Who Benefits Most in Disease Management Programs: Improving Target Efficiency*

Timothy Simcoe† Maryaline Catillon ‡ Paul Gertler§

December 27, 2017

Abstract

Disease Management Programs aim to save cost by improving the quality of care for chronic diseases. Evidence for their effectiveness is mixed. Reducing health care spending sufficiently to cover program costs has proved particularly challenging. This study uses a difference in differences design to examine the impact of a Diabetes Disease Management Program for high risk patients on preventive tests, health outcomes and cost of care. Heterogeneity is examined along the dimensions of severity (measured using the proxy of poor glycemic control) and preventive testing received in the baseline year. While disease management programs tend to focus on the sickest, the impact of this program concentrates in the group of people who had not received recommended tests in the pre-intervention period. If confirmed, such findings are practically important to improve cost effectiveness in disease management programs by targeting relevant subgroups defined both based on severity and on (missing) test information.

Keywords: Disease Management, Chronic Illness, Diabetes, Heterogeneity, Program Evaluation, Target Efficiency

JEL Codes: I10

*Thanks to Derek Newell, Yang Gao, Alex Exuzides and LifeMasters Supported Self Care, Inc. for providing data and insights into LifeMasters' disease management program. This paper is a very revised version of the working paper by Gertler and Simcoe, 2009, Disease Management: Helping patients (who don't) help themselves.

†Boston University Questrom School of Business and NBER. Email: tsimcoe@bu.edu

‡Harvard Business School and NBER

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

1 Introduction

Diabetes is the sixth leading cause of death worldwide (World Health Organization 2016). Care for people diagnosed with diabetes consumes more than 1 in 5 health care dollars in the U.S. (American Diabetes Association 2013). But although diabetes is relatively well understood and there is broad consensus on how to manage it, achievement of clinical guidelines is still far from adequate (Centers for Disease Control and Prevention 2014).

Diabetes management programs aim to contain cost and improve quality of care by helping patients and clinicians follow clinical guidelines and engender healthy behaviors. Evidence on the impact of diabetes management programs is limited. Most early studies showed modest improvements in compliance with guidelines and clinical outcomes, such as HbA1c levels, but they lacked conclusive evidence that disease management programs reduce healthcare costs (Congressional Budget Office 2004; Goetzel et al. 2005; Knight et al. 2005; Mattke et al. 2007; McDonald et al. 2007; Lairson et al. 2008). Although some studies find large effects across a broad range of outcomes, including financially relevant measures such as inpatient bed days (Berg and Wadhwa 2007; Bray et al. 2008; Tricco et al. 2012), most recent studies find neither evidence of large effects on clinical outcomes (Pimouguet et al. 2011) nor significant reductions in medical expenditures (Bott et al. 2009; Brown et al. 2012; Nelson 2012).

These apparently contradictory findings can be explained by patient response heterogeneity. The relationship between disease management program intensity and impact on healthcare costs is likely nonlinear and heterogeneous (Berg et al. 2012). Decreases in medical expenditures have been shown to be larger at the most expensive tail of the healthcare cost distribution (Kranker 2016). Different mechanisms can explain these findings: more severe patients may benefit more from more intensive care, a program may cut waste in utilization, or medical expenses of high utilizers may regress to the mean (Nolte et al. 2012). Heterogeneity in the effect of disease management programs is not well understood (Elissen et al. 2013). Identifying those patients most likely to benefit from an intervention can

improve targeting, rendering programs more cost efficient and clinically effective.

We estimate the impacts of a diabetes disease management program for patients at a large Massachusetts HMO, using a difference-in-differences research design, which compares changes in outcomes for enrolled and unenrolled patients before and after the start of the program. To address internal validity concerns, we show that the treated and untreated patients are demographically similar, and control for patient-specific time-invariant unobservables in the analysis. To address external validity concerns, we note that the program includes similar components to recent programs described in the diabetes disease management litterature (Meng et al. 2016), and that, although diabetes care has improved towards more compliance with guideline recommended tests, still more than 700,000 diabetic Medicare patients annually do not receive HbA1c screening (National Committee on Quality Assurance 2016).

The program reduced both HbA1c levels and expenditures among enrolled patients. The main contribution of our study is to measure response heterogeneity along two dimensions: baseline severity and (non)compliance with clinical guidelines. Not surprisingly, patients with high baseline severity exhibit a larger reduction in HbA1c scores, while those who were not tested at baseline exhibit a larger increase in compliance. Perhaps surprisingly, the financial benefits were highly concentrated among the patients who did not receive any HbA1c test during the baseline year. The latter result suggests important policy implications. In particular, disease management programs may benefit from targeting patients based on missing information in patient records as well as severity thresholds. Our review of the literature did not find any evidence that disease management programs used targeting strategies based on missing test information, although modern EHR systems would presumably have such a capability.

The balance of the paper proceeds as follows: Section 2 provides a brief overview of targeting strategies in disease management. Section 3 describes the intervention, and Section 4 describes our data. Section 5 discusses methods. Section 6 presents the results, and we conclude with a brief discussion of implications.

2 Targeting Strategies in Disease Management

Disease management programs collect data on patients with diabetes to identify intervention opportunities. Appropriate services are then provided to the patient, the healthcare practitioners or both (Arora et al. 2008; Conklin and Nolte 2011). Prior studies have acknowledged the importance of targeting precise population groups as a determinant of a disease management intervention's impact (Dusheiko et al. 2011). Surprisingly, target efficiency in this context has received little attention beyond the dimension of severity: "If a DM program targets too large a population, it may provide costly services to beneficiaries too healthy to show significant improvements, but the program leaves potential savings unrealized if services are limited to a small population" (Kranker 2016).

Current targeting strategies are based on varied criteria reflecting severity including HbA1c level, complications, specific cutoffs in healthcare expenditures or health utilization, including hospitalizations (Conklin and Nolte 2011). Targeting strategies often focus on identifying particularly severe cases, based on diagnostic tests results (Beaulieu et al. 2006; Berwick and Hackbarth 2012; Cebul et al. 2008; Richardson et al. 2001; McGlynn et al. 2003).

Diabetes disease management is a highly information intensive activity. The most important information sought about control over diabetes is based on the level of one blood test, glycosylated hemoglobin (HbA1c), which reflects average blood glucose levels over the prior 120 days. A score less than 7 is considered good control, while a score greater than 9.5 indicates poor control. Clinical trials show a strong correlation between HbA1c levels, complications and death, which makes HbA1c score a good proxy for severity (Zoungas et al. 2012). The probability of ordering an HbA1c test is used as a measure of the quality of diabetes care (Scott et al. 2009).

