Outcomes-Based Payments and Physician Productivity: Evidence from Diabetes Care in Hawaii *

Benjamin R. Handel[1], Jonathan T. Kolstad[1], Allyson B. Root[1], and Michael D. Whinston[2]

[1]UC Berkeley
[2]MIT

March 21, 2019

Abstract

Designing physician payments to incentivize low cost, high quality care is a core component of most health systems. The steady improvement in medical information technology over the past decade has directly led to (i) an increase in the ability of systems to more precisely link payments to health care and health outcomes of interest and (ii) deliver timely information to physicians about their performance on these care and outcomes metrics. We study physician productivity in response to quality-based incentives empirically using novel proprietary data from HMSA, the largest insurer in the state of Hawaii. Over the time period 2006-2015 HMSA implemented significant program changes in its pay-for-performance program including the money at stake and the types of measures being rewarded. We focus on payment incentives that HMSA provides for diabetic HbA1c control. Initially, from 2006-2010, the insurer provided very limited incentives for HbA1c control. From 2011-2013, HMSA provided substantial financial incentives to screen HbA1c while, for 2014-2015, they changed this program to only reward high quality HbA1c medical outcomes for patients with diabetes. We find that the shift to outcomes-based rewards (i) meaningfully increases prescriptions of more intensive interventions (insulin) for patients above the HbA1c quality threshold (ii) leads to additional visits to specialists (iii) leads to additional diagnosis of complications related to diabetes and (iv) improves HbA1c outcomes. We find that increased insulin prescribing is associated with (i) being part of a physician organization, as opposed to being a solo practitioner and (ii) older physicians. Finally, we show that high IT users have relatively higher baseline rates of insulin prescriptions among high A1c patients (before and after the program change), while low IT users only alter their insulin prescription rates after the introduction of outcomes-based payments.

*[We thank HMSA for use of their data and thank Anita Cabreros and Robert Eubanks at HMSA for their tireless work in supporting this project. We thank Dr. Elizabeth Murphy for her guidance. We thank Zarek Brot-Goldberg, Ishita Chordia, Natalie Cox, Nianyi Hong, Nir Kaldero, Eva Lyubich and Nadav Tadelis for excellent research assistance. We thank NSF CAREER Grant 1552824, Microsoft Research, and NIA P01 grant XXXXX for generous financial support. We thank seminar participants at Columbia University, UCLA, University of Texas and XXXXXXXX for their helpful comments.

1 Introduction

Spending on health care per capita in the United States is higher than in other developed nations, without evidence that that spending achieves better health outcomes. Historically, most health care in the U.S. has been reimbursed under fee-for-service payments, where physicians are paid more to do more. Over the past several decades, policymakers and payers have implemented alternative payment arrangements. These include (i) capitated payments, where providers receive a fixed amount per patient per time period, regardless of treatments performed (ii) bundled payments, where providers receive a fixed amount per patient per diagnosis, regardless of treatments performed (iii) Accountable Care Organizations (ACOs, see, e.g. Frech et al. (2015)), which incentivize providers to collaborate and share savings relative to fixed payments benchmarks and (iv) performance-based payments that compensate providers using performance measures related to key health care processes and/or key health outcomes of interest.[1]

As the data available in the health care industry improve, it has become increasingly possible to implement nuanced performance-based payment programs for providers. To date, there is limited economic evidence on how providers respond to such incentives in environments where (i) there is a lot of money at stake (ii) providers primarily face one incentive program and (iii) providers have the information technology to effectively assess their performance in incentive programs in real time. Further, in large part because of data and technological limitations, most pay-for-performance programs (and most studies of such programs) have focused on payments that reward physicians for care processes (e.g., screening diabetics for HbA1c levels) rather than for actual health outcomes (e.g., actual HbA1c test results). Understanding how providers respond to such incentive programs, and the mechanisms underlying those responses, is crucial for evaluating the benefits payers can hope to achieve with such programs as well as for assessing how to optimally design such programs.

