an optimal patient care process. Key characteristics of this process, as described in the Chronic Care Model (Wagner et al. 2001), include practitioners providing care consistent with clinical guidelines, preferably in multidisciplinary, collaborative teams, interacting with motivated patients, who take responsibility for self-management of their disease.

In terms of P4P, while Figure 1 was originally developed to describe a health plan–sponsored DM program, the components and processes are applicable to other sponsors such as physician groups, employers, or public purchasers interested in improving chronic care management. While DM is often thought of as being provided through health plans, either directly or through a DM vendor, physician groups may be able to perform as good or better, providing that they have the necessary infrastructure. Employers and government payers can opt to configure incentive programs with health plans via their DM programs, or they may find it more advantageous to contract directly with physician groups. Health plans can also structure incentive programs for participating physician groups, delegating all or a portion of the DM process. Figure 1 also identifies DM program components that might be measured in the event that an alternative to the patient-level measures currently used in P4P programs is desired.

P4P: STAKEHOLDER PERSPECTIVES
For public and private payers interested in improving the quality of care for their constituents with chronic conditions, two key issues emerge. First, who is the appropriate entity to be targeted for incentives—health plans, DM vendors, physicians, individuals with the disease, or a combination of the above? Second, should P4P rely on the commonly used patient-level "outcome" indicators, or should these programs incorporate process and structure measures? From the perspective of health plans and third party administrators, one of the key challenges is how plans can optimally influence the patient care process when plans do not directly deliver health care. For physicians or physician groups subject to P4P, their primary challenge is to implement the required infrastructure and make the appropriate changes to the patient care process to achieve the best outcomes, but there is little evidence of the relationship between specific processes and outcomes.

DATA

The data for this study were collected by the National Business Coalition on Health as part of their eValue8 Request for Information (RFI) project. The RFI is an annual survey issued to health plans providing coverage to coalition and individual employer members of the National Business Coalition on Health. It was developed with financial support and research expertise from the Centers for Disease Control and the Substance Abuse and Mental Health Services Administration. While the RFI covers a wide range of topics, several modules are devoted to DM for chronic conditions. Plan responses are scored by members of the coalition, and employers participating in this process use the results to facilitate value-based purchasing of health insurance coverage for their employees. Some employers, including the General Motors Corporation, factor the RFI scores in the establishment of employee health insurance copremium levels, with the higher scoring eValue8 plans being offered to employees at lower prices. While the primary goal of the survey is to provide comparative ratings for plans, another goal of the RFI is to establish a common set of purchaser expectations to motivate health plan quality improvement initiatives.

Plan responses to the 2002 RFI Diabetes Management module were used for this study. While the purposive sample is not representative of the universe of health plans in the United States, it consists of plans with a wide range of locations, affiliations, enrollment, and tax status. From the perspective of validity, responses from each plan are reviewed by a member of the coalition, and health plans must provide documentation (i.e., attachments) to substantiate responses to key questions. Coalition members also communicate with plan representatives to clarify and resolve uncertainties. Once the information is confirmed, scoring is performed by coalition members using a predefined scoring algorithm. Data in this study include plan responses, coalition reviewer comments, and assigned scores for the diabetes module; attachments were not available for review. A small portion of the data consisted of narrative responses requiring coding and abstracting via content analysis (Neuendorf 2002).

Two other data sources were employed including InterStudy Corporation's MSA Profiler and Competitive Edge (calendar year 2001), which was used to describe managed care organization characteristics, and the National Committee for Quality Assurance's HEDIS data (calendar years 2001 and 2002) used to assess diabetes care performance for plans in the RFI set. For this analysis, the six HEDIS Comprehensive Diabetes Care indicators (National Committee for Quality Assurance 2003), which are generally accepted measures of process and intermediate outcomes, were used. The indicators are as follows: glycated haemoglobin (HbA1c) testing (percentage with at least one HbA1c test in the current year), poor HbA1c control (percentage with HbA1c > 9.0), eye exams (percentage with annual diabetic retinal eye exam), low-density lipoprotein (LDL) screening (percentage with annual LDL-C screening), LDL control (percentage with LDL < 130 mg/dl) and monitoring nephropathy (percentage with microalbuminuria test). All indicators are rates based on the percentage of members with diabetes.