Information about severity and control of diabetes is prerequisite to patients and clinicians taking appropriate measures. Patients who know their HbA1c level understand their care better, but this knowledge alone does not necessarily lead to better self-management (Heisler

et al. 2005). Both provision of information to patients and efforts to actively involve them in treatment decision-making are associated with better overall diabetes self-management (Heisler et al. 2007).

Targeting strategies based on severity alone do not enable programs to be cost effective. In 2012, 11.2 million people over 65 had diabetes (Centers for Disease Control and Prevention 2015), and among Medicare diabetic patients, 7% did not receive HbA1c screening (National Committee on Quality Assurance 2016). These figures imply that more than 700,000 Medicare diabetic patients annually lack critical information to manage their care. This is a lower bound for the size of a potential target population based on missing information.

To our knowledge, the *information hypothesis* – the idea that targeting patients based on missing information in their records may be more cost-effective than relying only on severity thresholds – has not been studied. This study therefore helps fill a gap in the literature on target efficiency of disease management programs. The proposed underlying mechanism is that information about the case is a precondition to both appropriate medical decision making and patient engagement. When patients and clinicians already have more information about a patient's health, the patient is expected to exhibit a smaller response to a program than those with less pre-intervention information, because information allows patients to respond and clinicians to intervene.

3 The Intervention

We examine a diabetes disease management intervention for beneficiaries of a large Massachusetts HMO providing health insurance for over 200,000 members. The disease management program was managed by LifeMasters, a private company specializing in chronic care management.

LifeMasters used medical claims data to identify patients with diabetes and group them into risk categories. Patients were eligible to enroll in the program if they met one or more of

the following severity criteria: hospitalization for a diabetes-related complication, an HbA1c level above 9.5, or an established diabetic complication.

At the start of the enrollment period, 5,632 persons were eligible. Primary care physicians were asked for permission to enroll their patients in the program. Physician acceptance rate was over 95%. Eligible patients were contacted by mail and telephone, receiving up to five calls during a first six-month enrollment window, at the end of which LifeMasters stopped actively recruiting patients and focused on program administration and the delivery of benefits to enrollees.

The program combined commonly used disease management strategies. Information generated by enrolled patients was monitored and used in conjunction with claims data to generate reports that were shared with physicians whenever there was a specific opportunity to intervene. Enrolled patients received electronically connected test kits and instructions for self-monitoring blood sugar. They could call an automated phone system that would answer questions, record test results, or connect them to a nurse. They could schedule regular one-on-one interactions with a case-worker. Enrolled patients were also scheduled for a series of educational sessions. LifeMasters periodically contacted them to check on their health and satisfaction with the program.

The treatment group for this study consists of eligible patients who enrolled during the first recruitment wave.[1] The control group corresponds to eligible patients who were not enrolled during the first recruitment wave. Exogenous assignment is not assumed, since patients who were contacted could still decline the offer to enroll, and identification of contacted patients is not available. However, our empirical strategy leverages inaccurate phone records that created substantial random variation in offers to enroll. The company was only able to reach approximately half of all eligible patients during the first enrollment wave. Later waves of enrollees are excluded from the treatment group as they were targeted based on recent health shocks.

[1] Patients who enrolled in later waves were more likely to select into the program based on adverse health outcomes, but were retained in the control group up to their date of enrollment.

4 Data

We utilize an unbalanced panel data set containing 192,255 quarterly observations for 5,632 eligible patients over a four year period.[2] Table 1 reports patient demographics, test information, health outcomes, and costs for eligible patients during the baseline year, in the treated and control groups, with the corresponding t-statistic and normalized difference.[3]

The panel on demographics includes patient's age, sex, insurance status, residential area (urban or rural), race, attrition status and income. Patient's age, sex, insurance status (Medicare and/or Medicaid), mailing address, medical claims, lab test information and attrition status come from the HMO dataset. Race and income were generated by linking patients' zip codes to U.S. Census data.

The panel on preventive tests includes a set of four binary variables representing whether or not a patient received each of the four recommended tests (HbA1c Test, Eye Examination, Lipid Level and Renal Function Test) during the baseline (pre-enrollment) year. A variable "Total Tests", corresponds to the sum of the four binary tests variable and summarizes how many of the four tests were received in the pre-intervention period.

The panel on health outcomes provides four measures: the average HbA1c level in the baseline year, severity (measured following Zoungas et al. (2012) using HbA1c $\geq$ 9.5, the threshold for poor control, as a proxy), a binary variable for comorbidities and a binary variable indicating whether a patient had an inpatient admission in the pre-intervention period. The cost variable includes total healthcare expenditures including inpatient, outpatient and pharmaceutical claims.

The treatment and control groups have very similar baseline demographics. The eligible population is on average 65.5 years old, 53% male, 60% covered by Medicare and 2% covered

[2]The sample period begins in the third quarter of 1998 and ends in the third quarter of 2002.

[3]The normalized difference is the difference in treated versus control sample means, divided by the average standard deviation. It provides a scale-free measure of the difference in the location of the two distributions that does not depend on sample size in the same manner as a t-statistic. While there are no critical values for this statistic, Imbens and Wooldridge (2008) suggest that when it exceeds 0.25, linear regression methods tend to be sensitive to the specification.

by Medicaid, 84% urban, 90% white. Their average annual income is $49,279 and their attrition rate 12%. The treated group is slightly younger than the control group (64.6 versus 65.7 years old). All other demographic differences between the treated and control group are small and not significant.

The treatment and control groups are different with respect to preventive tests. In the baseline year 73% of the eligible population received an HbA1c test, 57% received an eye test, 48% received a lipid test and 8% received a kidney test. Treated patients are uniformly more likely to be tested than control patients, and received on average two or more tests in the baseline year (2.21) while control patients received on average less than two tests (1.80) in the baseline year. 89% of treated patients versus 71% of control patients received an HbA1c test. 63% of treated patients versus 56% of control patients received an eye test. 57% of treated patients versus 47% of control patients received a lipid test and 12% of treated patients versus 7% of control patients received a kidney test.