We study how providers respond to pay-for-performance rewards empirically using novel proprietary data from HMSA, the largest insurer in the state of Hawaii. HMSA covers over 700,000 lives annually (including well more than half of the commercial market in HI) and most providers in the state receive the majority of their revenue from HMSA. Over the time period 2006-2015, HMSA implemented significant program changes to its pay-for-performance program including the money at stake and the types of measures being rewarded. By 2012, a medium-sized primary care primary care physician could earn approximately $50,000 per year in payments linked directly to measures assessing care quality. In 2012, HMSA led the implementation of a system-wide IT program that substantially increased primary care physicians' information about their performance on the quality metrics being rewarded and also allowed for granular targeting of patients to close associated care gaps. Prior to 2014, HMSA's quality metrics were process-based, focusing on making sure primary care physicians provided specific types of recommended care to specific types of patients. In 2014, HMSA changed some of these metrics to outcomes-based metrics, specifically rewarding physicians for achieving better health care outcomes for their patients. Thus, by 2014, primary care physicians

[1] These supply-side policies can be complements or substitutes to other policies that seek to efficiently steer health care utilization including (i) demand-side incentives (see, e.g., Brot-Goldberg et al. (2017)) (ii) health technology assessment and (iii) queuing.

at HMSA operated in an environment where they could earn meaningful financial payments from one salient program rewarding both health care processes and health care outcomes.

Within this context, we focus our analysis on incentives that HMSA provided for diabetic HbA1c control. Initially, from 2006-2010, HMSA provided very limited incentives for HbA1c control. From 2011-June 2014 HMSA provided substantial financial incentives to screen HbA1c once yearly. For July 2014-2015, they changed this program to only reward HbA1c screening results falling under a threshold of 9, indicating a minimum level of blood glucose control. We study the implications of this programmatic shift for the 592 primary care providers treating patients with diabetes (N = 27,220) at HMSA. To do this, we use granular data on patients, physicians, payments, health care, and health care outcomes. Specifically, we observe (i) claims data for all patients from 2006-2015 (ii) linkages between patients and the primary care physicians they are attributed to (iii) physician payment data (both pay-for-performance rewards and fee-for-service payments) (iv) physician performance on key quality metrics (v) physician click data for the IT program HMSA implemented and (vi) patient HbA1c test outcomes.

Our analysis uses these data elements, together with time-series variation in HMSA payment program and IT availability, to study several key questions. First, we ask whether the shift to outcomes-based rewards for HbA1c levels impacts provider treatment of diabetics. We find that the shift to outcomes-based payments increases insulin prescriptions by 14% for patients above the quality threshold (HbA1c = 9) after the shift to outcomes-based payments, relative to the period prior to this shift and relative to patients with HbA1c levels below 9. Physicians only get reward payments if patients are below this threshold, so have an incentive to prescribe more intensive medications (insulin) that helps patients achieve this HbA1c target. Almost all of this increase comes from increases in insulin prescriptions from primary care physicians, rather than from specialists. There was no analogous increase in insulin prescription rates for patients with diabetes nationwide (see, e,g, Montvida et al. (2017)), giving us more confidence that the effect we find is due to the shift to outcomes-based rewards. In addition, Handel et al. (2019) investigate diabetes treatment with all-payer claims data from Utah and do not find any year on year increase in insulin prescriptions over our study period. We also run our primary regressions for the two year period prior to the introduction of outcomes-based rewards (2012-2013), as a placebo check, and find no effects on insulin prescribing year on year.

In addition, we find that the shift to outcomes-based rewards leads to meaningful increases in visits to to specialists, in particular visits to cardiologists (34%). We also investigate the impacts of this shift on prescriptions for oral diabetes medications, which are generally less intensive than insulin though some of these medications are quite strong and expensive. Though we find no meaningful impacts of the shift to outcomes-based payments on oral medication prescriptions overall, we do find meaningful impacts for select stronger and more expensive oral diabetes medications. We see a statistically significant 32% decrease in use of the oral medication Januvia, and a 170% increase in the use of the Tradjenta (from a baseline use rate of about 1.1%).

Next, we ask whether the shift to outcomes-based payments, along with the concordant shifts in treatments, impacted patient health. We study 14 relevant health care outcomes, including

9 related to complications associated with diabetes. Interestingly, we find that, on aggregate, diagnosed complications increased by 15% for patients above the HbA1c threshold of 9, after the shift to outcomes-based payments. Additionally, we document percentage increases in kidney disease diagnosis (31%), lower limb cellulitis (14%) and hypoglycemia (2%), though the latter two of these effects are not statistically significant, due to statistical power issues. While at first glance it might seem surprising that the shift to outcomes-based rewards increased complications diagnoses for diabetics, in fact the increased rate of referrals to specialists suggests than diagnoses of complications should increase, ceteris paribus, if endogenous diagnosis is an issue in our environment (as it is in many environments, see, e.g, Song et al. (2010)). For both visits to specialists and health care outcomes we run placebo regressions for the two year period prior to the introduction of outcomes-based rewards and find no systematic effects on our key outcomes of interest.