RESEARCH QUESTIONS AND METHOD

The method employed in this study was motivated by two research questions. First, what is the variation in the characteristics of managed care organization–sponsored DM programs targeted at diabetes? Second, is there a link between the eValue8 scores measuring DM program structures and processes and the outcomes achieved for chronically ill patients? The first question is assessed descriptively by examining statistics of frequency, measures of central tendency (mean), and measures of variation (standard deviation and coefficient of variation) for components of the DM model. In addition, agglomerated hierarchical cluster analysis (Everitt, Landau, and Leese 2001) was employed to identify patterns of response in the Member Interventions components.

To address the second question, we assessed the relationship between the eValue8 process scores and commonly accepted "outcome" measures for diabetes management by regressing the eValue8 2002 diabetes module process scores on each of the HEDIS Comprehensive Diabetes Care indicators, using ordinary least squares (OLS) and logistic regression. The intent was to test the ability of the 2002 eValue8 process scores (pertaining to activities conducted in calendar year 2001) to predict the 2002 HEDIS (calendar year 2001) results, the 2003 HEDIS (calendar year 2002) results, and the change in HEDIS results from 2002 to 2003, within the sample.

RESULTS
HEALTH PLAN CHARACTERISTICS
All 50 plans in the 2002 RFI database were health maintenance organizations (HMOs). The mean enrollment for commercial members was 405,857 (SD = 662,189), with a range of 8,105 to 2,794,895. Sixty percent were for-profit and the remaining 40 percent not-for profit plans. Health plan affiliation status is as follows: national managed care firms, 50 percent; Blue Cross/Blue Shield, 18 percent; and independent plans, 32 percent. Plans were distributed across the following regions of the country: Northeast, 18 percent; Southeast, 6 percent; Midwest, 52 percent; West, 14 percent; and Southwest, 10 percent. The large concentration of the midwestern health plan is a reflection of the active role of health care coalitions and purchaser members from this region.

TABLE 1 MCO Model Type: Plans in 2002 RFI Sample (in percentages)

| | | | Distribution by Enrollment | |
| | | | RFI | National[a] |
Distribution of RFI Plans	No.	%	%	%
IPA	22	44.0	34.60	40.30
Mixed	16	32.0	29.80	38.40
Network	4	8.0	27.90	11.20
Group	7	14.0	7.50	9.90
Staff	1	2.0	0.10	0.20
Total	50	100.0	100.00	100.00

Note: MCO = managed care organization; RFI = Request for Information; IPA = Independent Practice Association.
a. Source is Kaiser Family Foundation (2004).

Finally, the distribution of plans and enrollment by model type is depicted in Table 1. Compared to 2002 national HMO enrollment statistics (Kaiser Family Foundation 2004), the RFI sample is underrepresented in mixed and Independent Practice Association (IPA) models and overrepresented in network model plans.

VARIATION IN DM PROGRAMS AS MEASURED BY RFI SCORES
The RFI scoring system is derived from measures of DM program structures, processes, and outcomes, with a maximum total score of 100 points. Table 2 presents the results of the scores for the sample, broken out by the DM program components noted in Figure 1. The Quality Improvement Process section, accounting for a potential 37 points, consists entirely of outcome measures. Approximately half of these are composed of the previously described HEDIS diabetes measures. Since the intention of this article is to examine measures of the structure and process of DM programs as alternative measurements for DM P4P programs, our analysis focused on the structural and process components of the RFI.

If a purchaser was to differentiate between competing programs, they would most likely start with the aggregate measure. The mean aggregate structure and process score for this sample is 32.5 points (SD = 16.3), or 51.6 percent of the maximum possible score, with a range of 4 to 62.5 points and a coefficient of variation (CV) of 50.1. These results strongly suggest wide variation in the implementation of DM structures and processes within plans in the sample. Also, assuming that 63 points represent an ideal score, they also indicate that, on average, plans are implementing just more than half of the structures and processes that the National Business Coalition on Health deems important for DM programs.

TABLE 2   2002 RFI Diabetes Scoring Results by Disease Management (DM) Model Category (N = 50)

DM Model Component	Maximum Possible Score	Mean Score	Mean as Percentage of Maximum Possible Score	SD	Coefficient of Variation[a]	Minimum Score	Maximum Score
Structure and process							
Program philosophy and objectives	4	2.0	49.0	1.58	80.61	0	4
Evidence-based guidelines	4	2.6	65.0	1.4	53.85	0	4
Patient registry	3	1.7	56.0	0.72	42.86	0	3
Stratification	2	1.4	68.0	0.94	69.12	0	2
Member interventions	19	10.9	57.5	6.14	56.18	1	19
Practitioner interventions	31	14.0	45.0	11.29	80.87	0	30.5
Subtotal	63	32.5	51.6	16.28	50.09	4	62.5
Outcomes							
Quality improvement process	37	10.1	27.4	7.52	74.16	0	30.5
Total	100	42.7	42.7	21.29	49.88	7	84

Note: RFI = Request for Information.
a. The coefficient of variation is the (standard deviation ÷ mean) × 100. Higher values equate to greater variation.