The health outcome and cost panels show both contrasts and similarities between the treated and control patients. While the average HbA1c score in the pre-intervention period for all eligible patients is 8.32, this score is higher in the treated group (9.02) than in the control group (8.13). Similarly, 21% of the eligible population had poor HbA1c control, but this proportion is 33% in the treated group versus 18% in the control group. While this is a significant difference in HbA1c levels, this may reflect that treated patients are more likely to be tested. In contrast, the comorbidity rate, inpatient visits and healthcare costs are slightly higher in the treated group than in the control group, but these differences are small and not statistically significant. The comorbidity rate is 30% in the treated group versus 27% in the control group. Inpatient visits rate is 25% in the treated group versus 22% in the control group. Total annual cost is $9,629 in the treated group versus $8,683 in the control group.

Overall, treated and control patients have very similar demographics, comorbidity rates, inpatient visits and healthcare costs. They are however different with respect to preventive tests received. While this might raise concerns about self-selection into the program, treat-

ment effect heterogeneity along the dimension of HbA1c test received in the baseline year is examined below.

5 Methods

Our evaluation examines three outcome variables: the number of recommended tests received, severity (measured following Zoungas et al. (2012) using HbA1c ≥ 9.5, the threshold for poor control, as a proxy for severity) and total cost. Heterogeneity in the treatment effect is analyzed along the dimensions of preventive tests received in the baseline year and baseline severity.

We use a difference in differences (DD) design to measure the average impact of the disease management program on enrolled patients. A major concern is that patients who choose to enroll in the DM program may differ from those who decline, and these differences could correlate with observed outcomes. A DD model compares the change in outcomes for patients in the treatment group to the change in outcomes for patients in the control group. The treatment group difference controls for time-invariant patient and environmental characteristics that might be correlated with both enrollment and health. The control group difference captures any time-varying factors common to all patients.

The basic DD model (model (1) in tables 2, 3 and 4) can be specified as this two-way, fixed effect, linear regression:

$$y_{itk} = \alpha_k T_{ik} + \beta X_{it} + \gamma_i + \lambda_t + \epsilon_{it} \tag{1}$$

where y_{itk} is an outcome for individual i who is in the k^{th} quarter of enrollment at time t, T_{ik} is an indicator variable that takes the value of 1 if individual i is in their k^{th} enrollment quarter 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. The standard errors are clustered at the individual patient level to control for possible serial correlation.

The parameter α_k is an estimate of the average impact of the disease management program in the k^{th} quarter of treatment. The key assumption in this model is that the change in the control group is an unbiased estimate of the change the treatment group would have experienced had they not enrolled in the disease management program. While it is not possible to test this assumption directly, data from the pre-enrollment period can serve as supporting evidence. In particular, we conduct a test of the joint null hypothesis that $\alpha_k = 0$ for all $k < 0$. Rejecting this "parallel trends" null hypothesis would provide evidence against the key DD identification assumption.

Many of our results can be presented graphically, by plotting the α_k and their 95% confidence intervals. These graphs illustrate both the pre-intervention outcome trends, and any treatment effect that follows enrollment in the disease management program. Figures 1 and 2 below provide graphs for the complete sample, and also for samples stratified by either the number of tests received at baseline (Figure 1) and severity at baseline (Figure 2).

In addition to the flexible specification of model (1), we estimate a standard DD specification that pools all of the post-treatment α_k into a single parameter. This permits a more parsimonious exploration of treatment heterogeneity using regressions with interactions between a dummy for treatment T_{it} and indicators for baseline testing or severity. Once again, the model is specified as a two-way fixed effects linear regression:

$$y_{it} = \alpha T_{it} + \delta T_{it} Z_i + \beta X_{it} + \gamma_i + \lambda_t + \epsilon_{it} \tag{2}$$

where y_{it} is an outcome for individual i at time t, T_{it} is an indicator variable that takes the value of 1 if individual i is enrolled at time 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.

The treatment heterogeneity in this model comes from interacting T_{it} with Z_i: a binary variable indicating that a patient received at least one HbA1c test at baseline or, in the second specification, that the patient's baseline HbA1c score was less than 9.5 (the threshold for poor control).

In model (2), the parameter α represents the average treatment effect for people who did not receive any HbA1c test at baseline (alternatively, those with low severity). The estimated treatment effect for those with a baseline HbA1c test (or high severity) is equal to $\alpha + \delta$. The main effect of the time-invariant baseline status indicators is absorbed by the individual fixed effects, and we take into account the possibility of different time-trends for the heterogeneous sub-populations by allowing the λ_t to vary across groups defined by Z_i.

6 Results

This disease management program had large and significant effects on severity and total costs. The effects of the program appeared within one or two quarters of enrollment and persisted over time during the observed period. Analysis of response heterogeneity reveals that patients who received less than two of the four recommended tests at baseline (rather than the more severe patients at baseline) benefitted the most.

6.1 Graphical DD Results

The effects of the program on the number of recommended tests received, severity and total cost are presented in Figure 1 and Figure 2.[4] Each panel plots the α_k coefficients and associated 95% confidence intervals for a different sub-sample. Figure 1 shows all three outcomes (one per row) for the full sample (column 1), as well as the two subsamples receiving less than two recommended tests at baseline (column 2) and two or more recommended tests at baseline (column 3). Figure 2 shows the same graphs for all patients who received at least

[4]See Tables A-1 and A-2 in the Appendix for coefficient estimates from a similar specification.

one HbA1c test at baseline, as well as the subsample of more severe patients with HbA1c above the 9.5 threshold at baseline and less severe patients, with HbA1c below the 9.5 threshold at baseline.

For the full sample, we observe no effect of the program on the number of recommended tests received, but a large and rapid effect on severity and costs. The most interesting insight from Figure 1, however, is that the program's effect on cost is concentrated among patients who received less than two recommended tests at baseline. This can be seen by comparing the graphs in the second and third columns of the bottom row of Figure 1. The effect of the DM program on severity appears to be larger but noisier in the sample of patients who received less than two recommended tests at baseline. Most of these patients were not tested before enrolling, making it difficult to precisely measure any change in the mean HbA1c score.

Classifying patients into different severity subgroups based on their HbA1c score obviously requires that they received at least one HbA1c test during the baseline year. As a result, people with no HbA1c test at baseline cannot be classified as high or low severity. Figure 2 shows DD results for (1) the full sample excluding people with no HbA1c score at baseline, (2) the high severity at baseline subsample and (3) the low severity at baseline subsample. The most interesting insight from Figure 2 is that when people with no HbA1c score at baseline are excluded from the full sample (column 1), a statistically significant effect of the program on severity persists, but the impacts on cost disappear. Contrary to the assumption behind classic targeting strategies that the sickest may benefit the most and potentially generate the bigger cost savings, this program has no effect on total cost for high severity patients who received at least one HbA1c test at baseline.