We also look at health outcomes (HbA1c levels) directly and find in an event study analysis that starting or restarting insulin reduces A1c levels by 0.7 for all patients after 1 year, and by 1.5 for patients with HbA1c levels that are above 9 prior to this insulin use. So, evidence points to insulin prescriptions improving patient HbA1c control, despite the increase in complications diagnoses discussed above.

Next, we ask whether the shift to outcomes-based rewards induced heterogeneous provider responses, and, if so, what the determinants of those heterogeneous responses were. First, we investigate a differences-in-differences specification that tie changes in insulin prescription rates to physician observables. We find that physicians working in physician groups (51% of doctors in our sample) have an 11.3 percentage point larger effect than solo practitioners, implying that physicians in groups are responsible for essentially all of the documented increases in insulin prescribing. This is consistent with the hypotheses that it is easier to disseminate guidelines within an organization and easier to make payment changes salient.[2]

Finally, we investigate how physician responses to the shift to outcomes-based rewards relate to their use of IT. We categorize physicians into terciles of IT use based on the their total pageviews in the IT system to manage quality rewards that was implemented in mid-2012. We find that, for high IT users, insulin prescription rates increased for patients with diabetes above the HbA1c threshold of 9 *before* the shift to outcomes-based rewards while for low IT users insulin prescription rates increased only after the shift to outcomes-based rewards. This suggests that physicians who were high IT users shifted their care after the implementation of IT and didn't need the accompanying financial reward to start targeting the kinds of treatments that could help achieve better HbA1c outcomes. This suggests that IT use is a substitute for outcomes-based rewards patients, though caution is warranted with this interpretation since we only document the associated between IT and insulin prescribing and not a causal effect of IT.

Our study relates to a number of papers that study physician responses to financial incentives. Few papers study physician responses to pay-for-performance incentives in an economic framework. Mullen et al. (2010) study physician responses to process measure-based pay-for-performance and

[2]We also find that male physicians, physicians above the median age, and physicians who are internal medicine specialists have positive, but statistically insignificant treatment effects.

find that, while physicians make small improvements in the measures being rewarded (in response to relatively small rewards) there are no noticeable improvements in health care otherwise, suggesting that physicians target achieving the financial reward but don't use the incentive program to make deeper changes to their practices. Lee et al. (2010) employ a difference-in-differences design to study patients with diabetes for physicians who were voluntarily enrolled in a government run pay-for-performance program in Taiwan. They find that this program, which provided incentives for healthcare providers to enhance self-care education and annual screenings (similar to HMSA P4Q metrics), significantly increased testing/exam rates, physician visits, and reduced hospitalizations, with minimal net cost increases. Li et al. (2014) also uses a DID approach to evaluate a P4P program in Ontario targeted more generally to primary care services. They find that physicians respond to some of the financial incentives, but not to others, with differential response linked to the cost of responding and the strength of the evidence linking a service with quality. They also find that younger physicians and physicians from larger practices were most likely to respond.

Quite a few papers study physician responses to non-quality-based financial incentives. Clemens and Gottlieb (2014) study the impact of an exogenous change in Medicare reimbursements on physicians' treatment decisions. They model (i) physician altruism (ii) financial incentives and (iii) physician practice styles and find that physicians change their care patterns in response to financial incentives. Shafrin (2010) studies physician responses to fee-for-service payments vs. capitated (fixed lump sum) payments and finds no statistically significant effect of one payment mode vs. another on primary care physician treatments, though he does find a change in referrals to specialists under fee-for-service payments, which in turn leads to increased surgeries and increased expenses. Glied and Zivin (2002) study physician behavior when they have some patients who they receive fixed lump sum payments for and others who they get paid fee-for-service for. They find that physicians respond coarsely to the proportion of patients they have under each arrangement: physicians who treat most of their patients under fee-for-service incentives tend to use fee-for-service-style care patterns for *all* of their patients, and vice-versa for physicians who treat most of their patients under fixed lump sum payments. [3]

There is also a related growing literature on physician responses to non-financial organizational factors that impact their performance. Kolstad (2013) studies cardiac surgeons' responses to quality report cards about patient health outcomes and finds that physicians are intrinsically motivated to improve their performance due to the public revelation of information on their quality. Chan (2016) studies physician productivity in the ER as a function of their peer group productivity, finding that peer group productivity has a meaningful positive impact on physician productivity. Chan (2018) finds evidence that ER physicians "slack off" near the end of shifts, suggesting changes to schedules and patient assignment can add value. Abaluck et al. (2016) and Obermeter (2018) study heterogeneity in physician productivity, finding both substantial heterogeneity in physician

[3]Barro and Beaulieu (2003), Devlin and Sarma (2008), Dumont et al. (2008), and Gruber and Owings (1994) also study different questions related to physician responses to financial incentives. Eliason et al. (2018) and Einav et al. (2018) study how medical organizations respond to threshold-based financial incentives in the context of long-term care hospitals, finding meaningful care responses in response to large incentives to either keep patients in the facility or discharge/transfer them.

performance and substantial opportunities to improve physician performance in their respective contexts.

