

Disease Management: Using Standards and Information Technology to Improve Medical Care Productivity *

Paul Gertler

University of California at Berkeley, Haas School of Business and NBER

Tim Simcoe

University of California at Berkeley, Haas School of Business

DRAFT March 2004

*Thanks to Derek Newell, Yang Gao, Alex Exuzides and LifeMasters Supported Self Care, Inc. for providing data and insights into LifeMasters' disease management program. We also received useful comments from Florian Zettlemeyer, Meghan Busse, and Catherine Wolfram. Address for correspondence: Haas School of Business, UC Berkeley, Berkeley CA 94720-1900. E-mail: simcoe@haas.berkeley.edu

Disease Management: Using Standards and Information Technology to Improve Medical Care Productivity

Abstract

Patients with a chronic illness (such as diabetes or congestive heart failure) are one of the costliest and fastest growing populations within the U.S. health care system. Disease management programs use clinical standards and information technology to identify high-risk patients within this population and intervene before expensive treatments become necessary. The goal of DM programs is to improve the long-term health through education and preventive care. Despite the growing popularity of these programs, few studies have demonstrated that they actually change patient behaviors, improve health outcomes, or reduce costs. This paper describes the emerging market for disease management, and examines the impact of a disease management program for diabetics at a central Massachusetts HMO. We find evidence that the program led to increased compliance with Clinical Practice Guidelines (CPGs), improvements in patient health, and reductions in the total cost of care.

1 Introduction

Patients with a chronic illness (such as asthma, congestive heart failure, or diabetes) account for about three-quarters of the \$1.6 trillion spent on medical care in the United States (Hoffman and Rice 1995; Levit, Smith, Cowan et al 2004). As the population ages, and the obesity epidemic expands, the number of chronically ill and the cost of caring for them will grow dramatically (Sturm, Rongle and Andreyeva, 2004). The size and scope of this problem has led to a broad search for methods of improving chronic care while reducing overall costs. Disease management (DM) is one potential solution (Institute of Medicine, 2001). DM tries to reduce expenditures over the long-run by preventing the onset of complications that frequently lead to expensive hospitalizations and procedures. In this paper, we describe the recent rise of DM within the health care industry, and estimate the impact of DM on medical care productivity—measured as a reduction in average costs accompanied by improvements in health outcomes.

DM is a relatively simple technology that combines information monitoring and communication with interventions designed to change patient behavior. DM uses information technology to identify patients who are chronically ill, to monitor their compliance with clinical practice guidelines (standards of care), and to alert both patient and physician to problems needing medical intervention. DM also tries to remove the two major impediments to patients following prescribed treatment regimes and changing their lifestyles when they are diagnosed with a chronic disease. The first impediment is that patients may not have the necessary information about the costs and benefits of changing their behavior. The second impediment is the psychological barrier to change. DM uses nurse practitioners to educate patients about their condition, and to help them comply with self-care treatment regimes (e.g. monitoring vital statistics, taking medication, diet, and exercise).

The trend towards DM signals a remarkable change in the organization and delivery of medical services. One of DM's innovations is to shift the locus of intervention from physician to patient. Medical care productivity depends upon the joint effort of patients and physicians. However, the prevailing view has been that the primary responsibility for delivering treatment lies with the physician. DM programs aim to improve productivity by increasing the quantity and quality of patient involvement. This shift has occurred in other industries, such as retailing, where IKEA revolutionized furniture-selling by making the consumer responsible for transportation and assembly. Similarly, gas stations moved to self-service in the 1970s, and education has long relied on the input of parents and children. In order to work with chronically ill patients more directly, DM also shifts much of their care into centralized programs that are run or contracted by insurance companies. In the past, insurance companies simply provided financing. Even most HMOs have primarily contracted with independent medical providers

and used financial incentives such as capitation and utilization review to control costs. Only a small share of the HMOs are fully integrated insurance and provider organizations, and these organizations were among the early adopters of DM in the late 1990s.

While DM is likely to be most effective under the management of physicians, few physician groups have implemented DM (Shortell et al 2003). Rather, DM is rapidly being adopted by both private and public insurance companies (Felt-Lisk and Mays 2002; Foote 2003). In 2002 alone, private insurance companies contracted with DM companies to provide services valued at over \$600 million.¹ Currently, Insurance companies, through either direct provision or through contracting, provide DM to more than a million chronically ill patients. Moreover, half of the Medicaid programs have adopted some type of disease management. Finally, as part of the “Medicare Prescription Drug Act of 2003,” Congress authorized a \$100 million DM pilot program, to evaluate whether Medicare should offer DM to its 35 million beneficiaries.

These facts lead to a number of important questions. First, why are we seeing the adoption of a relatively simple technology like DM now? Second, why are insurance companies adopting DM rather than medical care providers? Finally, what is the effect of DM implemented by insurance companies on medical care productivity measured by costs and outcomes? Indeed, while a growing practitioner literature provides some indication of the potential benefits of DM, there have been to our knowledge no rigorous evaluations of the impact of a large scale DM intervention.

This paper begins to address this gap by evaluating the impact of a large DM intervention for diabetics. The data come from a population of more than 6,000 patients of a central Massachusetts Health Maintenance Organization, who were eligible to enroll in a DM intervention administered by a third-party provider. Our results show that patients who enrolled in DM became more likely to comply with clinical practice guidelines. We also find significant improvements in their health, measured in terms of blood sugar levels, hospitalizations, and mortality. Finally, we find statistically and economically significant reductions in the overall cost of care in less than one year’s time.

The remainder of the paper is organized as follows. Section 2 provides an overview of disease management and discusses its recent emergence as a potential solution to the growing costs of chronic care. That section also discusses why disease management is being adopted now, and by insurance companies rather than physicians. Section 3 describes the particular DM intervention for diabetes that we study in this paper. Section 4 discusses our methodology, Section 5 presents the results, and Section 6 offers several conclusions.

¹<http://www.healthaffairs.org/press/julyaug0305.htm>

2 The Organization of Health Care and The Rise of DM

Disease management offers a combination of clinical decision support to physicians and tools for self-care to patients. A DM program typically starts by trying to identify patients who might benefit from enrolling. These patients are those diagnosed with a chronic illness and who are not following recommended practices for self-care (e.g. taking their medication). However, if eligibility thresholds are set too low, costs become prohibitive. When they are set too high, the benefits are not large enough. Next, DM uses IT to monitor compliance with self-care and treatment regimes. Conditional on physician permission, patients are asked to enroll in the program. Those who accept, usually participate in an educational session designed to familiarize patients with their condition, and suggest potential improvements in lifestyle. Patients are also provided with a variety of tools for taking care of themselves, such as reference materials, testing and monitoring equipment, and access to a nurse-staffed call center. Regularly scheduled contact with nurse-coaches helps the DM provider monitor each patient and provide them with reminders about therapy compliance or warnings about potential problems. Finally, DM coordinates its activities with physicians, primarily by sharing information obtained from data on compliance monitoring outcomes and regular interactions with the patient.

The success of DM depends critically on successful information management, which requires three things: established standards for measuring therapy compliance, systems for collecting and analyzing patient data, and analytical tools that can assess the health and compliance of chronically ill patients. Two recent advances have created the conditions necessary for development of DM. The first is the development of Clinical Practice Guidelines (CPGs) for the treatment of chronic diseases. The second is the widespread adoption of increasingly sophisticated IT systems for tracking patients.

CPGs are a set of reference standards or “best practices.” They are an important innovation because they can be used to monitor therapy compliance and to identify patients who are likely to benefit from an intervention. The guidelines provide a baseline that can be used to identify patient- or physician-specific deviations from a recommended course of treatment. The guidelines were created following research the documentation of substantial geographic variation in clinical practice and health outcomes that could not be explained by differences in epidemiology or the demographics of local populations (Wennberg 1973). That is to say that the same patient with the same problem would be treated differently depended on where he or she was treated. The development of CPGs represents an attempt to remedy the situation by providing “evidence-based” methods for maximizing the probability of achieving the best possible outcome.² However, studies found that CPG availability had little or no impact on

²The work of creating these standards was carried out in the mid 1990s under the auspices of the National

medical practice, primarily because *physician* oriented interventions did not lead to increases in *patients'* compliance with the recommended guidelines (Iliadis 1999; Grimshaw 1999). The idea behind DM is to work directly with patients to increase their probability of CPG compliance.