6.1.1 Pre-trend falsification tests

Overall, we find small differences between the treated and control patients before the intervention. Graphically, we see that zero falls within the 95% confidence interval for most

pre-intervention coefficients plotted in Figures 1 and 2. F-tests found no significant difference in the pre-treatment trend for log costs or severity. The pre-intervention difference in the number of tests received is addressed by splitting the sample along this dimension.

6.1.2 Timing of DM program impacts

Figures 1 and 2 also show how quickly the DM program works. The response is remarkably rapid for patients receiving less than two recommended tests at baseline (Figure 1, middle column). By the second quarter of enrollment, these patients are 20% more likely to receive two or more tests; their total costs decline by 65 percent, and poor HbA1c scores fall by 20 percent relative to the unenrolled patients. Moreover, all of these improvements persist for several years.

For patients receiving two or more tests at baseline (Figure 1, right column), the benefits are smaller. For example, after the fourth enrollment quarter information increases by 0.1 tests per patient, HbA1c control improves by 10 to 15 percent, and costs are not significantly reduced. These differences in the timing and magnitude of the estimated treatment effects suggest that DM interventions are more effective for patients who receive fewer recommended tests at baseline.

Overall, we interpret the results in Figure 1 and 2 as evidence both that treated patients are not systematically different from those who did not enroll, and that they experienced sharp changes in recommended tests received, severity and cost following admission to the program. There is also evidence that the program's impact was greater for patients who received fewer recommended tests at baseline.

6.2 Response heterogeneity

Tables 2, 3 and 4 report estimates from regression models with a single treatment dummy, and add interactions to explore response heterogeneity along the dimensions of baseline

HbA1c test, and baseline severity.[5] Three outcomes are examined: number of tests received (Table 2), severity (Table 3) and total cost (Table 4).

In each table, column 1 presents the regression of the outcome variable on a single treatment dummy. Column 2 adds an interaction between treatment and a dummy variable for receiving at least one HbA1c test at baseline. Column 3 presents the regression of the outcome variable on a single treatment dummy and adds an interaction between treatment and baseline severity, using a dummy variable for HbA1c less than the 9.5 threshold. Column 3 has a smaller number of observations because information on baseline severity is only available for the subset of patients with at least one HbA1c test in the pre-intervention period.

Table 2 presents the effect of the program on number of tests received. In the basic model, the program has small and non-significant effects on the number of tests received. However, the specification with interaction between treatment and an indication for receiving at least one HbA1c test at baseline shows a large and significant improvement in the number of tests received among the group of patients who did not receive any HbA1c test at baseline.

Table 3 shows large and significant effects on health outcomes in models (1) and (3) (without interaction and with interaction between treatment and baseline severity). The larger effect on severity appears in the high severity group. However, comparing effects in this panel needs caution: in model (2) pre-post comparison using HbA1c score is likely to underestimate the effect of the program on severity in the people who did not receive any HbA1c tests at baseline. In this table, standard errors in column 2 are much larger than standard errors in column 1 and 3 because there is no data on severity at baseline for patients who did not receive an HbA1c test in the pre-intervention period.

The most interesting insights from these regressions are presented in Table 4, showing

[5]We checked the balance of the covariates in the subgroup of patients with no HbA1c test at baseline. Overall, the balance looks good (see Table A-3), but with some evidence of inbalance in baseline comorbidities and inpatient visits. To verify that this inbalance is not driving the results, we ran regression models in the subsample without HbA1c test at baseline, adding interactions between treatement and baseline comorbidites and between treatment and baseline inpatient visits. We did not find significant heterogeneity in the treatment effect along these dimensions.

the impact of the program on cost. Since these variables enter in logs, we can interpret the coefficient estimates as percentage changes. An estimate of the savings per patient-quarter in US dollars is obtained from a separate regression run in levels.

The program has a large and significant impact on total cost, with an overall decrease of 22 percent, corresponding to $209 per quarter. The regression specification adding an interaction term between treatment and an indicator for receiving at least one HbA1c test at baseline shows that the effect is large and highly significant for the patients who did not receive any HbA1c test at baseline (coefficient is -0.66 corresponding to a decrease of $617 per quarter). However, the regression specification adding an interaction term between treatment and baseline severity shows that the effect in the high severity group is much smaller and not significant. The effects of the program concentrate among the group of patients who did not receive any HbA1c test at baseline, corresponding to a sample of 91 patients, 10% of the treated group, where the cost savings are concentrated. This finding has significant implications for targeting interventions towards patient populations that are likely to yield a large financial return.

7 Discussion

Using data from a large Massachusetts HMO, this paper explores the impact of a disease management program and its response heterogeneity. The disease management program had large and significant impacts on the number of recommended tests received, health outcomes and cost. Most of the benefits, especially the financial ones, were realized in patients who did not receive any HbA1c test at baseline, which suggests important policy implications.

Most studies use changes in HbA1c levels as the outcome measure to assess the effect of disease management programs, which requires access to baseline HbA1c scores. As a result, although the importance of diagnostic tests in diabetes disease management is well-known, the effect of disease management programs on people who do not get HbA1c tests in the

pre-intervention period – the population we find most likely to get large benefits from such programs – is not known.

One important implication of our study is that observational studies assessing the impact of disease management programs based on changes in observable variables like HbA1c levels may underestimate the effect of the programs they aim to evaluate, because these effects concentrate in people about whom this information is not available in the pre-intervention period. Thus the findings suggest that in order to maximize returns disease management programs should target patients who do not receive recommended tests.

Using a published cost estimate of $117 per patient-year (FROM WHERE? CITE NEEDED), simple calculations suggest that, for patients who did not receive recommended tests at baseline the program saved $2,351 per patient year.[6] Had all the patients who did not receive HbA1c test at baseline been treated, this estimate suggests that the program could have saved $3.4 million in the first year alone.[7] Thus, our results indicate that, focusing on patients who do not receive recommended tests at baseline, disease management programs could produce significant cost savings in the short run. If similar savings were achievable for the 700,000 Medicare diabetic patients who do not receive HbA1c screening, more than $1.5 billion could be saved annually on diabetes care.