The rest of the paper proceeds as follows. Section 2 describes the HMSA environment and the different pay-for-performance programs introduced at HMSA over our study period. Section 3 presents descriptive statistics related to our study. Section 4 presents our primary analysis of the impact of the switch to health outcomes-based rewards on health care and health outcomes. Section 5 concludes.

2 Environment

Our study uses detailed micro-level data from the Hawaii Medical Service Association (HMSA) the largest health insurer in the state of Hawaii. HMSA, a not-for-profit insurer that is an independent licensee of the Blue Cross and Blue Shield Association, HMSA insures the majority of patients in the state of Hawaii and, as a result, was able to implement a large-scale financial incentive program to primary care physicians that offered meaningful money (more than other private insurers have typically been able to offer in the U.S.) to achieve certain quality targets.

Figure 1 shows the breakdown of HMSA enrollees in 2012, at the beginning of our study period. HMSA covered just over 700,000 lives, compared to a total population of approximately 1.35 million in the state of Hawaii overall. Most of the patients ($\approx$ 525,000) are covered under an HMSA-run employer provided plan, with a sizable portion ($\approx$ 131,000) covered by Hawaii QUEST, a managed care program for Medicaid beneficiaries. The remainder are covered under a Medicare Advantage plan, or some other special program.

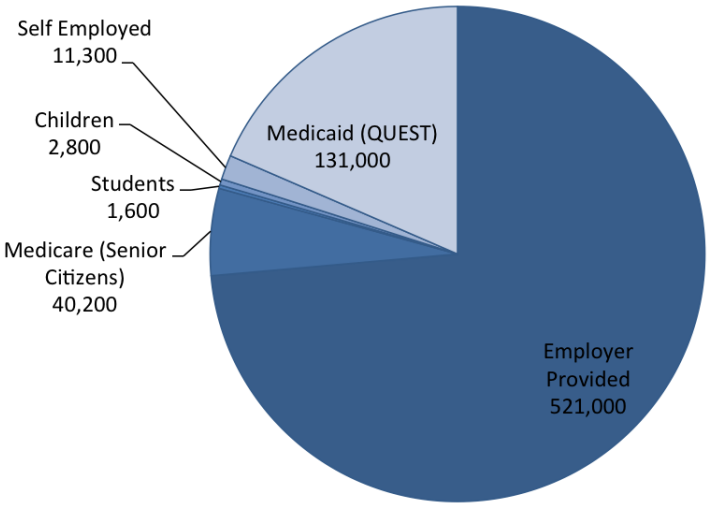

Figure 1: Composition of the HMSA Patient Population

In terms of population health, when compared to the nation as a whole Hawaii tends to perform

at or slightly below the national average on metrics related to chronic illnesses and preventive care take-up. As of 2012, the Center for Disease Control (CDC) reported that 56.1% of Hawaii residents were classified as being overweight or obese (slightly below the national average), while 8.3% of respondents in a 2010 CDC survey reported having been diagnosed with diabetes. This compares to 69.2% and 8.7% for these respective metrics in the national overall suggesting that, as with many health metrics, Hawaiian population health is similar to population health nationwide.

Despite some evident similarities there are certain aspects of Hawaiian healthcare that do differentiate it from other states, and make it an interesting case study in how chronic illness and preventable disease can be better monitored, managed, and controlled. Traditionally, the management of chronic illness amongst Hawaiian residents has been complicated by the state's unique geography. The seven islands are separated by considerable aquatic distance and vary in population density, creating a number of rural, isolated communities. This dispersed population, along with a relative shortage of primary care physicians, leads to substantial population heterogeneity in terms of access to care for chronic illness management and preventive care otherwise. Segments of the population living in under-serviced or particularly remote locations tend to have infrequent doctor visits and large "gaps in care."