While the development CPGs was important, a critical question was how the CPGs could be used to monitor compliance. Many of the CPGs cover diagnostic tests and drug therapy, which rely not only on a physician prescribing them, but also patients actually following physician instructions. Application of CPGs for monitoring requires information on patient compliance behavior. This information does not exist in physician offices, but does exist in insurance company databases. Because insurance companies pay claims, they have data on diagnostic tests it paid for and drug prescriptions filled.

Investments in information technology (IT) played a key role in DM's ability to use insurance company claims data to monitor CPG compliance. Advances in processing power, storage technology, and database software have made it possible to access and analyze the large quantities of information residing in most health plans' computerized information systems. DM providers have developed a variety of systems that incorporate the rules and guidelines provided by CPGs into a set of tools for monitoring behaviors and outcomes, and distributing information to patients and providers. Ideally, these systems can access patient-level clinical data, apply decision support tools to that information, and generate recommendations in a manner that is acceptable to patients and/or physicians. These analytical tools are complemented by a set of technologies that are used to manage patient relationships in an integrated fashion across call-centers, online channels, and face-to-face interactions.³

However, in order to earn a financial return on their fixed IT development costs, DM providers need to exploit scale and enroll a large number of patients. These costs are high for a general practitioner who treats only a handful of patients with any particular chronic disease. However, managed care organizations and traditional insurers often have both the internal IT capabilities and the scale required to benefit from a DM intervention. Indeed, many of the early adopters of DM have been large insurers, state health plans, or large managed care organizations.

The incentives created by the traditional health-care delivery model pose a challenge for DM. Since the cost savings from are achieved through preventing complications that lead to hospitalizations and complex procedures, physicians have little financial incentive to adopt DM. Physicians are compensated for primary care services and would not benefit from reduced

Commission on Quality Assessment, the National Institutes of Health and the Institute of Medicine. These organizations convened panels of experts who reviewed the scientific literature to identify a set of treatment guidelines.

³The underlying technology is quite similar to that deployed in Customer Relationship Management (CRM) applications by large marketing operations.

hospitalizations. However, insurance companies do benefit from keeping their beneficiaries healthy and from reduced hospitalizations.

Thus, in the late 1990s the development of CPGs (standards) and sophisticated IT set the stage for insurance companies to adopt IT. Insurance companies had the databases, the financial incentives, and the scale to make DM profitable. On the other hand, physicians, who might most productively use DM, had neither the scale nor the databases to easily implement DM and no financial incentive to make large fixed investments in the technology required to run a program. Hence the stage was set for insurance companies, and not physicians, to adopt DM.

However, because of beneficiary turnover, insurance companies often need to achieve financial returns quickly to benefit from DM. On average beneficiaries switch insurance companies every two years. Hence the returns to prevention must accrue within two years. Only when cost savings materialize quickly or enrolled patients are likely to be retained do the long-run benefits of preventive care accrue to the organization that invested in providing those services.

While the enabling technology was available in the late 1990s, it was a change in financial incentives that prompted insurance companies to adopt DM. In the late 1980s and 1990s, most insurance companies expanding their managed care programs. Managed care controlled costs by paying physicians capitated payments-i.e. a fixed payment per person per month for all care. Physicians then were able to keep the difference between the capitation rate and the costs of care. In essence, insurance companies shifted the risk to providers. In fact, capitation led to a levelling of medical care cost inflation in the mid 1990s. However, both physicians and patients grew tired of the restrictions and incentives in capitation and rebelled against it (Blenden et al 1998). As a result insurance companies backed off of capitation and medical costs began to rise quickly again around the turn of the century. At this point insurance companies began to search for other mechanism to stem rapidly rising medical care costs and they began to examine DM as an alternative.

The use of IT and CPGs to establish large-scale DM interventions that address these issues is a recent phenomenon. Our research is motivated by the fact that—despite the increasing level of interest in these programs—there is little empirical evidence on their effectiveness at changing behaviors, improving the quality of care, or reducing costs. In the next section, we describe the institutional setting, patient population, and DM intervention for our study.

3 Data

Our study examines the impact of a DM program for diabetes at the Fallon Community Health Plan, a large central Massachusetts HMO. This program provides a unique opportunity to

evaluate the potential impact of DM on medical care productivity because of its size and duration. We follow more than 6,000 patients over a four-year period beginning in September 1998—one year before the start of the intervention. The program was managed by LifeMasters, a private company that specializes in chronic care. Through LifeMasters, we have obtained data on patient demographics, medical claims, prescriptions, and lab test results. In this section, we discuss the DM program for diabetes and describe our data set.

Fallon Community Health Plan provides health insurance for more than 214,000 members, and operates a large a large group medical practice.⁴ While about two-thirds of Fallon’s members receive their primary care through the Fallon Clinic, the remaining one-third are insured by Fallon, but participate through an affiliated network of independent providers. In May 1999, Fallon contracted with LifeMasters to introduce a disease management program for diabetes.

At the start of the intervention, LifeMasters identified patients with diabetes, and grouped them into three categories based on historical claims data. The medium and high-risk patients eligible to enroll in the program met one of the following criteria: hospitalization for a diabetes-related complication, an HbA1c level greater than 9.5, or an established diabetic complication such as retinopathy, neuropathy, nephropathy, amputation, or skin ulcers.⁵ Hemoglobin A1C, or HbA1c, is the most important measure of a patient’s control over their diabetes. The test measures long-term variability in a patient’s blood sugar level. An HbA1c score less than 7.0 is considered good control, while a score greater than 8.0 suggests that a patient has poor control over their condition.

In May 1999, LifeMasters began to ask physicians for permission to enroll their patients in the DM program. With the explicit support of Fallon’s management, the resistance by physicians was minimal. Their initial acceptance rate was 82 percent, and it increased rapidly as the program was introduced. In September, the program started to enrol patients. Enrollment was conducted by mail and telephone, with LifeMasters attempting to contact eligible patients up to five times. Many patients declined for personal reasons. Others did not enroll because they could not be reached—usually because of an incorrect telephone number. Over the course of the study, 27 percent of the eligible patients decided to enroll in the DM program.

When a patient did choose to enter the program, they were immediately scheduled for a series of educational sessions. These could be taken either in person or over the phone. The lessons provided basic information about living with diabetes, and also explained how to use the resources available through the program’s call center and web site. After completing

⁴At the beginning of our study, Fallon’s membership consisted of 166,000 commercial members, 36,000 Medicare+Choice participants, and 12,000 on medical assistance.

⁵Patients in the low-risk category received educational mailings and promotional materials.

the classes, enrolled patients could call in to an automated phone system that would answer questions, record test results, or connect them to a nurse. They could also schedule regular one-on-one interactions with a case-worker. Enrolled patients also received test kits and instructions for monitoring their symptoms. The kits were designed for use with the internet or telephone, so patients could easily send test results to LifeMasters. This information was monitored and used in conjunction with claims data to generate “exception reports” that were screened and potentially brought to the attention of physicians. While some physicians requested this data on a regular basis, it was generally used only when there appeared to be a specific opportunity for preventive intervention. Finally, LifeMasters periodically contacted enrolled patients to check on their health and satisfaction with the program.

Our primary source of information for evaluating the impact of the DM program are individual medical claims and lab test results provided by Fallon. These data can be used to measure costs, compliance with clinical practice guidelines, and health outcomes—such as HbA1c levels or inpatient hospitalization rates. We also have a limited set of patient demographics, including age, sex, and a mailing address. We created several additional demographic variables (e.g. race and income) by linking the patients’ zip codes to data from the U.S. Census. Finally, we used social security death records for all patients in the sample to measure mortality. Table 1 provides a short definition for each of the variables used in our analysis, and Table 2 provides means and standard deviations.

The first two columns in Table 2 provide summary statistics for the entire sample of eligible patients. In the remaining columns, we split the sample by baseline-year compliance and co-morbidity. We call a patient compliant if they received an HbA1c test during the baseline year. Co-morbidity indicates the presence of a serious cardiac condition. In particular, co-morbid patients have been diagnosed with either Coronary Artery Disease (CAD), Chronic Obstructive Pulmonary Disease (COPD), or Congestive Heart Failure (CHF). In our sample, 61 percent of the eligible patients were compliant, and 27 percent were co-morbid. In our analysis, we will examine how the impact of enrolling in the DM program varies with these two baseline characteristics.