A review of the individual component scores indicates similar patterns. Mean scores, expressed as a percentage of the maximum possible, range from a low of 45 percent for Practitioner Interventions to a high of 68 percent for Stratification. Using the coefficient of variation as the comparative metric, Patient Registry has the lowest variation with a CV of 42.86, and Practitioner Interventions have the highest with a CV of 80.87. On the surface, the variation in scores would indicate that the RFI scores provide a viable means of differentiating plans that would be amenable to P4P. However, to assess the validity of this approach, it is necessary to investigate details of the National Business Coalition on Health scoring system to assess if it measures the appropriate content and uses an appropriate metric. The remainder of this section will address scoring for each DM component, working from right to left in the model depicted in Figure 1.

Member and Physician Interventions

Member and practitioner interventions play a key role in the management of chronic conditions. In a DM program, they represent the means by which health plans attempt to improve the patient care process. The goal of member interventions is to motivate and facilitate self-management, a critical component of chronic care (Glasgow et al. 2002). Self-management occurs when patients are engaged in activities to promote health, adhere to treatment protocols, monitor their own physical and emotional status, and manage the impact of the disease on their life (Wagner, Austin, and Von Korff 1996). Given the significance of self-management, member interventions strategies should play a particularly important role both in differentiating between programs and also in deciding the most appropriate entity to provide the services.

The Member Interventions component is scored primarily on the basis of a plan's provision of the interventions noted in Table 3. Looking at intervention usage across all plans, one of the most interesting findings is that the most frequently provided intervention, printed educational materials, is generally not believed to be very effective at motivating self-management. Since plans typically provide multiple interventions, patterns were analyzed using cluster analysis (Everitt, Landau, and Leese 2001) to categorize plans into groups based on the types of interventions used. Interventions were classified as binary events (yes or no), and the simple matching coefficient was employed as the measure of similarity. Clustering was accomplished via the average linkage

method, and the four-cluster solution was selected. Table 3 indicates the frequency of intervention use, the average number of interventions used, and the average intervention score for each cluster and across all plans. Plans in Cluster 1 were the most aggressive, using more interventions than the others. In contrast, plans in Cluster 2, the largest on average, provided less than half the number of interventions, primarily relying on printed materials and reminders for services. Cluster 3 differentiates itself by heavy reliance on electronic technologies including the telephone (inbound and outbound), and e-mail/Web. Finally, Cluster 4 is characterized by the total absence of interventions; however, since plans were only credited as providing an intervention based on provision of eligibility and participation data, this absence may be partially the result of their inability to submit the necessary documentation, but is still nonetheless informative, particularly for purchasers who compare and evaluate plans.

TABLE 3   Cluster Analysis Results: 2002 RFI Member Interventions

	Cluster				All Plans
	1	2	3	4	
Plans in cluster	16	17	6	11	50
M (SD) number of interventions	5.75 (0.77)	2.4 (0.79)	4 (2.00)	0 (0.0)	4.02 (0.83)
M (SD) RFI section scores	13.28 (1.39)	5.29 (2.14)	9.67 (2.16)	0.0 (0.0)	7.21 (5.12)
Frequency of intervention use (in percentages)					
Printed educational materials	100	88	100	0	74
Patient reminder notices	100	76	83	0	68
Education programs	100	35	0	0	44
Diabetes related home visits	63	6	0	0	22
Inbound phone services	50	6	100	0	30
Outbound phone calls	88	18	83	0	44
Web/e-mail services	38	6	100	0	26
Other	38	6	0	0	14

Note: RFI = Request for Information.

On the positive side, the RFI assesses the most commonly used member interventions and assigns an appropriate relative weighting to each. However, given the importance of member interventions, the maximum availability of 19 points for this component is not sufficient. Realistically, it should be allocated 30 to 35 points. Perhaps most important from the purchaser's perspective, the scoring system fails to capture differences in intervention content, the quality of delivery, and participation rates. Regarding the latter, while plans must provide eligibility and participation rates to receive credit, scores are assigned on the provision of the data rather than on the level of participation. Therefore, a plan providing an intervention to 5 percent of its eligible population receives the same score as one delivering it to 20 percent of its eligible population. Using this component score as a means for purchasers to differentiate programs might lead to erroneous conclusions since a plan providing multiple interventions with mediocre implementation and minimal participation would receive a higher score than a plan that delivers limited, but well-executed interventions to the majority of its members with the disease.