Our study also has several limitations. We are not able to quantify the net financial impacts of the program because we lack information on administrative costs. The sub-sample of patients with missing test information who enrolled in the DM program is also relatively small (94 patients), suggesting the need for follow-up research to assess the replicability and generalizability of our results for that group. Another promising avenue for future research is to examine how to develop methods to identify high risk patients who do not receive recommended diagnostic tests, based on electronic health records. These patients may be the most promising target for health programs. At the same time, because they do not

[6]Using the estimate from Table 4 for patients who did not receive HbA1c test at baseline, the program reduced patient quarter expenditures by $617, thus by $2,468 annually.

[7]There were 1,430 patients who did not receive HbA1c test at baseline in the control group

receive recommended diagnostic tests, they may not be readily identified as high risk using available information.

8 Conclusion

Data from a large Massachusetts HMO show that a disease management program had large and significant effects on health outcomes and costs. Patients who benefit the most are not the more severe patients, but rather patients who were not recently tested, thus preventing themselves and their doctors from taking appropriate measures. This suggests an important new strategy to improve target efficiency in disease management programs.

References

Abadie, A. and G. W. Imbens (2011). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics 29*(1), 1–11.

American Diabetes Association (2013). Economic costs of diabetes in the us in 2012. *Diabetes care 36*(4), 1033–1046.

Arora, R., J. Boehm, L. Chimento, L. Moldawer, and C. Tsien (2008). Designing and implementing medicaid disease and care management programs: A user's guide. *AHRQ Publication* (07), 08.

Beaulieu, N., D. M. Cutler, K. Ho, G. Isham, T. Lindquist, A. Nelson, P. O'Connor, et al. (2006). The business case for diabetes disease management for managed care organizations. In *Forum for Health Economics & Policy*, Volume 9, pp. 1–36. bepress.

Berg, G. D., S. Donnelly, M. Miller, W. Medina, and K. Warnick (2012). Dose-response effects for disease management programs on hospital utilization in illinois medicaid. *Population health management 15*(6), 352–357.

Berg, G. D. and S. Wadhwa (2007). Health services outcomes for a diabetes disease management program for the elderly. *Disease Management 10*(4), 226–234.

Berwick, D. M. and A. D. Hackbarth (2012). Eliminating waste in us health care. *Jama 307*(14), 1513–1516.

Bott, D. M., M. C. Kapp, L. B. Johnson, and L. M. Magno (2009). Disease management for chronically ill beneficiaries in traditional medicare. *Health Affairs 28*(1), 86–98.

Bray, K., R. S. Turpin, K. Jungkind, and G. Heuser (2008). Defining success in diabetes disease management: digging deeper in the data. *Disease Management 11*(2), 119–128.

Brown, R., D. R. Mann, et al. (2012). Best bets for reducing medicare costs for dual eligible beneficiaries assessing the evidence. Technical report, Mathematica Policy Research.

Brown, R. S., D. Peikes, G. Peterson, J. Schore, and C. M. Razafindrakoto (2012). Six features of medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. *Health Affairs 31*(6), 1156–1166.

Cebul, R. D., J. B. Rebitzer, L. J. Taylor, and M. E. Votruba (2008). Organizational fragmentation and care quality in the us healthcare system. *The Journal of Economic Perspectives 22*(4), 93–113.

Centers for Disease Control and Prevention (2014). National diabetes statistics report: estimates of diabetes and its burden in the united states, 2014. *Atlanta, GA: US Department of Health and Human Services 2014*.

Centers for Disease Control and Prevention (2015). Diabetes report card, 2014, atlanta, ga: Centers for disease control and prevention, us dept. of health and human services, 2015.

Congressional Budget Office (2004). An analysis of the literature on disease management programs. washington, dc: Congressional budget office; 2004 october 13.

Conklin, A. and E. Nolte (2011). Disease management evaluation: A comprehensive review of current state of the art. *Rand health quarterly 1*(1).

Dusheiko, M., H. Gravelle, S. Martin, N. Rice, and P. C. Smith (2011). Does better disease management in primary care reduce hospital costs? evidence from english primary care. *Journal of health economics 30*(5), 919–932.

Elissen, A. M., L. M. Steuten, L. C. Lemmens, H. W. Drewes, K. M. Lemmens, J. A. Meeuwissen, C. A. Baan, and H. J. Vrijhoef (2013). Meta-analysis of the effectiveness of chronic care management for diabetes: investigating heterogeneity in outcomes. *Journal of Evaluation in Clinical Practice 19*(5), 753–762.

Goetzel, R. Z., R. J. Ozminkowski, V. G. Villagra, and J. Duffy (2005). Return on investment in disease management: a review. *Health care financing review 26*(4), 1.

Heisler, M., I. Cole, D. Weir, E. A. Kerr, and R. A. Hayward (2007). Does physician communication influence older patients' diabetes self-management and glycemic control? results from the health and retirement study (hrs). *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 62*(12), 1435–1442.

Heisler, M., J. D. Piette, M. Spencer, E. Kieffer, and S. Vijan (2005). The relationship between knowledge of recent hba1c values and diabetes care understanding and self-management. *Diabetes care 28*(4), 816–822.

Imbens, G. and J. Wooldridge (2008). Recent developments in the econometrics of program evaluation. *NBER Working Paper No. 14251*.

Knight, K., E. Badamgarav, J. M. Henning, V. Hasselblad, A. D. Gano Jr, J. J. Ofman, and S. R. Weingarten (2005). A systematic review of diabetes disease management programs. *Am J Manag Care 11*(4), 242–50.

Kranker, K. (2016). Effects of medicaid disease management programs on medical expenditures: Evidence from a natural experiment in georgia. *Journal of health economics 46*, 52–69.

Lafata, J. E., H. L. Morris, E. Dobie, M. Heisler, R. M. Werner, and L. Dumenci (2013). Patient-reported use of collaborative goal setting and glycemic control among patients with diabetes. *Patient education and counseling 92*(1), 94–99.

Lairson, D. R., S.-J. Yoon, P. M. Carter, A. J. Greisinger, K. C. Talluri, M. Aggarwal, and O. Wehmanen (2008). Economic evaluation of an intensified disease management system for patients with type 2 diabetes. *Disease Management 11*(2), 79–94.

Mattke, S., M. Seid, and S. Ma (2007). Evidence for the effect of disease management: is $1 billion a year a good investment? *American Journal of Managed Care 13*(12), 670.

McDonald, K., C. Conlon, and M. Ledwidge (2007). Disease management programs for heart failure: not just for the 'sick'heart failure population. *European journal of heart failure 9*(2), 113–117.