Table 2 organizes our data into a set of baseline patient characteristics, followed by measures of compliance behavior, health outcomes, and the cost of care. The table indicates that 27 percent of the eligible patients enrolled in the DM program during the study period. It also shows that 29 percent of the eligible patients left the sample during the same three-year period (15 percent died and 14 percent left through attrition). The table provides a number indications that this is an expensive and high-risk population. The average age of an eligible patient was 64, and more than half were Medicare beneficiaries. Based on HbA1c results, only 29 percent of the patients had good control of their diabetes ($\text{HbA1c} \leq 7.0$) while 44 percent had poor

control over the condition ($\text{HbA1c} \geq 8.0$). (The average baseline-year HbA1c was 8.44.) The quarterly inpatient admission rate was 10 percent and the quarterly mortality rate was almost 1 percent. Finally, the average cost of care was almost \$2,700 per patient-quarter, rising to between \$4,400 and \$4,600 for those with co-morbidities.

Table 2 shows the dramatic impact that a chronic condition can have on health and the cost of medical care. Given these statistics, we might expect this group to benefit from preventive care. However, before estimating the impact of the DM program, we should consider patients' enrollment decision and examine the characteristics of enrolled and unenrolled patients. We turn to these issues in the next section as part of our discussion of methods.

4 Methods

Our goal is to identify the average impact of a DM program on enrolled patients (i.e. the average impact of the treatment on the treated). Specifically, we are interested in comparing the health of a patient in the DM program to the health of the same patient at the same point in time had they not enrolled. Since this counterfactual is never observed, we must estimate it. In principle, we would like to do this by randomly assigning patients to the program and comparing the average outcome among the enrolled to the average outcome among the unenrolled. In the absence of a controlled randomized-trial, we are forced to turn to non-experimental methods that mimic this setup under reasonable conditions.

A major concern with our approach is that patients who choose to enroll may differ from those who decline, and that these differences may be correlated with observed outcomes. For example, individuals who are more likely to “take care of themselves” may also be more likely to sign up for a DM program. In this case, an observed correlation between enrollment and health could be caused by the impact of better self-care. More generally, we can classify the various individual characteristics that might confound identification along two dimensions—observability and variation over time. In this section, we discuss our empirical specification and describe several methods that control for fixed or time-varying, observed or unobserved patient characteristics that might bias our results.

A common method of controlling for time-invariant heterogeneity (both observed and unobserved) is to use panel data and estimate difference in differences models.⁶ This approach compares the change in outcomes within the treatment group before and after the intervention to the same change in outcomes within the control group. By using within-group changes, we eliminate any time-invariant patient characteristics that might be correlated with both enroll-

⁶The difference in difference estimator is one of the most widely used in the evaluation literature (see, among others, Angrist 1995; and Heckman et al 2000).

ment and health. The average change in outcomes within the control group serve as an estimate of the true counterfactual—i.e. what would have happened to a treated patient if there was no intervention.⁷ The basic difference-in-differences model can be specified as a two-way fixed effect linear regression:

$$y_{it} = \alpha Treatment_{it} + \beta X_{it} + \gamma_i + \lambda_t + \varepsilon_{it} \quad (1)$$

where y_{it} is an outcome for individual i in quarter t , $Treatment_{it}$ is an indicator variable that takes the value of 1 if individual i is enrolled in the DM program during quarter t and 0 otherwise, X_{it} is a vector of control variables that vary across both individuals and time, γ_i is a fixed-effect unique to individual i , λ_t is a time effect common to all individuals in period t , and ε_{it} is an individual time-varying error distributed independently across patients and independently of all γ_i and λ_t (see Chamberlain, 1984; and Heckman and Robb 1985). The parameter α is an estimate of the average impact of the disease management program.

Although our fixed-effects specification controls for the impact of time-invariant individual characteristics, we should be concerned about the validity of our underlying assumptions if we find that the control and treatment groups are considerably different from one another. In particular, we have assumed that the change in outcomes for patients in the control group is an unbiased estimate of the true counterfactual. However, a difference-in-differences regression will produce biased estimates when the treatment effect varies with individual characteristics that differ across these two groups (Heckman et al 1997, and Heckman et al 1998a). This bias is caused by the presence of enrolled patients in the treatment group with no comparable counterpart in the control group (and vice versa), or by a different distribution of the relevant characteristics within the two groups.⁸

We control for this potential source of bias by constructing a matched sample of patients (both enrolled and unenrolled) that fall within the region of common support in the distribution of X . Rosenbaum and Rubin (1983) show that matching treated and untreated observations on the basis of X is equivalent to matching them using a balancing score $B(x)$.⁹ Conditional on $B(X)$, the counterfactual outcome distribution for the treated patients is the same as the observed outcome distribution for the controls. The coarsest balancing score is the propensity score—the probability of enrollment conditional on the pre-treatment values of the vector X ,

⁷Another way to state this is that the change in outcomes for enrolled patients controls for fixed characteristics, while the change in outcomes for unenrolled patients controls for time-varying factors that are common to both groups.

⁸Heckman et al (1997) suggests that, in practice, the first of these two sources of bias is likely to be the most severe.

⁹This result is very important in practice since it reduces the potential problem of matching on a high dimensional vector X to matching on a scalar.

(i.e. $B(X) = Pr(Treated = 1|X)$).

We estimate propensity scores from a logit model of the enrollment decision. The results of these enrollment regressions are presented in Table 3. We find that the effect of age on enrollment is concentrated among the older patients with co-morbidities. Women are more likely to enroll than men. Patients whose primary care physician is part of the Fallon Clinic are more likely to join than patients associated with the independent network. Patients who had their HbA1c level checked during the baseline year, or who spent more on medications prior to the intervention signed up more frequently. Finally, patients with high HbA1c test results at baseline were more likely to enroll.

Using the estimated enrollment probabilities from these regressions (i.e the fitted values) we found observations on the common support of X by excluding all unenrolled patients whose estimated probability of enrollment fell below the first percentile of the probability distribution for the enrolled population. (Similarly, we exclude all treatment observations whose propensity score was greater than the 99th percentile of the unenrolled population.) This matching procedure trimmed 285 patients, or 4.6 percent of the overall population, from our sample. Table 4 compares the baseline-year means of the enrolled and unenrolled patients in the matched sample for several variables.

By constructing a matched sample, we have attempted to rule out the possibility that our estimate of the true counterfactual (based on changes within the control group) was biased by different distributions of the observed patient characteristics. However, we might also be concerned that unobserved characteristics will lead to a similar bias. While we cannot examine these characteristics directly, we can use the fact that we have several periods of pre-enrollment data to search for evidence that they exist. In particular, we can use the pre-intervention data to test whether the control and treatment groups had similar secular trends prior to enrollment (Heckman and Hotz 1989). If the secular trends were indistinguishable prior to the intervention, it seems reasonable to assume that they would have been indistinguishable in the post-intervention period, so that the unenrolled controls provide an unbiased estimate of the counterfactual.

To test for equality of the pre-intervention secular trends, we use a simple variation on the difference-in-differences regression of Equation 1. Specifically, we estimate the following model:

$$y_{it} = \gamma_i + \lambda_t + Treated_i \delta_t + \varepsilon_{it} \tag{2}$$

where $Treated_i$ is an indicator variable that takes a value of 1 if individual i ever enrolls in the program, γ_i are a set of time-invariant individual fixed-effects, λ_t are a set of period-specific effects common to all individuals, and ε_{it} is an independently distributed individual time-

varying error term. (We restrict attention to the pre-treatment period by removing any post-enrollment observations from the treatment group.) The coefficients of interest are the δ_t , which measure period-specific differences in outcomes between the treatment and control populations prior to enrollment. If we cannot reject the null hypothesis that the δ_t are jointly equal to zero, then the pre-treatment trends in the control group are statistically indistinguishable from those in the treatment population.

We estimated Equation 2 on both the matched and unmatched samples using four different dependent variables (HbA1c Compliant, Inpatient visit, HbA1c, and log Total Costs). All of the data comes from the four pre-intervention quarters prior to the start of LifeMasters' enrollment efforts. Table 5 reports F-statistics for a test of the joint null hypothesis $\delta_t = 0$, based on each of these regressions.¹⁰ With the potential exception of HbA1c scores, the results on the unmatched sample generally indicate that the pre-treatment trends were the same for enrolled and unenrolled patients. For the matched sample, we cannot reject the null hypothesis of equal pre-treatment trends for any one the dependent variables. These results suggest that correlations between enrollment and unobserved patient heterogeneity will not bias our results. For the remainder of the paper, we will report results based on the matched sample.