From the perspective of the P4P issue of who should be incentivized, the provision of member interventions is a resource-intensive process. Successful strategies for facilitation of patient self-management require frequent and ongoing collaboration between practitioners and patients to elicit patient behavior change (Norris, Engelgau, and Narayan 2001). While the relative effectiveness of individual and specific combinations of interventions to motivate self-management has yet to be proven empirically, multifaceted interventions appear to be more effective than single interventions (Shojania et al. 2004). Time demands and training issues suggest that physicians may not be the best individuals to provide self-management training, and the process may be more effectively implemented through the use of nurses, dieticians, and other practitioners (Larme and Pugh 1998).

This would suggest that optimal delivery of member interventions would be cost prohibitive for solo and small group practitioners and might be more appropriately provided by health plans. On the other hand, the introduction of the health plan as a third party in the patient care process presents communication and coordination issues that might impede collaboration.

In the context of DM, practitioner interventions are actions taken to encourage and facilitate adherence to clinical guidelines. The Practitioner Interventions component scores are primarily determined by the plan's provision of provider guideline adherence reports. Sixty-eight percent of plans reported that they provide diabetes guideline adherence reports to primary care physicians, comparing their performance to peers and/or benchmarks. The RFI scoring system differentiates plans based on the number of indicators reported, frequency of reports, inclusion of benchmarks, and categories of physicians receiving reports. There was variation in these report attributes among the plans in the sample; however, interpretation of this variation yields little in the way of practical information for differentiating plans. Even when the responses were analyzed with cluster analysis, no clear patterns emerged. What the scores fail to reveal is whether the physicians read the reports and whether they modify their practice patterns on the basis of the feedback.

From the perspective of P4P, one of the interesting findings in this section is that 30 percent of the plans in the sample provided financial incentives to physicians for guideline adherence. While details of the incentive programs were not available, comments by the scorers suggest that most plans incorporate diabetes indicators along with other indicators in their compensation program, rather than providing a specific diabetes care bonus. It is also not feasible to determine if the dollars are sufficient to motivate behavior change. However, this raises interesting options for the use of P4P. Even if purchasers decided to incentivize health plans, rather than providers, plans may choose to implement their own P4P program for their provider network.

For a purchaser, the results of the scores in the Practitioner Interventions component would probably not be that useful for either discriminating between programs or for use as a P4P metric. First, the section is overweighted at 31 points, which is almost one half of the RFI structure and process scores. While practitioner interventions are important, a more appropriate weight would be 15 to 20 points. Second, the section focuses too heavily on guideline reports at the expense of other interventions. In general, guideline adherence reports have been found to be one of the less effective interventions for motivating changes in physician behavior (Stone et al. 2002).

Patient Registry, Stratification, and Evidence-Based Guidelines

In a DM program, the patient registry captures key clinical and administrative data for members identified with the disease to enable tracking over time (Metzger 2004). The registry typically provides the information required for stratification and can also facilitate clinical decision making by highlighting needed or overdue services. The RFI scores for the Patient Registry component are primarily related to patient identification. To receive full credit for this component, a plan must identify at least 5 percent of its members as having diabetes. The mean prevalence of diabetes identified by the plans in the sample was 3.6 percent (SD =1.01 percent), with only 8 percent meeting the 5 percent threshold. Absent information on plan distribution of enrollee age and race/ethnic composition, it is not feasible to calculate precise expected prevalence rates. Nonetheless, based on the CDC's (2004) estimates of national prevalence, plans in the sample appear to be significantly under-identifying members with diabetes.

The ability to identify persons with the disease may be an important factor in deciding who to hold accountable for chronic disease management. In order for plans to do this, they must rely either on claims, encounter, pharmacy, and laboratory data or on referrals to the DM program from practitioners. Our findings suggest that plan systems for patient identification do not function that well. Practitioners are more likely to be aware of the patient's condition as a result of ongoing encounters, access to the medical record, and diagnostic test results. However, compared to practitioners, a health plan is more likely to have the resources to implement a sophisticated registry system.