2.1 HMSA and Quality Improvement

The programs we study are a subset of a number of initiatives undertaken by HMSA. Programs aimed at primary care physicians (PCPs) and physician organizations — the main focus of our paper —- are a supplement, rather than replacement, to existing traditional fee-for-service programs. HMSA's first quality improvement program began in 1998 and has since expanded / changed in a manner that both refined the metrics physicians are judged on and increased the money they could earn through the program. In addition to the programs that have targeted primary care physicians, there are now programs targeted at hospitals, physician organizations (POs), and specialists. Over time, there have been yearly revisions (both minor and major) to the quality indicators, reward incentives, and quality goals that underlie these programs.

The primary focus of our research is the Pay-For-Quality (P4Q) program, which in 2011 replaced another, smaller, incentive program (PQSR) as the central quality improvement program for individual primary care providers at HMSA. P4Q attempted to both simplify the requirements for physicians and increase the money at stake, relative to the prior program. All eligible physicians were automatically enrolled in the program, with eligibility gradually expanding to include physicians serving all HMO, PPO, Medicare, and Medicaid patients. As of December 2012, 506,097 commercial patients and 988 primary care providers were part of the commercial P4Q program (essentially all commercial patients and primary care providers associated with HMSA).

The automated nature of the P4Q program meant the payment of a P4Q bonus award requires little active participation on the part of the physician. The bonus amount is automatically calculated and dispensed by HMSA using claims data, although physicians can actively petition to change what claims are used in the calculation of their award (in practice, not many did so).

Almost all PCPs were eligible by 2012 and all those eligible are automatically enrolled and receive payments based on their performance. This suggests that, in cases where physicians are unaware of the program to begin with, they should become somewhat aware at a minimum after receiving payment with a detailed description of their performances.

2.2 Performance Measures: From Process to Outcome

The P4Q program bases its assessment of physician performance on a set of widely accepted "quality measures." These measures focus on the care of chronic illness (namely diabetes and heart diseases) and preventative care (for example, cancer screening or immunizations). Using claims data, HMSA determines how many of a physician's patients were "eligible" to receive a measure in the previous quarter, and then calculates the rate of how many patients actually received recommended care (or achieved recommended health outcomes). There are three components to this: first HMSA determines which patients should be attributed to which physicians, second HMSA determines if a patient meets specific criteria that make them eligible for a given metric, and third HMSA determines if the patient received recommended care (or achieved the recommended health outcome). The main "quality metric" or "success rate" for each measure is defined as:

$$\frac{Number\ of\ Eligible\ Patients\ Satisfying\ Quality\ Metric}{Number\ of\ EligiblePatients\ for\ Quality\ Metric} \tag{1}$$

The physician's rate for each measure is then compared to the national performance rate on this measure (provided by National Committee for Quality Assurance (NCQA)) and to physician's own rate in the previous period. Thus, performance quality for a given metric depend both on a physician's absolute performance (relative to national standards) and their own recent improvement in performance. HMSA then translates performance across the *set* of rewarded metrics into a monetary award. The key features of the financial reward structure are:

- Physicians receive a total dollar amount per patient per month that determines the maximum money they can earn. In 2012, a physician earned $4 per patient per month, such that a physician with 1,000 eligible patients could earn 1,000 X $2 X 12 months = $48,000 in bonus payments for the year.

- HMSA assesses performance on each measure and determines whether physicians have passed specific national performance thresholds (10th, 25th, 50th, 75th, 90th) for that measure. Payment related to a given measure jumps at each threshold but does not increase in between thresholds.

- Once the maximum award and performance level for each measure are determined, HMSA determines weights for each of 14 reward measures (in 2012) based on (i) the proportion of a physician's patients eligible for that measure and (ii) the ex ante importance HMSA assigns to a measure. If most of a physician's patients are eligible for one measure but not for others, the physician will be paid primarily on their performance for that measure. HMSA combines these program elements to determine a final bonus payment for each physician.

Since this algorithm to determine the P4Q monetary award is fairly complex, we describe the entire program in more detail in Appendix B. Table 1 below shows the number of physician and patient participants in P4Q as well as the distribution of actual and maximum possible awards (for all measures and diabetes related measures only) for each year in our sample.