Finally, it is possible that changes over time in the composition of the sample might bias the results. Correlations between enrollment and either death or attrition could have a large effect, since 27 percent of the eligible patients leave our sample before the end of our study period. A common method for dealing with this issue is to construct a balanced panel, excluding all patients who either die or leave during the study period. Using a balanced panel, we found no significant changes to any of the results presented below.¹¹ Another way to check for a potential problem is to examine the correlation between enrollment and either death or attrition. Table 7 presents the results of a discrete-time hazard model of attrition (excluding all patients who died during the study). While these results show a slight decline in attrition during the first six months of treatment followed by a mild increase, neither of the effects is statistically significant. If we interact enrollment with co-morbidity, we find that joining the DM program increases retention for co-morbid patients, offsetting a higher average attrition rate among this group. If anything, this retention effect for the sicker and more expensive co-morbid patients will bias our results towards no effect or a negative impact. (We discuss the results of a similar regression for mortality, presented in Table 13, in the Results section.) However, based on analysis of the balanced panel, we do not believe that death or attrition lead to any bias in our results.

¹⁰We also examined whether the results varied with severity of illness by estimating the co-morbid and non-comorbid populations separately.

¹¹All of these results may be obtained from the authors on request.

5 Results

In this section we estimate the average impact of the DM program on compliance with clinical practice guidelines, health outcomes, and the cost of care. First, however, Table 6 provides a count of enrolled patients for each quarter in the analysis, showing how the size of the program changed over time. Initial recruitment efforts led to a quick surge in participation during the third and fourth quarters of 1999. Enrollment then levelled off until the second quarter of 2001, when it began to grow gradually through the end of the study period. The second half of Table 6 shows the number of patients that we observe in a given quarter of enrollment. While we observe almost 2,000 patients during their first year of the program, these sample sizes fall to almost half that number by the third year.¹²

Our main results are found in Tables 8 through 11.¹³ The first three tables show how the impact of the DM program evolves over time for various measures of compliance, health, and cost. The fourth table focuses on the long-run (seven months and beyond) effects of enrollment, and how these vary with baseline compliance and co-morbidity.

5.1 Compliance Behavior

Clinical practice guidelines suggest that high-risk diabetes patients should have an HbA1c exam every six months, and a retinal exam, lipids panel, and proteinuria exam annually. Based on these guidelines, we created four indicators of therapy compliance. Each variable was coded as one for “in compliance” quarters and 0 for “out-of-compliance” quarters (i.e. any quarter when the elapsed time since a patient’s last exam exceeded the recommended guideline). To assess the impact of the DM program on compliance behavior, we estimated a series of linear probability models using the difference-in-difference specification of Equation 1. Table 8 provides estimates of the DM program’s impact for each quarter of the first year of enrollment, each semester of the second year, and thereafter.

Our results show that there was a significant improvement in patients’ compliance following enrollment. The average effects were around 4.6 percent for HbA1c exams, 3.3 percent for retinal exams, and 3.0 percent for proteinuria tests. (In terms of the percentage change in baseline compliance rates, these numbers represent an 8.2 percent change for HbA1c, 9.0 percent for retinal exams and, and 13.6 percent for proteinuria tests.) There also appears to be a small but positive enrollment effect for lipids. While the impact appears to be long-lived for retinal and proteinuria tests, they disappear after 9 months for HbA1c tests. However, the short-

¹²Since we find that most of the estimated effects stabilize after six months to one year in the program, we pool observations from these later periods in our analysis.

¹³All estimation was done using Stata version 8.

run nature of the HbA1c result may reflect the fact that Fallon started its own initiative to encourage HbA1c compliance, which probably reached many of the unenrolled patients and their physicians.

5.2 Health Outcomes

Table 9 shows the impact of the DM program on several health outcomes. The first panel in this table uses HbA1c test results as a measure of patients’ control over their diabetes. The second two measures are indicator variables for an HbA1c score greater than 7 and 8 respectively (where less than 7 represents “good” control over diabetes and greater than 8 represents “poor” control). Finally, we created an indicator variable coded as one for quarters when a patient had an inpatient hospital admission to measure the occurrence of severely adverse outcomes. Again, all of the results are based on a differences-in-differences specification with individual and quarterly fixed effects.

In the first panel of Table 9, we see a statistically significant drop of 0.3 to 0.4 points for HbA1c test results beginning in the second quarter of enrollment. On average, this represents a 4 to 5 percent improvement that remains quite stable over time. From the second and third panels, we can see that the average decline in HbA1c results led to a substantial decline in the number of patients with “poor” control over their condition. The results show that enrolled patients were 10 to 12 percent more likely than unenrolled patients to bring their HbA1c scores below the threshold value of 8.0. The fourth panel in Table 9 shows that there is a long-term decline in inpatient admissions for patients in the DM program relative to the controls. This 1 percent drop in the quarterly admission rate corresponds to a 10 percent decline in the probability of admission, which can lead to dramatic cost savings.

Our final measure of the program’s health impact is its effect on mortality. However, because death is not an event for which we have repeated observations, we could not use the difference-in-differences specification. Instead, we estimated a discrete-time hazard model, assuming that the patient-specific hazard rate (i.e. probability of death in a given period) takes the logistic form $h_{it} = 1/\{1 + \exp(-\lambda_t - X_{it}\beta)\}$ (e.g. Allison 1982; Jenkins 1995). In this expression, the λ_t are a set of period-specific effects common to all individuals, and the X_{it} are a set of patient-quarter covariates that may or may not change over time.¹⁴ We are interested in the coefficient β on a time-varying indicator variable corresponding to enrollment in the disease management program. Table 13 presents the parameter estimates from this model of mortality. We included sex and a set of baseline medical conditions as time-invariant explanatory variables along with

¹⁴Given these assumptions, the hazard model has the same likelihood function as a logistic regression where the dependent variable is an indicator for death in the current quarter.

enrollment and age as time-varying explanatory variables. We found that enrolling in the DM program lead to a 21 percent drop in average quarterly mortality. This effect is concentrated among baseline non-compliant patients with co-morbidities, for whom the average mortality rate falls by almost 50 percent.

5.3 Cost of Care

Table 10 illustrates the impact of the DM program on health care expenditures. We divide these expenditures into four categories—total costs, inpatient claims, outpatient claims, and pharmaceutical claims.¹⁵ Each of the four panels presents results from a difference-in-differences specification with individual and quarterly fixed effects.

Our primary result appears in the first panel of Table 10. We find that patients’ total cost of care declines by 10 to 15 percent after six months in the DM program. These savings are the result of changes in both inpatient and outpatient claims. There is a large short-term drop in inpatient claims that becomes smaller in the second and third year of enrollment (although it remains significant in economic terms). on the other hand, outpatient savings relative to the control population grow over time to between 6 and 13 percent of baseline costs.

The decline in medical claims is partly offset, at least in the short-term, by an increase in pharmaceutical expenditures. Pharmaceutical spending is a measure of both therapy compliance and overall health, and our results are consistent with a short run cost increase (associated with improved therapy compliance), followed by a long-run decline in spending (the result of improved health). However, we do not wish to place too strong an interpretation on these estimates. The most important result is the appearance of significant aggregate savings for enrolled patients after six months in the program. This finding helps to explain why insurers have been adopting DM programs despite their often high levels of patient turnover.

5.4 Long-run Impacts

Table 11 examines the long-run impact of enrollment and how that impact varies with patients’ baseline compliance and co-morbidity status. The table is broken into three panels for compliance, health, and financial impacts. Within each panel, the first set of coefficients shows the average impact of the DM program starting in the 7th month of enrollment. These coefficients are based on a difference-in-differences model with individual and quarterly fixed effects that excludes observations from the first six months of program participation. The results uniformly

¹⁵For each regressions, we used log expenditures as the dependent variable for ease of interpretation and to reduce the importance of outliers. Running the regressions in levels produced qualitatively similar results.

show that the DM program led to statistically significant long-run improvements in health and compliance, along with reductions in the total cost of care.

We also investigated the impact of baseline compliance and co-morbidity on our results by interacting these variables with program enrollment (and including a set of quarterly fixed-effects interacted with compliance and co-morbidity). These regressions produced a number of interesting results that are reported in the second half of each panels in Table 11. First, we see that impact of enrollment on compliance behaviors is generally larger for patients with baseline co-morbidities. There is also some evidence that baseline compliance predicts a compliance effect following enrollment in the DM program. In three out of four cases, it is the non-compliant non-comorbid patients who have marginal and statistically insignificant compliance effects. This suggests that it is harder to change the behavior of someone who is neither sick nor predisposed to comply with the recommended therapy.