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), 2635–2645.

Meng, Y.-Y., A. Diamant, J. Jones, W. Lin, X. Chen, S.-H. Wu, N. Pourat, D. Roby, and G. F. Kominski (2016). Racial and ethnic disparities in diabetes care and impact of vendor-based disease management programs. *Diabetes care 39*(5), 743–749.

Nathan, D. M. (2015). Diabetes: advances in diagnosis and treatment. *Jama 314*(10), 1052–1062.

National Committee on Quality Assurance (2016). The state of health care quality report.

Nelson, L. (2012). *Lessons from Medicare's demonstration projects on disease management and care coordination*. Congressional Budget Office Washington, DC.

Nolte, E., A. Conklin, J. L. Adams, M. Brunn, B. Cadier, K. Chevreul, I. Durand-Zaleski, A. Elissen, A. Erler, M. Flamm, et al. (2012). Evaluating chronic disease management.

Pimouguet, C., M. Le Goff, R. Thiébaut, J. F. Dartigues, and C. Helmer (2011). Effectiveness of disease-management programs for improving diabetes care: a meta-analysis. *Canadian Medical Association Journal 183*(2), E115–E127.

Richardson, W. C., D. M. Berwick, J. Bisgard, L. Bristow, C. Buck, C. Cassel, et al. (2001). Crossing the quality chasm: a new health system for the 21st century.

Scott, A., S. Schurer, P. H. Jensen, and P. Sivey (2009). The effects of an incentive program on quality of care in diabetes management. *Health economics 18*(9), 1091–1108.

Tricco, A. C., N. M. Ivers, J. M. Grimshaw, D. Moher, L. Turner, J. Galipeau, I. Halperin, B. Vachon, T. Ramsay, B. Manns, et al. (2012). Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. *The Lancet 379*(9833), 2252–2261.

World Health Organization (2016). *Global report on diabetes*. World Health Organization.

Zoungas, S., J. Chalmers, T. Ninomiya, Q. Li, M. Cooper, S. Colagiuri, G. Fulcher, B. De Galan, S. Harrap, P. Hamet, et al. (2012). Association of hba1c levels with vascular complications and death in patients with type 2 diabetes: evidence of glycaemic thresholds. *Diabetologia 55*(3), 636–643.

Tables and Figures

Table 1: Baseline Summary Statistics: Eligible Patients, Treated versus Controls

	Eligible	Treated	Control	Norm. Diff.	T-stat
			Demographics		
Age	65.5	64.6	65.7	0.09	2.28
Male	0.53	0.51	0.54	0.05	1.39
Medicare	0.60	0.59	0.61	0.04	1.20
Medicaid	0.02	0.02	0.03	0.06	1.48
Urban	0.84	0.84	0.84	0.00	0.10
White	0.90	0.90	0.89	0.05	1.41
Attriter	0.12	0.12	0.13	0.01	0.15
Income	49,279	49,447	49,249	0.01	0.40
			Tests		
HbA1c Test	0.73	0.89	0.71	0.48	11.51
Eye Exam	0.57	0.63	0.56	0.16	4.18
Lipids	0.48	0.57	0.47	0.18	4.94
Kidney	0.08	0.12	0.07	0.16	4.67
Total Tests	1.85	2.21	1.80	0.41	10.91
			Health Outcomes		
HbA1c Level	8.32	9.02	8.13	0.51	12.09
HbA1c $\geq$ 9.5	0.21	0.33	0.18	0.34	8.72
Comorbidities	0.27	0.30	0.27	0.07	1.79
Inpatient visit	0.22	0.25	0.22	0.07	1.94
			Healthcare Costs		
Total Cost	8,825	9,629	8,683	0.06	1.60
Patients	5,632	848	4,784		

Table 2: DM Impact on Total Tests

Outcome: Total Tests (Count) Unit of Analysis: Patient-Quarter			
	(1)	(2)	(3)
Treated_{it}	0.04	0.23	0.05
	(0.03)	(0.09)**	(0.04)
Treated_{it} x 1[HbA1c Test]		-0.18	
		(0.09)*	
Treated_{it} x 1[HbA1c<9.5]			0.003
			(0.06)
Patient Effects	Y	Y	Y
Quarter Effects	Y	N	Y
Quarter x HbA1c Effects	N	Y	n/a
Observations	70,625	70,625	45,018
Patients	4,966	4,966	3,098
Joint F test		2.35	1.41
Prob >F		0.12	0.23

*Significant at 5%; **significant at 1%. Robust standard errors clustered on patient in parentheses. Quarter x HbA1c effects allow separate time-trends for patients with and without a baseline HbA1c test.

Table 3: DM Impact on Severity (HbA1c $\geq$ 9.5)

Outcome: Severity (HbA1c > 9.5) Unit of Analysis: Patient-Quarter			
	(1)	(2)	(3)
Treated_{it}	-0.10	0.12	-0.30
	(0.01)**	(0.19)	(0.02)**
Treated_{it} x 1[HbA1c Test]		-0.21	
		(0.19)	
Treated_{it} x 1[HbA1c<9.5]			0.41
			(0.02)**
Patient Effects	Y	Y	Y
Quarter Effects	Y	N	Y
Quarter x HbA1c Effects	N	Y	n/a
Observations	23,569	23,569	20,752
Patients	3,938	3,938	3,098
Joint F test		43	101
Prob >F		0	0