Patient stratification is used to divide members with chronic conditions into categories based on severity and/or the probability of experiencing acute complications, and it is also used to prioritize interventions based on patient needs. RFI scoring for the Stratification component is contingent on the plan's provision of data about the number of strata, the criteria used for stratum assignment, and the distribution of members among these categories. Major findings from the data are that 84 percent of the plans stratify their patients, and with the exception of standardized systems in plans affiliated with national managed care firms, they all do it very differently. Given the lack of published empirical data on the relative effectiveness of different approaches in the stratification process (Cousins, Shickle, and Bander 2002), it is not feasible to make definitive statements about the meaning of this variation.

Evidence-based clinical guidelines are key to the DM process, providing the roadmap for the clinical management of persons with the disease (Weingarten and Graber 1998). Scoring of the Evidence-Based Guidelines component is contingent on the presence or absence of guidelines, participation in regional collaborative efforts to standardize guidelines among competing plans, and provision of education to physicians on guideline topics. Realistically, the last factor is more of a practitioner intervention than a guideline issue, complicating interpretation of the RFI's components scores.

The major findings related to the guideline component are that 98 percent of the plans had implemented diabetes guidelines, and 54 percent participated in some form of collaborative guideline effort. Since virtually all plans in the sample report guidelines in place, an indicator of their presence or absence does not provide a lot of information to distinguish program quality (except in the case of the one plan in the sample that reported not having formally adopted guidelines). Also, given the availability of diabetes guidelines, at no cost, from sources such as the American Diabetes Association (2003) and the Institute for Clinical Systems Improvement (2001), it is easy for even solo practitioners to obtain a copy for their practice.

For the purchaser, a general observation of the scoring criteria for the Stratification, Patient Registry, and Evidence-Based Guidelines components is that the questions primarily target structures and do not get at the functionality of the processes. The ideal DM program would incorporate key guideline requirements into the registry system, use the registry information to feed the stratification process, and employ the registry and stratification information to facilitate decision support by practitioners (e.g., provision of reminders for needed services). While the RFI questions and scores related to these components may be effective at ruling out programs that do not have key structures (e.g., the plan without guidelines), the scores fail to provide information that discriminates a program's ability to integrate these components into well-functioning programs.

The final component, Program Philosophy and Objectives, is a miscellaneous category with scoring contingent on the cultural and linguistic appropriateness of services and the provision of innovative services. In general, plans performed well on the first topic, with a mean score of 83 percent of the maximum available points. The Innovations section allows for open-ended descriptions of unique aspects of programs not covered in other sections of the RFI. While 40 percent of the plans received full or partial credit for providing innovative services, descriptions were not available for review. In terms of plan discrimination, the scores in this section should be interpreted with caution since the assignment of scores for innovations is up to the judgment of the RFI scorers.

The RFI scoring system provides a means of rank ordering plans on the basis of an assessment of DM structures and processes. While this system has the potential to be used as a P4P measurement device, its practical use is contingent on the validity of the metrics. One of the challenges facing DM is the lack of conclusive empirical studies indicating the factors that contribute to effective functioning of each component. For example, despite multiple studies, we still do not have a clear picture of the most effective intervention strategies for motivating changes in either physician or patient behavior (Shojania and Grimshaw 2005). Given the absence of good empirical data, the RFI scoring system was developed primarily through expert consensus. Ultimately, if the RFI structure and process scores are to be used to discriminate between plans, either for value-based purchasing

or P4P, the scores should positively correlate with patient outcomes. In the next section, we assess this correlation.

CORRELATION BETWEEN RFI STRUCTURE AND PROCESS SCORES AND HEDIS RESULTS

To test the relationship between the RFI scores and patient outcomes, we used the HEDIS Comprehensive Diabetes Care indicators as the outcome measures. We regressed each of the HEDIS indicators for calendar years 2001 and 2002, and the change in results between the two years, on the aggregate 2002 RFI structure and process scores (covering calendar year 2001) using OLS regression. Both years were assessed since there is likely a lag between the delivery of DM services and their effect on patient outcomes. Within the models, we hypothesized that there would be a positive and significant association between the RFI scores and the HEDIS results.