When we turn to health effects, there is a slightly different story. It appears that the DM program had a fairly uniform impact on HbA1c levels (although there were larger shifts in the tail of the distribution for patients who were baseline compliant). But the decline in inpatient admissions was concentrated among patients who were non-compliant at baseline. For these patients, the admission rate declined by almost 50 percent following enrollment. Mortality rates also declined more for patients who were non-compliant at baseline. With mortality, however, the only group whose odds of survival improved by a statistically significant margin were those with co-morbidities. Taken together, the long-term impacts on health suggest that while all of the DM participants realized some benefits, the biggest beneficiaries were the non-compliant—who naturally stand to benefit the most from a preventive intervention—and the co-morbid, who are at much greater risk.

The third panel shows that the financial impacts of the DM program were overwhelmingly concentrated among the patients who were non-compliant at baseline. For non-comorbid non-compliant patients, there were reductions in inpatient, outpatient and pharmaceutical claims. For non-compliant co-morbid patients, we saw decrease in inpatient and outpatient claims accompanied by an increase in pharmaceutical spending. Finally, for the co-morbid compliant patients, we actually find an increase in expenditures. These cost patterns are consistent with the notion that preventive care works best for those who are not already taking care of themselves. Moreover, if these individuals can be reached before they become sick, the effects are even larger. For co-morbid patients, on the other hand, disease management appears to increase costs (particularly pharmaceutical) at the same time as it prolongs life.

Finally, to get some feel for the substantial economic benefits of DM, we can apply the estimated long-run quarterly savings of 15 percent to the baseline total quarterly costs of \$2,694. This results in an average savings of \$404 per enrolled patient per quarter. During 2000, when

Fallon had enrolled approximately 1,000 patients, our estimates suggest that LifeMasters' DM program saved them about \$1.6 million.

6 Conclusions

We find that the DM program for diabetes introduced by Fallon had a significant impact on compliance behaviors, health outcomes (including HbA1c levels, inpatient hospitalization, and mortality rates), and the overall cost of care. Specifically, we find increases of 10 to 20 percent in patients' compliance with a variety of clinical practice guidelines—particularly among the comorbid. We also find a reduction of 5 percent in average HbA1c levels, and a decline in inpatient admissions and mortality rates. Finally, we found statistically and economically significant reductions in the overall cost of care for enrolled patients. On average, there was a 15 percent cost reduction that appeared after about six months in the DM program.

The primary implication of these results is that disease management appears to be a promising method for providing effective and financially viable preventive care. The growing popularity of DM reflects its promise as a way out of the dilemma posed by rapidly increasing health care costs and an aging population. The rise of DM may also herald broader changes in the industrial organization of health care. We have argued that DM is being developed by insurers and managed care providers because of financial incentives and scale. Innovations such as DM, electronic medical records, and clinical practice guidelines are beginning to be applied in ways that take responsibility for individual patients out of the physicians' hands and give it to a variety of new kinds of health-care workers, as well as to the patients themselves. Costs and demographics will continue to create pressure for productivity enhancing changes in the delivery of health care.

References

- Allison, P.D. (1982) "Discrete-time methods for the analysis of event-histories," *Sociological Methodology*, pp.61-97.
- Angrist, J. (1985) "Introduction to the JBES Symposium on Program and Policy Evaluation," *Journal of Business and Economic Statistics* 13, pp. 249-88.
- Bindels R, de Clercq PA, Winkens RA, Hasman A., (2000) "A test ordering system with automated reminders for primary care based on practice guidelines," *Int J Med Inf*;58-59:219-33.
- Cabana, MD, Reed, CS, et al (1999) "Why Don't Physicians Follow Clinical Practice Guidelines?" *JAMA*, 282(15): 1458-1465.
- Chamberlain, G. (1984) "Panel Data," in Griliches, Z. and M. Intrilligator (eds.) *Handbook of Econometrics*, Volume 2. North-Holland.
- Chassin, MR; Kosecoff, J; Park, et al (1987) "Does inappropriate use explain geographic variations in the use of health care services? A study of three procedures" *JAMA*, 258(18):2533-7.
- Christakis DA, Zimmerman FJ, Wright JA, et al, (2001) "A Randomized Controlled Trial of Point-of-Care Evidence to Improve the Antibiotic Prescribing Practices for Otitis Media in Children," *Pediatrics* 107(2): e15-e19.
- Grimshaw JM, Russell IT (1993) "Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations." *Lancet*; 342: 1317-1322.
- Heckman, J., R. LaLonde and J. Smith (2000): "The Economics and Econometrics of Active Labor Market Programs," in *The Handbook of Labor Economics*, Vol. 3A. Ashenfelter, O. and D. Card (eds.), North-Holland.
- Hermens RP, Hak E, Hulscher ME, Mulder J, Tacke MA, Braspenning JC, Grol RP. (1999) "Improving population-based cervical cancer screening in general practice: effects of a national strategy." *Int J Qual Health Care*; 11(3):193-200.
- Hoffman, C. and D. Rice (1995): "Estimates based on the National Medical Expenditure Survey", UCSF Institute for Health and Aging.
- Iliadis EA, Klein LW, Vandenberg B et al (1999) "Clinical practice guidelines in unstable angina improve clinical outcomes by assuring early intensive medical treatment" *Am Coll Cardiol* 15;34(6):1689-95.
- IOM (2000) "To Err is Human: Building a Safer Health System," Washington DC: National Academy Press.
- IOM (2001) "Crossing the Quality Chasm: A New Health System for the 21st Century," Washington DC: National Academy Press.
- Jenkins, S. (1995) "Easy Estimation Methods for Discrete-Time Duration Models," *Oxford Bulletin of Economics and Statistics*, 57(1): pp. 129-137.
- Leviton LC, Goldenberg RL, Baker CS, Schwartz RM, Freda MC, Fish LJ, Cliver SP, Rouse DJ, Chazotte C, Merkatz IR, Raczynski JM. (1999) "Methods to encourage the use of antenatal corticosteroid therapy for fetal maturation: a randomized controlled trial." *JAMA*; 281(1):46-52.

McPhee, SJ; Bird, JA; Fordham, D; Rodnick, JE; Osborn, EH. (1991) "Promoting cancer prevention activities by primary care physicians. Results of a randomized, controlled trial." *JAMA*, 266(4):538-44.

Ray WA, Stein CM, Byrd V, Shorr R, Pichert JW, Gideon P, Arnold K, Brandt KD, Pincus T, Griffin MR, (2001) "Educational Program for Physicians to Reduce Use of Non-Steroidal Anti-Inflammatory Drugs Among Community-Dwelling Elderly Persons: A Randomized Controlled Trial," *Medical Care*; 39(5): 425-435.

Rosenbaum, P. and D. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika* 70, pp. 41-55.

Schriger DL (1996) "Emergency medicine guidelines: we can make them, but will we use them?" *Annals of Emergency Medicine*. 27: 655-657.

Schriger DL, Baraff LJ, Rogers WH, and Cretin S. (1997) "Implementation of Clinical Guidelines Using a Computer Charting System: Effect on the Initial Care of Health Care Workers Exposed to Body Fluids," *JAMA*; 287(19): 1585-1590.

Schuster MA, McGlynn EA, and Brook RH (1998): "How Good is the Quality of Health Care in the United States" *The Milbank Quarterly*, 76(4): 517-563.

Shiffman RN, Liaw Y, Brandt et al (1999), "Computer-Based Guideline Implementation Systems: A Systematic Review of Functionality and Effectiveness," *Journal of American Medical Informatics Association*; 6(2):104-114.

Stone TT, Kivlahan CH, Cox KR. (1999) "Evaluation of physician preferences for guideline implementation" *Am J Med Qual*; 14(4):170-7.

Tang PC, LaRosa MP, Newcomb C, Gorden SM. (1999) "Measuring the effects of reminders for outpatient influenza immunizations at the point of clinical opportunity." *J Am Med Inform Assoc*; 6(2):115-21.

Wennberg, J. and A. Gittlesohn (1973): "Small Area Variations in Health Care Delivery" *Science*; 182: 1102-8.