P4Q Eligibility and Award Distribution per Quarter (Q3)

Year	Measures	Physician N	Patient N	Payments	Mean	SD	Min	P25	Median	P75	Max
2011	All	681	176524	Actual	$1,811	$2,243	$0	$233	$930	$2,631	$15,087
				Max Possible	$3,967	$3,270	$6	$1,184	$3,383	$6,139	$16,318
	DM Only	544	28474	Actual	$734	$887	$0	$90	$399	$1,048	$5,275
				Max Possible	$1,798	$1,482	$7	$536	$1,498	$2,696	$6,971
2012	All	826	198809	Actual	$3,807	$4,937	$0	$227	$1,497	$5,917	$27,479
				Max Possible	$6,624	$6,647	$4	$848	$4,458	$10,969	$32,604
	DM Only	623	29322	Actual	$1,401	$1,934	$0	$75	$462	$2,191	$9,709
				Max Possible	$2,611	$2,588	$9	$368	$1,787	$4,231	$12,509
2013	All	866	178277	Actual	$4,485	$5,505	$0	$295	$1,970	$7,099	$31,576
				Max Possible	$7,159	$6,759	$12	$1,090	$5,284	$11,670	$36,127
	DM Only	592	27220	Actual	$1,779	$2,255	$0	$122	$678	$2,830	$10,316
				Max Possible	$2,943	$2,637	$9	$609	$2,229	$4,685	$12,221
2014	All	926	328952	Actual	$4,191	$5,278	$0	$226	$1,620	$6,941	$35,394
				Max Possible	$7,073	$6,750	$14	$1,212	$5,100	$11,675	$37,514
	DM Only	661	35593	Actual	$739	$973	$0	$42	$269	$1,140	$5,415
				Max Possible	$1,558	$1,399	$5	$375	$1,180	$2,473	$7,677
2015	All	926	329222	Actual	$4,125	$5,206	$0	$224	$1,510	$6,946	$28,099
				Max Possible	$7,565	$7,222	$14	$1,381	$5,176	$12,587	$37,409
	DM Only	701	35058	Actual	$816	$1,077	$0	$55	$284	$1,354	$5,839
				Max Possible	$1,682	$1,597	$8	$368	$1,161	$2,644	$10,443

Table 1: This table describes P4Q Eligibility and Award Distribution from 2011-15.

After the inception of P4Q in 2011 HMSA made several meaningful changes to the program, including to the magnitude and calculation of the awards and to the quality measures incentivized. Figure 2 outlines the major program changes. Of particular importance for our study, HMSA changed some important measures from process-based to outcomes-based in mid-2014. Prior to this change, HMSAs quality measures were based strictly on services performed: screenings, lab tests and scans for which a primary care physician would bill themselves or refer the patient elsewhere. In mid-2014, some of these quality measures were changed to reflect downstream outcomes rather than just the provision of a service. For example, our study focuses on a change from rewarding *testing* patients for A1c levels to rewarding *actual A1c outcomes*.

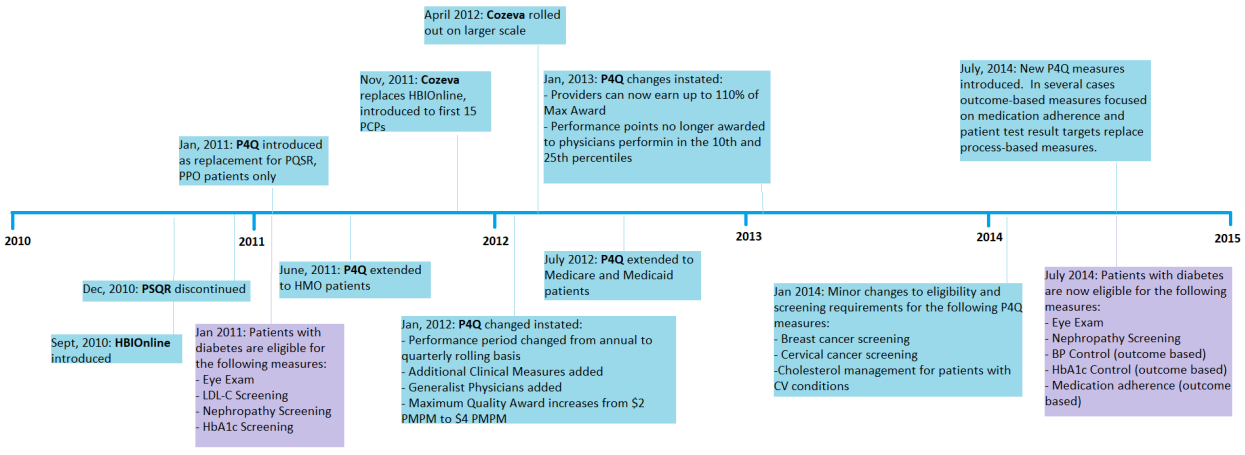

Figure 2: Timeline of P4Q Program Changes

One other key environment change is the HMSA-led introduction of Cozeva, an in-office IT platform designed specifically to help physicians manage their performance in the HMSA P4Q program. This platform provided key summary metrics related to performance as well as the ability to assess their most valuable improvement opportunities, alongside functionalities to help them achieve those improvements on a patient-by-patient basis. We study the introduction and use of this IT platform in more depth in other ongoing work. For the purposes of this study, we discuss performance in the program as a function of IT use. We describe this platform in more depth in Appendix C.