The results of these regressions are summarized in Table 4. Rather than report the coefficient estimates themselves, we report the estimated effect of a 10-point difference in the RFI score on the HEDIS indicator results, the adjusted R^2 value, and the significance of the estimated effects. There are three key findings from these results. First, more than half of the regressions resulted in insignificant coefficients, indicating that the estimated effect is not statistically different from zero, implying the absence of a relationship. Second, even among the significant models, the effects are very small. For example, in the 2002 Monitoring Nephropathy indicator, while the effect was statistically significant, a 10-point increase in the RFI process score equates to just a 1.71 percentage point increase in the HEDIS result. Finally, the adjusted R2 values ranged from .0652 to 0.12, indicating that the combined scores explain very little variance in the HEDIS results within the sample. While the results are not reported in this article, we also employed logistic regression to test the ability of the RFI scores to predict the probability of a plan being in either the top or bottom quartile (nationally) of the HEDIS results in calendar years 2001 and 2002. Our findings were similar to the OLS models, namely, a high percentage of insignificant coefficients, and low effect sizes in the models with significant results. Finally, we also regressed the HEDIS measures on the individual structure and process component scores listed in Table 2, using multivariate OLS and logistic models. The findings failed to reveal a consistent relationship between any of the individual component scores and the HEDIS measures.

Taken together, our regression analyses suggest that there is not a strong or consistent relationship between the HEDIS results and the RFI scores either at a point in time or when considering changes over time. Because of limitations in degrees of freedom, our models might not have been able to detect the effect if there was a small or modest correlation. It is also possible that the effect varied by types of plans or market characteristics, and since we were not able to control for these other possibly important covariates because of degrees of freedom constraints, we cannot rule this out. Still, if there was a very strong correlation between the two, whether positive or negative, our simple regressions would show significance.

DISCUSSION

Our analysis of the National Business Coalition on Health's eValue8 RFI data suggests there is significant variation in managed care organization– sponsored diabetes DM programs, both within the individual program components and in the degree to which the managed care organizations are able to effectively structure the components into viable programs. Despite this variation however, the structural and process scores from the RFI were not substantively related to the HEDIS measures that are currently being used to assess health plan diabetes care performance. This begs the question of why there is not an observed relationship between DM program structure/process and chronic care outcomes and what implications the absence of such a relationship has for the development of P4P programs and the stakeholders of these programs.

TABLE 4   Regression Results: *HEDIS* Comprehensive Diabetes Care Measures on 2002 RFI Aggregate Structure and Process Scores[a]

	HbA1c Testing		Poor HbA1c Control[b]		Eye Exams		LDL Screening		LDL Control		Monitoring Nephropathy	
	Effect[c]	Adj. R^2	Effect	Adj. R^2	Effect	Adj. R^2	Effect	Adj. R^2	Effect	Adj. R^2	Effect	Adj. R^2
HEDIS 2002[d]	−0.708	.012	−2.33**	.120	1.170	.007	1.280	.052	1.510*	.089	1.713*	.098
HEDIS 2003[e]	1.262*	.065	−0.156	.042	2.64*	.090	0.940	.043	1.170	.030	3.115*	.111
Change in *HEDIS*[f]	2.026*	.0838	0.9586	.0126	1.721*	.0980	−0.433	.0027	0.6485	.0072	−0.042	.0108

Note: RFI = Request for Information; LDL = low-density lipoprotein; HbA1c = glycated haemoglobin.
a. 2002 RFI aggregate structure and process scores reflect activities occurring in calendar year 2001.
b. For poor HbA1c control, a lower value is better.
c. Effect indicates the impact of a 10-point difference in RFI scores on *HEDIS* results.
d. 2002 *HEDIS* data reflect activities occurring in calendar year 2001 (*N* = 48).
e. 2003 *HEDIS* data reflect activities occurring in calendar year 2002 (*N* = 46).
f. Change is measured by subtracting 2002 results from 2003 results from 2003 (*N* = 46).
*$p < .05$. **$p < .01$.

In many respects, the absence of a relationship between the eValue8 RFI and the HEDIS diabetes results is not surprising given the challenge of using a survey tool to systematically assess the structure and process of DM programs. There are practical limitations on both the number and scope of questions that can be included in the RFI, making it difficult to cover each component of the DM model in sufficient detail to render the results scientifically valid. A significant reason for the limitation pertains to the ability to make distinctions based on the content, breadth, or the effectiveness of program implementation. For example, member interventions were scored as binary "yes" or "no" events, rendering mediocre and excellent interventions as equivalent in the eValue8 scoring system. Decisions regarding the assignment of weights to various components of DM programs, while necessary for a scoring system, can also serve to generate arbitrary variation in the overall assessment of DM programs. Finally, despite the fact that the RFI includes a data auditing function, our analysis

also found concerns with the quality of voluntary self-reported data submission in the form of missing or implausible data.