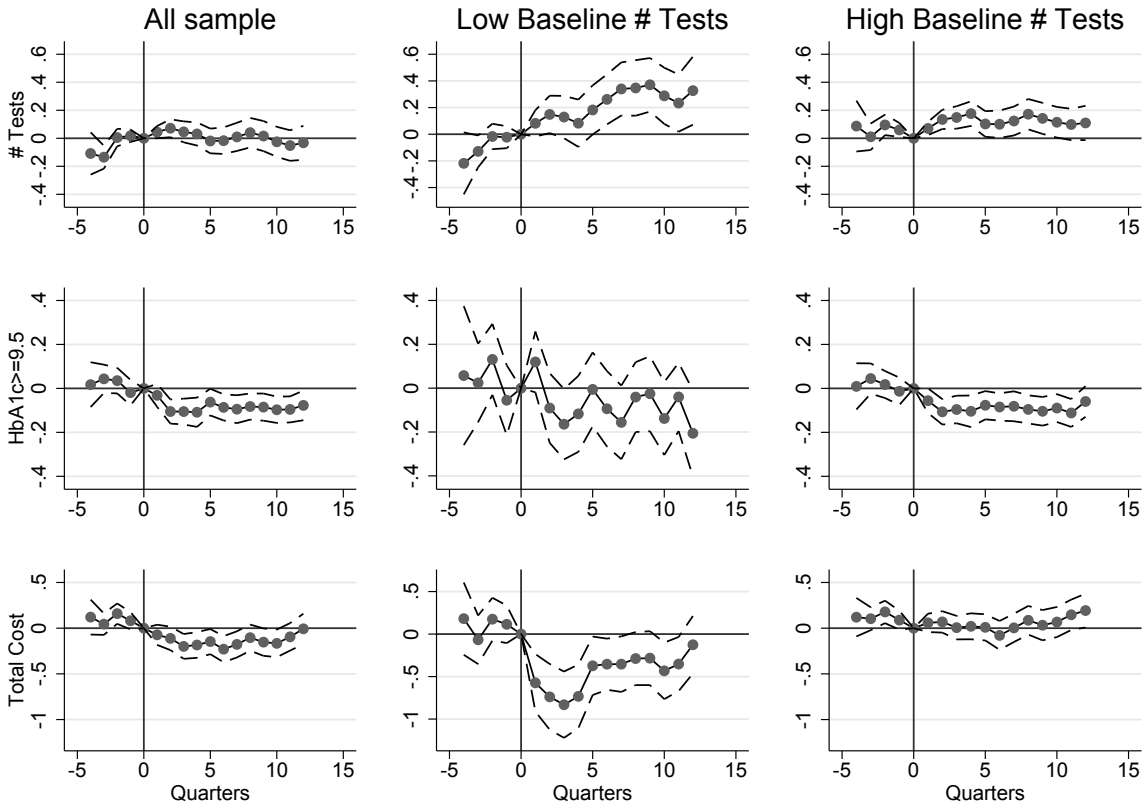
*Significant at 5%; **significant at 1%. Robust standard errors clustered on patient in parentheses. Quarter x HbA1c effects allow separate time-trends for patients with and without a baseline HbA1c test.

Table 4: DM Impact on Healthcare Cost

Outcome: log(Total Cost+$1) Unit of Analysis: Patient-Quarter			
	(1)	(2)	(3)
$Treated_{it}$	-0.22	-0.66	-0.09
	(0.04)**	(0.11)**	(0.06)
	[-$209]	[-$617]	[-$202]
$Treated_{it}$ x 1[HbA1c Test]		0.60	
		(0.12)**	
		[+$511]	
$Treated_{it}$ x 1[HbA1c<9.5]			0.06
			(0.07)
			[+$183]
Patient Effects	Y	Y	Y
Quarter Effects	Y	N	Y
Quarter x HbA1c Effects	N	Y	n/a
Observations	70,624	70,624	45,017
Patients	4,966	4,966	3,098
Joint F test		1.94	0.25
Prob >F		0.16	0.61

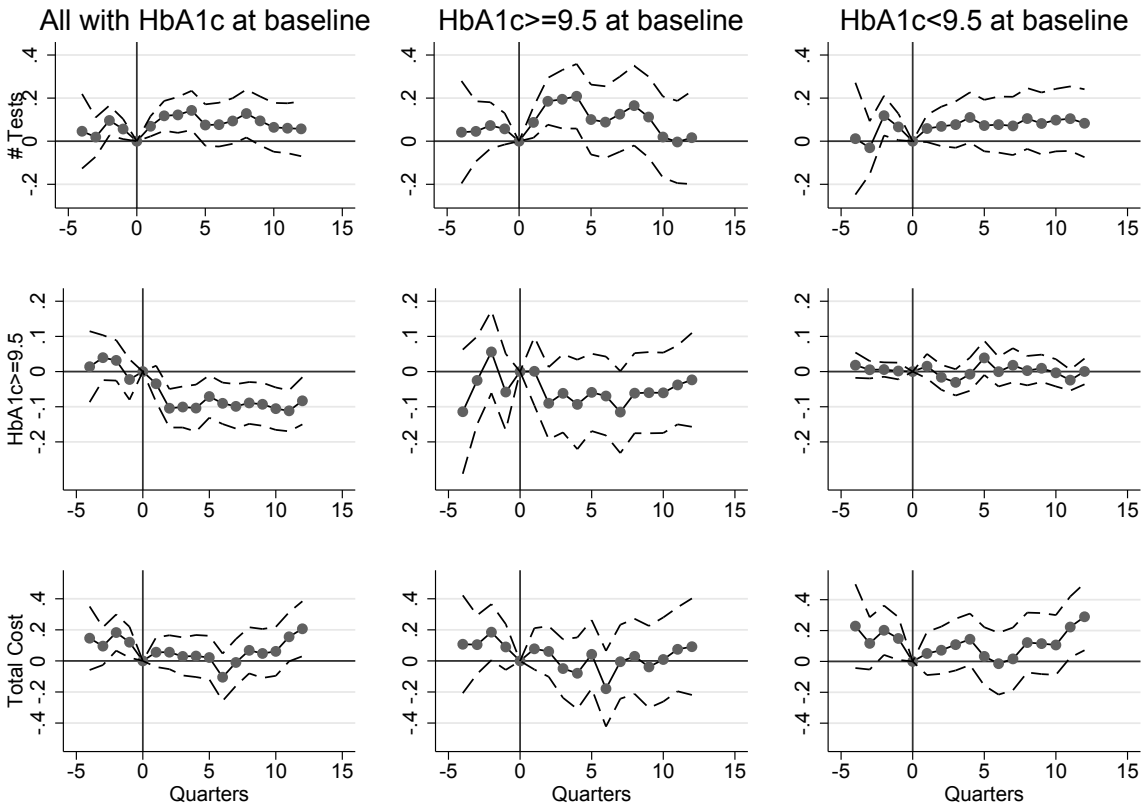
*Significant at 5%; **significant at 1%. Robust standard errors clustered on patient in parentheses. Quarter x HbA1c effects allow separate time-trends for patients with and without a baseline HbA1c test. OLS coefficients are presented in brackets for ease of interpretation.

Figure 1: Difference in Differences: Heterogeneity in Baseline Testing



Each panel graphs the coefficients α_k from OLS estimation of equation (1), along with a 95% confidence interval (based on robust SE's clustering on patient). We normalize $\alpha_0 = 0$ in the quarter before DM program enrollment begins. Each row is a different outcome and each column is a different sample. See text for more details.