Table 1: Variable Definitions

	Baseline Characteristics	
Age	Fixed	Age in years at September 1999
Medicare	Fixed	Indicator variable: 1 if patient insured by Medicare
Medicaid	Fixed	Indicator variable: 1 if patient insured by Medicaid
Core Treated	Fixed	Patient's primary care physician was a member of the Fallon Clinic
Died	Fixed	Patient enrolled in the DM program
Attrition	Fixed	Patient died before the end of the study
Baseline HbA1c	Fixed	Patient left Fallon before the end of the study
Compliant	Fixed	Pre-intervention HbA1c Test results
Comorbid	Fixed	Indicator variable: 1 if patient received a pre-intervention HbA1c exam
CHF	Fixed	Indicator variable: 1 if patient had CAD, COPD, or CHF prior to intervention
COPD	Fixed	Indicator variable: patient had Congestive Heart Failure
CAD	Fixed	Indicator variable: patient had Chronic Obstructive Pulmonary Disease
Hypertension	Fixed	Indicator variable: patient had Coronary Artery Disease
		Indicator variable: patient had Hypertension
	Compliance Measures	
HbA1c Compliant	Quarterly	Indicator variable: 1 if received HbA1c exam in last six months
Eye Compliant	Quarterly	Indicator variable: 1 if received retinal exam in last year
Lipid Compliant	Quarterly	Indicator variable: 1 if received Lipids panel in last year
Malb Compliant	Quarterly	Indicator variable: 1 if received Proteinuria test in last year
	Health Outcomes	
HbA1c	Quarterly	HbA1c test results
HbA1c7	Quarterly	Indicator variable: 1 if HbA1c results are above 7.0
HbA1c8	Quarterly	Indicator variable: 1 if HbA1c results are above 8.0
Inpatient visit	Quarterly	Indicator variable: 1 if had an inpatient admission
Death	Quarterly	Indicator variable: 1 if died in current quarter
	Financial Outcomes	
Total Cost	Quarterly	Dollar value of claims (Inpatient, outpatient or pharmaceutical)
Inpatient Claims	Quarterly	Dollar value of inpatient claims
Outpatient Claims	Quarterly	Dollar value of outpatient claims
Pharma Claims	Quarterly	Dollar value of pharmaceutical claims

Table 2: Summary Statistics

Variable	Eligible Population		Non-comorbid Non-compliant		Non-comorbid Compliant		Comorbid Non-compliant		Comorbid Compliant	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
	Baseline Characteristics									
Treated	0.27	0.44	0.21	0.41	0.30	0.46	0.19	0.39	0.33	0.47
Died	0.15	0.36	0.09	0.29	0.10	0.30	0.29	0.45	0.30	0.46
Attrition	0.14	0.35	0.19	0.39	0.11	0.32	0.21	0.41	0.06	0.23
Age	64.64	14.63	60.19	16.76	63.73	14.31	70.07	11.55	70.80	9.58
Male	0.53	0.50	0.51	0.50	0.52	0.50	0.56	0.50	0.57	0.50
Urban (Zip)	0.84	0.23	0.85	0.23	0.84	0.24	0.85	0.20	0.83	0.23
White (Zip)	0.90	0.09	0.89	0.09	0.90	0.09	0.89	0.10	0.90	0.09
Income (Zip)	49,440	13,469	49,462	13,918	49,366	13,109	49,828	14,164	49,293	13,170
Medicare	0.58	0.49	0.44	0.50	0.55	0.50	0.70	0.46	0.80	0.40
Medicaid	0.03	0.17	0.05	0.22	0.02	0.12	0.05	0.22	0.01	0.11
Core	0.89	0.31	0.77	0.42	0.99	0.11	0.68	0.46	0.99	0.09
Baseline HbA1c	8.44	1.88	8.50	1.61	8.57	2.11	8.17	1.04	8.14	2.03
Compliant	0.61	0.49	0	0	1	0	0	0	1	0
Comorbid	0.27	0.45	0	0	0	0	1	0	1	0
CHF	0.11	0.31	0	0	0	0	0.40	0.49	0.41	0.49
COPD	0.09	0.28	0	0	0	0	0.37	0.48	0.27	0.44
CAD	0.18	0.38	0	0	0	0	0.62	0.49	0.68	0.47
Hypertension	0.31	0.46	0.25	0.43	0.31	0.46	0.39	0.49	0.37	0.48
	Compliance Measures									
HbA1c Compliant	0.56	0.50	0.23	0.42	0.75	0.43	0.15	0.36	0.73	0.44
Eye Compliant	0.42	0.49	0.34	0.47	0.44	0.50	0.41	0.49	0.50	0.50
Lipid Compliant	0.49	0.50	0.27	0.45	0.62	0.49	0.23	0.42	0.66	0.48
Malb Compliant	0.22	0.42	0.12	0.32	0.32	0.47	0.06	0.23	0.25	0.43
	Health Outcomes									
HbA1c	8.09	1.72	8.11	2.00	8.15	1.68	7.60	1.69	7.92	1.67
HbA1c ≥ 7	0.71	0.45	0.66	0.47	0.74	0.44	0.56	0.50	0.67	0.47
HbA1c ≥ 8	0.44	0.50	0.44	0.50	0.46	0.50	0.32	0.47	0.39	0.49
Inpatient visit	0.09	0.29	0.07	0.25	0.06	0.24	0.19	0.40	0.17	0.37
Death	0.01	0.10	0.01	0.08	0.01	0.08	0.02	0.13	0.02	0.14
	Financial Outcomes									
Total Cost	2,694	7,590	2,123	7,049	2,020	5,539	4,666	10,833	4,400	10,222
Inpatient Claims	946	5,766	712	5,329	552	3,939	2,137	8,838	1,708	7,857
Outpatient Claims	1,307	2,792	1,047	2,566	1,052	2,391	1,988	3,259	2,091	3,665
Pharma Claims	441	460	363	452	415	396	540	555	600	528
Patients	6,142		1,601		2,866		793		882	
Patient-Quarters	97,306		26,452		45,748		11,720		13,386	

Table 3: Logit Estimates of Enrollment in DM Program

	Entire Sample	No Death or Attrition
Age	-0.004 (0.003)	-0.008 (0.003)**
Male	-0.203 (0.060)***	-0.232 (0.064)***
Urban	-0.103 (0.145)	-0.134 (0.152)
White	0.493 (0.503)	0.658 (0.536)
Core	0.615 (0.132)***	0.254 (0.151)*
log Income	0.024 (0.164)	-0.074 (0.173)
Medicare	-0.013 (0.095)	0.018 (0.101)
log Total Costs	0.032 (0.025)	0.037 (0.028)
log Pharma	0.117 (0.020)***	0.112 (0.022)***
Inpatient Visit	-0.123 (0.097)	-0.138 (0.103)
Baseline Compliant	0.282 (0.070)***	0.247 (0.073)***
Baseline HbA1c	0.104 (0.015)***	0.098 (0.016)***
Cholesterol Test	0.268 (0.081)***	0.264 (0.085)***
Baseline Comorbid	0.082 (0.157)	0.084 (0.163)
CAD	0.113 (0.143)	0.115 (0.148)
CHF	-0.064 (0.123)	-0.073 (0.128)
COPD	-0.308 (0.143)**	-0.335 (0.149)**
Hypertension	-0.091 (0.066)	-0.106 (0.070)
Observations	6,112	5,265

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Attrition: Beneficiaries who leave Fallon before the end of the study period.

Table 4: Comparison of Means for Matched Sample

Variable	Control	Treatment	T-test
Age	65.22	64.13	2.56
Male	0.54	0.50	3.01
Medicare	0.59	0.56	2.25
Medicaid	0.02	0.02	0.09
Urban	0.84	0.83	1.46
White	0.90	0.90	1.70
Income	49,397	49,539	0.36
Core	0.90	0.95	6.26
Compliant	0.59	0.71	8.20
Comorbid	0.27	0.27	0.38
CHF	0.11	0.11	0.67
COPD	0.09	0.07	2.55
CAD	0.17	0.19	1.04
Hypertension	0.32	0.31	1.11
HbA1c Exam	0.59	0.71	8.20
Eye Exam	0.63	0.70	5.30
Lipid Panel	0.36	0.46	7.12
Proteinuria	0.07	0.09	3.20
Base HbA1c	8.27	8.60	6.27
HbA1c7	0.77	0.86	6.24
HbA1c8	0.54	0.71	9.79
Inpatient Visit	0.24	0.23	0.69
Total Cost	8,605	8,055	0.95
Inpatient Claims	3,579	3,064	1.24
Outpatient Claims	5,025	4,990	0.15
Pharma Claims	1,749	2,123	6.75
Patients	4,280	1,577	

Comparison based on two sample t-test with equal variances; * significant at 10%; **significant at 5%; ***significant at 1%.