Regarding the issue of who should be held accountable (and incentivized) for chronic care outcomes, this analysis raises more questions than answers. The conceptual framework noted in Figure 1 makes it clear that optimal management of persons with chronic conditions is a complex process requiring significant infrastructure. While health plans are likely to have more resources to devote to infrastructure than physician groups, plans also have a more difficult task in terms of coordinating this process among multiple independent physicians and physician groups.

While there is evidence that health plan–sponsored DM programs can be effective at improving patient outcomes (Fireman, Bartlett, and Selby 2004; Villagra and Ahmed 2004), DM has been described as a "red pill" that seems to work, but the ingredients are uncertain (Shojania and Grimshaw 2005). Quite simply, while there is strong belief that the Chronic Care Model and DM are effective for improving chronic illness care, there is little evidence about the most influential or important components that lead to improved chronic care outcomes (Villagra 2004) and little evidence regarding which stakeholders should assume responsibility for implementing the various components of DM. This fact, coupled with the documented variance in DM program design, as well as the increasing prevalence of P4P programs, suggests the need for carefully designed data collection and research that demonstrates the effectiveness of DM program components so as to provide physicians, health plans, and purchasers with the information necessary to improve chronic care delivery and to achieve P4P program objectives.

Because many factors can influence chronic care outcomes, not the least of which is patient self-management and patient compliance, there is good reason in theory to measure the effectiveness of DM structure and process rather than outcomes. In practice, however, the complexity of measuring DM structure and process, and the lack of empirical evidence regarding the optimal characteristics of these structures and processes, suggests that standardized intermediate outcome and process measures such as those captured by HEDIS are probably the best targets for P4P programs, regardless of the entity being held accountable. Nonetheless, the eValue8 RFI provides a starting point for thinking about how to advance the concept of a broad measure of DM program quality to the point where it is possible that DM structure and process measures might replace or augment the measures that are currently being used in P4P programs.

There are two important limitations of our study. First, the RFI sample was not randomly selected, and therefore, results from our analyses are not generalizable to the entire population of managed care organizations. Second, even though our data represent the largest sample of health plan–sponsored DM programs that we are aware of, it is small enough that it does not provide much statistical power to determine effect sizes.

While the prevalence of health plan–sponsored DM programs has grown substantially since the late 1990s and appears to be yielding positive clinical results, it is still uncertain if DM has been successful in reducing health costs. The emergence of P4P programs incentivizing physicians provides another alternative strategy for improving chronic care delivery. It is reasonable to postulate that P4P programs have the potential to motivate either health plans or physicians to take more active roles in chronic care management. Both approaches have strengths and weaknesses, and it is likely that both will emerge as viable alternatives, at least until more definitive evidence becomes available to distinguish the most effective manner for encouraging improvements in the care of chronically ill patients.

REFERENCES

American Diabetes Association. 2003. Standards of medical care for patients with diabetes mellitus. Diabetes Care 26 (Suppl. 1): S33-50.

Anderson, G., and J. Horvath. 2004. The growing burden of chronic disease in America. Public Health Reports 119 (May/June): 263-70.

Beich, J. 2005. Diabetes care improvement initiatives in managed care organizations. University Park: Health Policy and Administration, Pennsylvania State University. http://etda.libraries.psu.edu/theses/approved/WorldWideIndex/ETD-1022/

Bridges to Excellence. 2005. Bridges to excellence: Overview. http://www. bridgestoexcellence.org/bte/bte_overview.htm (accessed August 3, 2005). Centers for Disease Control and Prevention. 2004. National diabetes fact sheet: General information and national estimates on diabetes in the United States 2003. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention. http:// www.cdc.gov/diabetes/pubs/estimates.htm#prev3 (accessed November 14, 2004).

Couch, J. B. 1998. Disease management: An overview. In The health care professional's guide to disease management, edited by J. B. Couch. Gaithersburg, MD: Aspen. Cousins, M. S., L. M. Shickle, and J. A. Bander. 2002. An introduction to predictive modeling for disease management risk stratification. Disease Management 5 (3): 3-27.

Disease Management Association of America. 2002. DMAA definition of disease management. http://www.dmaa.org/definition.html (October 29, 2002).

Everitt, B. S., S. Landau, and M. Leese. 2001. Cluster analysis. 4th ed. London: Arnold. Ferman, J. H. 2004. Pay for performance gaining popularity. Healthcare Executive 19 (4): 50-52.