Figure 2: Difference in Differences: Heterogeneity in Baseline HbA1c



Each panel graphs the coefficients α_k from OLS estimation of equation (1), along with a 95% confidence interval (based on robust SE's clustering on patient). We normalize $\alpha_0 = 0$ in the quarter before DM program enrollment begins. Each row is a different outcome and each column is a different sample. See text for more details.

Appendix: For Online Publication

Table A-1: DM Impact by Enrollment Quarter: Information Heterogeneity

Sample	Baseline Tests <2			Baseline Tests ≥ 2		
Outcome	Tests	Severity	log(Cost)	Tests	Severity	log(Cost)
$\text{Pre}_{k\leq -2}$	0.07	-0.12	0.12	0.00	-0.04	-0.04
	(0.04)*	(0.07)	(0.09)	(0.03)	(0.02)	(0.05)
$\text{Pre}_{k=-1}$	0.10	-0.08	-0.00	-0.06	-0.03	-0.13
	(0.05)	(0.06)	(0.12)	(0.04)	(0.03)	(0.05)*
$\text{Pre}_{k=0}$	0.0	0.0	0.0	0.0	0.0	0.0
	(na)	(na)	(na)	(na)	(na)	(na)
$\text{Post}_{k=1}$	0.18	-0.02	-0.57	0.00	-0.09	-0.07
	(0.06)**	(0.06)	(0.16)**	(0.04)	(0.02)**	(0.05)
$\text{Post}_{k=2}$	0.21	-0.20	-0.65	0.07	-0.14	-0.04
	(0.08)**	(0.06)**	(0.17)**	(0.04)	(0.02)**	(0.05)
$\text{Post}_{k=3}$	0.22	-0.19	-0.75	0.09	-0.12	-0.12
	(0.08)**	(0.06)**	(0.18)**	(0.04)*	(0.02)**	(0.06)*
$\text{Post}_{k=4}$	0.16	-0.23	-0.67	0.11	-0.14	-0.10
	(0.08)	(0.07)**	(0.16)**	(0.04)*	(0.03)**	(0.06)
$\text{Post}_{k\geq 5}$	0.31	-0.16	-0.34	0.05	-0.11	-0.15
	(0.08)**	(0.06)**	(0.12)**	(0.05)	(0.02)**	(0.06)*
Observations	20,476	4,105	20,476	38,311	14,819	38,310
Patients	1,979	1,281	1,979	3,653	3,011	3,653

Table reports coefficients from OLS estimation of equation (1). *Significant at 5%; **significant at 1%. Robust standard errors clustered on patient in parentheses. All models include Quarter x HbA1c effects to allow separate time-trends for patients with and without a baseline HbA1c test.

Table A-2: DM Impact by Enrollment Quarter: Severity Heterogeneity

Sample	Baseline HbA1c $\geq$ 9.5			Baseline HbA1c <9.5		
Outcome	Tests	Severity	log(Cost)	Tests	Severity	log(Cost)
$\text{Pre}_{k\leq -2}$	0.01	-0.07	-0.05	0.03	-0.01	-0.01
	(0.05)	(0.04)	(0.07)	(0.04)	(0.01)	(0.07)
$\text{Pre}_{k=-1}$	-0.05	-0.01	-0.14	-0.04	-0.02	-0.15
	(0.06)	(0.05)	(0.08)	(0.05)	(0.01)	(0.07)*
$\text{Pre}_{k=0}$	0.0	0.0	0.0	0.0	0.0	0.0
	(na)	(na)	(na)	(na)	(na)	(na)
$\text{Post}_{k=1}$	0.04	-0.02	-0.04	0.02	-0.00	-0.09
	(0.07)	(0.04)	(0.08)	(0.05)	(0.02)	(0.06)
$\text{Post}_{k=2}$	0.15	-0.11	-0.08	0.03	-0.03	-0.07
	(0.07)*	(0.04)**	(0.08)	(0.06)	(0.02)	(0.07)
$\text{Post}_{k=3}$	0.14	-0.07	-0.21	0.05	-0.04	-0.07
	(0.07)*	(0.04)	(0.09)*	(0.06)	(0.02)*	(0.07)
$\text{Post}_{k=4}$	0.14	-0.11	-0.24	0.06	-0.03	-0.01
	(0.07)*	(0.05)*	(0.11)*	(0.06)	(0.02)	(0.07)
$\text{Post}_{k\geq 5}$	0.06	-0.09	-0.19	0.02	0.01	-0.14
	(0.07)	(0.04)*	(0.10)	(0.06)	(0.01)	(0.07)
Observations	11,222	5,487	11,222	25,365	11,682	25,364
Patients	1,073	1,073	1,073	2,385	2,385	2,385

Table reports coefficients from OLS estimation of equation (1). *Significant at 5%; **significant at 1%. Robust standard errors clustered on patient in parentheses. All models include Quarter x HbA1c effects to allow separate time-trends for patients with and without a baseline HbA1c test.

Table A-3: Baseline Summary Statistics: Eligible Patients without HbA1c test at baseline, Treated versus Controls

	Eligible (no HbA1c)	Treated	Control	Norm. Diff.	T-stat
		Demographics			
Age	63.6	63.0	63.6	0.05	0.38
Male	0.53	0.51	0.53	0.05	0.49
Medicare	0.51	0.37	0.52	0.30	2.73
Medicaid	0.04	0.03	0.04	0.04	0.37
Urban	0.85	0.89	0.85	0.16	1.38
White	0.89	0.89	0.89	0.03	0.26
Attriter	0.11	0.10	0.12	0.06	0.55
Income	48,761	48,359	48,787	0.03	0.29
		Tests			
HbA1c Test	0	0	0	.	.
Eye Exam	0.47	0.58	0.47	0.22	1.99
Lipids	0.25	0.12	0.26	0.36	2.96
Kidney	0.01	0.04	0.01	0.23	3.37
Total Tests	0.73	0.74	0.72	0.02	0.20
		Health Outcomes			
HbA1c Level	N/A	N/A	N/A	N/A	N/A
HbA1c $\geq$ 9.5	N/A	N/A	N/A	N/A	N/A
Comorbidities	0.29	0.35	0.29	0.13	1.23
Inpatient visit	0.25	0.34	0.25	0.21	2.02
		Healthcare Costs			
Total Cost	8,638	10,713	8,506	0.13	1.14
Patients	1,521	91	1,430		