The analysis in this paper focuses on the P4Q quality measures related to patients with diabetes. Measures specifically targeting patients with diabetes represent 4/10 of the adult patient quality measures in 2011, and some of these measures are, ex ante, given double or four times the weight of other measures in the ultimate payment calculation. Figure 3 describes the four process-based measures related to diabetes care prior to mid-2014 including (i) testing for patient A1c (ii) LDL-C cholesterol screenings (iii) testing for nephropathy and (iv) eye exams.

The mid-2014 shift from process to outcomes-based rewards described above was particularly evident for diabetic patients. Most significantly, the program changed such that, instead of rewarding physicians based on the proportion of diabetic patients receiving A1c tests each year, physicians were rewarded based on the proportion of these patients receiving a test below the 9.0 percent A1c threshold.[4] This constituted a marked shift where physicians were now no longer rewarded in the P4Q program for providing a specific form of health care, but were instead rewarded to achieve better health outcomes, by whatever means they chose. After mid-2014, for diabetic patients, in addition to being rewarded for patient A1c outcomes, physicians were rewarded for (i) blood pressure control for diabetics (ii) medication adherence for oral diabetes medications and (iv) the same eye exam and nephropathy screening measures as they were before the program change. Figure 4 describes the set of rewarded measures after mid-2014 in more detail.

[4]After the switch to outcomes-based rewards, if patients were not tested within the past year, they did not count as having successfully fulfilled the HbA1c value requirement, regardless of their scores on prior tests.

MEASURE	HIGH-LEVEL DEFINITION
Eye exam	Percentage of patients with diabetes 18–75 years of age who received a retinal or dilated eye exam by an eye care professional (optometrist or ophthalmologist) in the current measurement period, or a negative retinal exam (no evidence of retinopathy) by an eye care professional in the prior measurement period. (American Diabetes Association guideline)
LDL-C screening	Percentage of patients with diabetes 18–75 years of age receiving at least one lipid profile (or ALL component tests) during the current measurement period. (American Diabetes Association guideline)
HbA1c testing	Percentage of patients with diabetes 18–75 years of age receiving one or more HbA1c test per measurement period. (American Diabetes Association guideline)
Nephropathy	Percentage of patients with diabetes 18–75 years of age who had at least one test for microalbumin during the current measurement period or who had evidence of medical attention for existing nephropathy (diagnosis of nephropathy or documentation of microalbuminuria or albuminuria; ACE inhibitor/ARB therapy during the measurement period is also acceptable evidence). (American Diabetes Association guideline)

Figure 3: Quality Measures for Treatment of Patients with Diabetes: Pre-July 2014

MEASURE	HIGH-LEVEL DEFINITION
Blood pressure control <140/90	The percentage of patients with diabetes 18–75 years of age whose most recent blood pressure reading during the measurement period was <140/90.
Eye exam	The percentage of patients with diabetes 18–75 years of age who received a retinal or dilated eye exam by an eye care professional (optometrist or ophthalmologist) in the current measurement period, or a negative retinal exam (no evidence of retinopathy) by an eye care professional in the prior measurement period. (American Diabetes Association guideline)
HbA1c poor control	The percentage of patients with diabetes age 18–75 years whose most recent HbA1c test during the measurement period is >9.0 percent or whose HbA1c wasn't measured. (Note: A lower score indicates better performance.)
Medication adherence for oral diabetes medications	The percentage of patients 18 years of age or older who adhered to their prescribed drug therapy across four classes of oral diabetes medications – biguanides, sulfonylureas, thiazolidinediones, and dipeptidyl peptidase-IV (DPP-IV) inhibitors – by meeting the proportion of days covered threshold or 80 percent during the measurement period.
Medical attention for nephropathy	The percentage of patients with diabetes 18–75 years of age who had at least one test for microalbumin during the current measurement period or who had evidence of medical attention for existing nephropathy (diagnosis of nephropathy or documentation of microalbuminuria or albuminuria; ACE inhibitor/ARB therapy during the measurement period is also acceptable evidence). (American Diabetes Association guideline)

Figure 4: Quality Measures for Treatment of Patients with Diabetes: Post-July 2014