Table 5: F-tests for Equality of Control and Treatment Pre-intervention Trends

	Entire Sample				
	Patients	Compliant	Inpatient	HbA1c	log Costs
Pooled	6,142	0.64	1.47	4.84***	1.50
Non-Comorbid	4,467	0.64	0.21	3.27**	0.63
Comorbid	1,675	0.13	1.47	1.89	1.24
	Matched Sample				
	Patients	Compliant	Inpatient	HbA1c	log Costs
Pooled	5,888	0.90	0.92	0.30	1.30
Non-comorbid	4,280	0.77	0.19	1.02	0.53
Comorbid	1,586	0.03	1.00	0.35	0.84

Entries are F-statistics for the null hypothesis that quarterly pre-treatment dummies are jointly equal to zero.

* significant at 10%; ** significant at 5%; *** significant at 1%..

Table 6: Sample Sizes for Control and Treatment

	Enrollment by Calendar Quarter											
Quarter	Q399	Q499	Q100	Q200	Q300	Q400	Q101	Q201	Q301	Q401	Q102	Q202
Control	5788	5414	5207	5026	4845	4741	4566	4216	3946	3779	3395	3216
Treatment	717	962	1015	1018	1029	1025	1013	1209	1292	1314	1433	1463
	Enrollment by Treatment Quarter											
Quarter	1	2	3	4	5	6	7	8	9	10	11	12
Observations	2407	2287	2034	1924	1750	1434	1336	1285	1192	1134	1052	755

Table 7: Logit Estimates of Attrition Probability

	Entire Sample	Matched Sample	Matched Sample
Months1-6	-0.161 (0.194)	-0.261 (0.206)	-0.260 (0.206)
Months7+	0.110 (0.111)	0.073 (0.113)	0.141 (0.229)
Baseline Comorbid * Months7+			-0.449 (0.234)*
Baseline Compliant * Months7+			0.096 (0.245)
Baseline Comorbid	0.316 (0.188)*	0.379 (0.202)*	0.459 (0.207)**
Baseline Compliant	-0.129 (0.088)	-0.054 (0.091)	-0.070 (0.100)
Male	-0.000 (0.074)	-0.048 (0.079)	-0.052 (0.079)
Age	-0.003 (0.001)***	-0.004 (0.001)***	-0.004 (0.001)***
Core	-1.751 (0.093)***	-1.787 (0.103)***	-1.783 (0.104)***
Medicare	-0.598 (0.107)***	-0.635 (0.114)***	-0.636 (0.115)***
Medicaid	1.133 (0.156)***	1.016 (0.197)***	1.009 (0.197)***
Baseline HbA1c	0.081 (0.019)***	0.098 (0.020)***	0.096 (0.020)***
log Total Costs	0.118 (0.023)***	0.090 (0.028)***	0.089 (0.028)***
log Pharma	0.007 (0.018)	0.044 (0.026)*	0.044 (0.025)*
Hypertension	-0.002 (0.082)	0.019 (0.088)	0.011 (0.088)
CAD	-0.139 (0.169)	-0.126 (0.183)	-0.137 (0.183)
CHF	0.017 (0.144)	-0.008 (0.158)	0.005 (0.159)
COPD	-0.051 (0.173)	-0.093 (0.192)	-0.111 (0.192)
Observations	65387	62785	62785

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%.

Attrition: Beneficiaries who leave Fallon before the end of the study period.

Table 8: Difference-in-differences Estimates of DM on Compliance Behavior

	Qtr1	Qtr2	Qtr3	Qtr4	Qtr5-6	Qtr7-8	Qtr9+
HbA1c Exam							
Unmatched	0.047 (0.010)***	0.069 (0.010)***	0.050 (0.011)***	0.016 (0.011)	-0.002 (0.010)	0.017 (0.011)	0.008 (0.010)
Matched	0.046 (0.011)***	0.068 (0.011)***	0.043 (0.012)***	0.007 (0.012)	-0.000 (0.011)	0.015 (0.012)	0.007 (0.010)
Retinal Exam							
Unmatched	0.038 (0.009)***	0.031 (0.009)***	0.043 (0.010)***	0.054 (0.010)***	0.036 (0.009)***	0.043 (0.010)***	0.036 (0.009)***
Matched	0.033 (0.010)***	0.027 (0.010)***	0.043 (0.011)***	0.053 (0.011)***	0.041 (0.010)***	0.047 (0.011)***	0.036 (0.010)***
Lipids Panel							
Unmatched	0.003 (0.009)	0.014 (0.009)	0.024 (0.010)**	0.032 (0.011)***	0.004 (0.009)	0.022 (0.010)**	0.000 (0.010)
Matched	-0.006 (0.010)	0.004 (0.011)	0.015 (0.011)	0.024 (0.012)**	0.006 (0.010)	0.016 (0.011)	-0.003 (0.011)
Proteinuria Exam							
Unmatched	0.031 (0.009)***	0.037 (0.009)***	0.038 (0.010)***	0.037 (0.010)***	0.020 (0.009)**	0.056 (0.010)***	0.098 (0.010)***
Matched	0.030 (0.010)***	0.035 (0.010)***	0.040 (0.011)***	0.032 (0.011)***	0.024 (0.010)**	0.053 (0.011)***	0.095 (0.011)***

Each cell contains the average treatment effect for patients in a given enrollment period.

Individual and quarterly fixed effects were included in each model.

Excluding death and attrition produces negligible differences (results available on request).

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 9: Difference-in-differences Estimates of DM on Health Outcomes

	Qtr1	Qtr2	Qtr3	Qtr4	Qtr5-6	Qtr7-8	Qtr9+
HbA1c (Mean = 8.09)							
Unmatched	-0.083 (0.045)*	-0.362 (0.047)***	-0.451 (0.052)***	-0.229 (0.059)***	-0.381 (0.045)***	-0.328 (0.049)***	-0.395 (0.046)***
Matched	-0.073 (0.048)	-0.318 (0.050)***	-0.446 (0.055)***	-0.203 (0.063)***	-0.346 (0.048)***	-0.326 (0.050)***	-0.360 (0.046)***
HbA1c ≥ 7							
Unmatched	-0.026 (0.014)*	-0.045 (0.015)***	-0.060 (0.016)***	-0.015 (0.019)	-0.053 (0.014)***	-0.016 (0.015)	-0.030 (0.014)**
Matched	-0.024 (0.016)	-0.030 (0.016)*	-0.059 (0.018)***	-0.016 (0.020)	-0.054 (0.015)***	-0.022 (0.016)	-0.028 (0.015)*
HbA1c ≥ 8							
Unmatched	-0.043 (0.016)***	-0.122 (0.016)***	-0.150 (0.018)***	-0.086 (0.021)***	-0.145 (0.016)***	-0.125 (0.017)***	-0.154 (0.016)***
Matched	-0.031 (0.017)*	-0.112 (0.018)***	-0.155 (0.019)***	-0.076 (0.022)***	-0.140 (0.017)***	-0.124 (0.018)***	-0.147 (0.017)***
Inpatient Visit							
Unmatched	-0.016 (0.007)**	-0.015 (0.007)**	-0.026 (0.008)***	-0.023 (0.008)***	-0.006 (0.007)	-0.008 (0.008)	-0.005 (0.007)
Matched	-0.016 (0.007)**	-0.013 (0.007)*	-0.015 (0.008)*	-0.030 (0.008)***	-0.014 (0.007)**	-0.013 (0.008)*	-0.008 (0.007)

Each cell contains the average treatment effect for patients in a given enrollment period.

Individual and quarterly fixed effects were included in each model.

Excluding death and attrition produces negligible differences (results available on request).