Fireman, B., J. Bartlett, and J. Selby. 2004. Can disease management reduce health care costs by improving quality? Health Affairs (Millwood) 23 (6): 63-75.

Glasgow, R. E., M. M. Funnell, A. E. Bonomi, C. Davis, V. Beckham, and E. H. Wagner. 2002. Self-management aspects of the improving chronic illness care breakthrough series: Implementation with diabetes and heart failure teams. Annals of Behavioral Medicine 24 (2): 80-87.

Institute for Clinical Systems Improvement. 2001. Management of Type 2 diabetes mellitus. http://www.icsi.org/knowledge/detail.asp?catID=29&itemID=182 (accessed February 14, 2003).

Institute of Medicine. 2000. To err is human: Building a safer health system. Washington, DC: National Academy Press.

-----. 2001. Crossing the quality chasm: Anew health system for the 21st century. Washington, DC: National Academy Press.

Joint Commission on Accreditation of Healthcare Organizations. 2002. Disease-specific care certification manual. Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations.

Kaiser Family Foundation. 2004. Trends and indicators in the changing health care marketplace. http://www.kff.org/insurance/7031/index.cfm (accessed November 22, 2004).

Larme, A. C., and J. A. Pugh. 1998. Attitudes of primary care providers toward diabetes: Barriers to guideline implementation. Diabetes Care 21 (9): 1391-96.

McGlynn, E. A., S. M. Asch, J. Adams, J. Keesey, J. Hicks, A. DeCristofaro, and E. A. Kerr. 2003. The quality of health care delivered to adults in the United States. New England Journal of Medicine 348 (26): 2635-45.

Metzger, J. 2004. Using computerized registries in chronic disease management. Oakland: California Healthcare Foundation.

National Committee for Quality Assurance. 2001. 2002 disease management standards and guidelines for accreditation. Washington, DC: National Committee for Quality Assurance.

----. 2003. The state of health care quality. http://www.ncqa.org/sohc2003/comprehensive_diabetes_care.htm (accessed November 22, 2004).

Neuendorf, K. A. 2002. The content analysis guidebook. Thousand Oaks, CA: Sage.

Norris, S. L., M. M. Engelgau, and K. M. Narayan. 2001. Effectiveness of self-management training in Type 2 diabetes: A systematic review of randomized controlled trials. Diabetes Care 24 (3): 561-87.

Rosenthal, M. B., R. Fernandopulle, H. R. Song, and B. Landon. 2004. Paying for quality: Providers' incentives for quality improvement. Health Affairs (Millwood) 23 (2): 127-41.

Shojania, K. G., and J. M. Grimshaw. 2005. Evidence-based quality improvement: The state of the science. Health Affairs (Millwood) 24 (1): 138-50.

Shoj ania, K. G., K. M. McDonald, R. M. Wachter, and D. K. Owens. 2004. Closing the quality gap: A critical analysis of quality improvement strategies. Volume 2. Diabetes mellitus care. Stanford University-UCSF Evidence-based Practice Center. http://www.ahrq. gov/downloads/pub/evidence/pdf/qualgap/front.pdf (accessed March 5, 2005).

Stone, E. G., S. C. Morton, M. E. Hulscher, M. A. Maglione, E. A. Roth, J. M. Grimshaw, B. S. Mittman, L. V. Rubenstein, L. Z. Rubenstein, and P. G. Shekelle. 2002. Interventions that increase use of adult immunization and cancer screening services: A meta-analysis. Annals of Internal Medicine 136 (9): 641-51.

URAC. 2002. Accreditation program guide for disease management standards: Version 1. Washington, DC: URAC.

Villagra, V. 2004. Strategies to control costs and quality: A focus on outcomes research for disease management. Medical Care 42 (4, Suppl.): III24-30.

Villagra, V., and T. Ahmed. 2004. Effectiveness of a disease management program for patients with diabetes. Health Affairs 23 (4): 255-66.

Wagner, E. H., B. T. Austin, C. Davis, M. Hindmarsh, J. Schaefer, and A. Bonomi. 2001. Improving chronic illness care: Translating evidence into action. Health Affairs 20 (6): 64-78.

Wagner, E. H., B. T. Austin, and M. Von Korff. 1996. Organizing care for patients with chronic illness. Milbank Quarterly 74 (4): 511-44.

Weingarten, S. R., and G. Graber. 1998. Outcome-validated clinical practice guidelines: A scientific foundation for disease management. In The health professional's guide to disease management, edited by J. B. Couch. Gaithersburg, MD: Aspen.