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 10: Difference-in-differences Estimates of DM on Cost of Care

	Qtr1	Qtr2	Qtr3	Qtr4	Qtr5-6	Qtr7-8	Qtr9+
	log Total Costs						
Unmatched	-0.031 (0.036)	-0.041 (0.037)	-0.134 (0.040)***	-0.140 (0.041)***	-0.144 (0.036)***	-0.092 (0.039)**	-0.089 (0.036)**
Matched	-0.041 (0.040)	-0.066 (0.041)	-0.121 (0.044)***	-0.146 (0.045)***	-0.128 (0.038)***	-0.089 (0.042)**	-0.102 (0.037)***
	log Inpatient						
Unmatched	-0.149 (0.061)**	-0.134 (0.063)**	-0.230 (0.067)***	-0.185 (0.069)***	-0.053 (0.060)	-0.053 (0.066)	-0.042 (0.060)
Matched	-0.138 (0.061)**	-0.109 (0.063)*	-0.142 (0.067)**	-0.252 (0.069)***	-0.120 (0.059)**	-0.086 (0.064)	-0.062 (0.057)
	log Outpatient						
Unmatched	0.005 (0.049)	-0.019 (0.050)	-0.166 (0.054)***	-0.125 (0.056)**	-0.115 (0.048)**	-0.113 (0.053)**	-0.135 (0.048)***
Matched	-0.018 (0.054)	-0.046 (0.056)	-0.155 (0.060)***	-0.131 (0.061)**	-0.059 (0.052)	-0.073 (0.057)	-0.124 (0.051)**
	log Pharmaceutical						
Unmatched	0.162 (0.036)***	0.150 (0.037)***	0.070 (0.040)*	-0.021 (0.041)	-0.051 (0.035)	0.046 (0.039)	0.110 (0.036)***
Matched	0.091 (0.038)**	0.073 (0.040)*	0.005 (0.042)	-0.070 (0.043)	-0.094 (0.037)**	-0.038 (0.040)	0.019 (0.036)

Each cell contains the average treatment effect for patients in a given enrollment period.

Individual and quarterly fixed effects were included in each model.

Excluding death and attrition produces negligible differences (results available on request).

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 11: Difference-in-differences Estimates of Long-run Impacts of DM

	Compliance Effects			
	HbA1c Test	Retinal Exam	Lipids Panel	Proteinuria
Mean of Dependent Variable	0.56	0.42	0.49	0.22
Average Treatment Effect	0.01 (0.04)**	0.04 (0.00)***	0.01 (0.06)*	0.03 (0.00)***
Non-comorbid Non-compliant	0.02 (0.17)	0.02 (0.15)	0.05 (0.00)***	0.01 (0.53)
Non-comorbid Compliant	0.05 (0.00)***	0.04 (0.00)***	0.00 (0.91)	0.02 (0.07)*
Comorbid Non-compliant	0.06 (0.00)***	0.05 (0.00)***	0.10 (0.00)***	0.06 (0.00)***
Comorbid Compliant	0.08 (0.00)***	0.07 (0.00)***	0.05 (0.00)***	0.06 (0.00)***
	Health Effects			
	HbA1c Level	HbA1c ≥ 8	Inpatient Visit	Mortality [†]
Mean of Dependent Variable	8.09	0.44	0.09	0.0071
Average Treatment Effect	-0.38 (0.00)***	-0.14 (0.00)***	-0.01 (0.00)***	-0.0015 (0.04)**
Non-comorbid Non-compliant	-0.38 (0.02)**	-0.05 (0.39)	-0.03 (0.01)**	-0.0020 (0.26)
Non-comorbid Compliant	-0.41 (0.00)***	-0.16 (0.00)***	-0.01 (0.26)	-0.0001 (0.64)
Comorbid Non-compliant	-0.31 (0.06)*	0.01 (0.87)	-0.04 (0.00)***	-0.0035 (0.04)**
Comorbid Compliant	-0.34 (0.00)***	-0.11 (0.00)***	-0.02 (0.13)	-0.0016 (0.08)*
	Financial Effects			
	log Costs	log Inpatient	log Outpatient	log Pharma
Average Treatment Effect	-0.15 (0.00)***	-0.13 (0.00)***	-0.16 (0.00)***	-0.00 (0.98)
Non-comorbid Non-compliant	-0.37 (0.00)***	-0.24 (0.01)**	-0.41 (0.00)***	-0.16 (0.00)***
Non-comorbid Compliant	-0.05 (0.13)	-0.05 (0.33)	-0.00 (0.97)	0.01 (0.75)
Comorbid Non-compliant	-0.21 (0.00)***	-0.31 (0.00)***	-0.30 (0.00)***	0.19 (0.00)***
Comorbid Compliant	0.11 (0.03)**	-0.12 (0.16)	0.10 (0.14)	0.36 (0.00)***

“Long-run” is defined as greater than or equal to seven months.

P-values in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%.

Individual and quarter fixed effects are not reported.

[†]Values are marginal effects from the discrete-time hazard model in Table 13.

Table 12: Supporting Regressions

		Compliance							
	HbA1c Exam	HbA1c Exam	Retinal Exam	Retinal Exam	Lipids Panel	Lipids Panel	Lipids Panel	Protein-uria	Protein-uria
Months7+	0.017 (0.007)**	0.002 (0.033)	0.043 (0.007)**	0.164 (0.030)**	0.017 (0.007)**	-0.018 (0.032)	0.042 (0.007)**	-0.061 (0.030)**	
Baseline Compliant		0.022 (0.019)		0.057 (0.017)**		-0.059 (0.018)**		-0.013 (0.018)	
Baseline Comorbid		0.040 (0.016)**		0.022 (0.016)		0.055 (0.016)**		0.052 (0.017)**	
Observations	82814	79308	79770	76264	78790	75284	74962	71456	
Patients	5889	5888	5889	5889	5891	5890	5889	5888	
		Health Outcomes							
	log HbA1c	log HbA1c	HbA1c ≥ 7	HbA1c ≥ 7	HbA1c ≥ 8	HbA1c ≥ 8	HbA1c ≥ 8	Inpatient Visit	Inpatient Visit
Months7+	-0.045 (0.004)**	-0.040 (0.018)**	-0.037 (0.010)**	-0.018 (0.050)	-0.134 (0.011)**	-0.048 (0.055)	-0.014 (0.005)**	-0.028 (0.010)**	
Baseline Compliant		-0.006 (0.019)		-0.028 (0.051)		-0.117 (0.056)**		0.021 (0.011)*	
Baseline Comorbid		0.009 (0.009)		0.014 (0.025)		0.057 (0.027)**		-0.008 (0.011)	
Observations	31660	29037	30634	29037	30634	29037	93080	89574	
Patients	4830	4688	4706	4688	4706	4688	5891	5891	
		Costs							
	log Costs	log Costs	log Inpatient	log Inpatient	log Outpatient	log Outpatient	log Pharma	log Pharma	log Pharma
Months7+	-0.122 (0.024)**	-0.371 (0.053)**	-0.112 (0.041)**	-0.235 (0.089)**	-0.131 (0.033)**	-0.410 (0.072)**	0.027 (0.024)	-0.164 (0.053)**	
Baseline Compliant		0.323 (0.057)**		0.182 (0.097)*		0.408 (0.078)**		0.174 (0.057)**	
Baseline Comorbid		0.159 (0.056)**		-0.070 (0.095)		0.106 (0.076)		0.355 (0.056)**	
Observations	93080	89574	93080	89574	93080	89574	93080	89574	
Patients	5891	5891	5891	5891	5891	5891	5891	5891	

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%. Individual and quarter fixed effects are not reported.

Table 13: Mortality Regressions

	Matched Sample	Matched Sample
Months1-6	-0.658 (0.240)***	-0.657 (0.240)***
Months7+	-0.234 (0.114)**	-0.319 (0.285)
Compliant	0.050 (0.083)	0.029 (0.087)
Comorbid	-0.037 (0.142)	-0.010 (0.144)
Months7+ * Compliant		0.245 (0.281)
Months7+ * Comorbid		-0.227 (0.216)
Male	0.451 (0.076)***	0.453 (0.077)***
Age	0.017 (0.001)***	0.017 (0.001)***
Core	0.126 (0.170)	0.133 (0.171)
Medicare	0.072 (0.160)	0.069 (0.160)
Medicaid	0.778 (0.437)*	0.773 (0.438)*
Baseline HbA1c	0.076 (0.022)***	0.076 (0.022)***
log Total Costs	0.212 (0.035)***	0.211 (0.035)***
log Pharma	-0.027 (0.024)	-0.026 (0.024)
Hypertension	-0.117 (0.081)	-0.118 (0.081)
CAD	-0.029 (0.120)	-0.025 (0.120)
CHF	0.817 (0.115)***	0.826 (0.116)***
COPD	0.633 (0.121)***	0.632 (0.121)***
Observations	67165	67165

Standard errors in parentheses; * significant at 10%;
 ** significant at 5%; *** significant at 1%.