2.3 National Trends: Treatment of Patients with Diabetes

Our study focuses on treatment of patients with diabetes at HMSA, given the environment and P4Q reward program just described. There are a number of approaches to glycemic treatment for patients with diabetes. Most patients with Type I diabetes should be treated with both long and

rapid acting insulin injections. For patients with Type II diabetes, metformin, an oral diabetes drug, is the preferred initial pharmacological agent. According to the 2015 American Diabetes Association Standards of Care Guide, patients who are unable to achieve or maintain the target A1c over 3 months with metformin only should be prescribed a second oral diabetes drug or a rapid acting insulin regimen (ADA (2015)). Due to the progressive nature of type II diabetes, many patients should eventually be prescribed insulin. Additionally, patients with newly diagnosed type II diabetes who have especially pronounced symptoms or elevated A1c may be immediately prescribed insulin therapy.

Higgins et al. (2016) analyzes national trends in type II diabetes treatments by PCPs and specialists from 2000 to 2015 and finds that though the average number of antidiabetic drugs prescribed per patients has remained relatively stable since 2006, there have been several shifts in regimens of oral medication commonly prescribed. Prescription of DPP-4 inhibitors, SGLT-2 inhibitors, and GLP-1 RAs became more common over this period at the expense of metformin and sulfonylureas, a trend holding true throughout our study period (2012-15). See Figure 5 below, which outlines the marketing periods for new oral diabetes and insulin treatments in our data. The proportion of specialist patients on insulin-based treatment regimens nearly doubled between 2000 and 2006 (from 15 to 30%) but has since remained stable over time. Montvida et al. (2017) also found evidence for national stability in insulin prescription rates over our study period. There has been a trend toward reduced time between diagnosis and initiation of insulin therapy, though this is stronger among specialists. Glucose control has also been relatively stable at the national level since 2008. Taken together, these national trends suggest that there is no broad reason we should expect insulin use to increase in the HMSA population over our 2012-2015 study period, apart from the specific changes made in the HMSA environment itself.

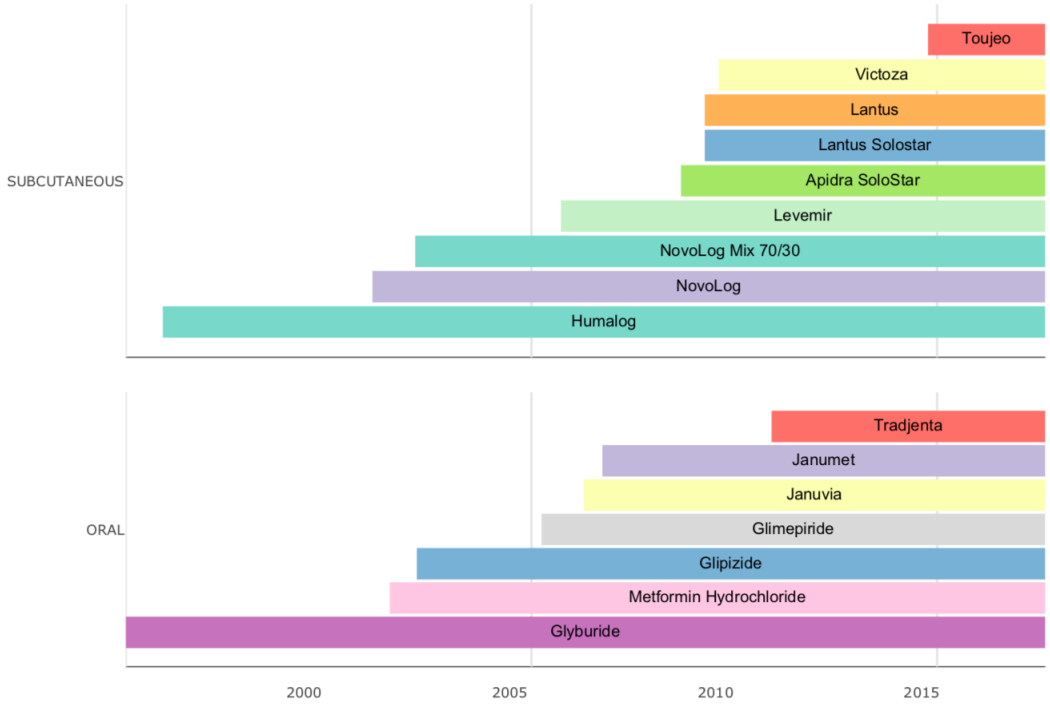

Figure 5: This graph describes the marketing periods for some of the most common drugs in our data.

3 Descriptive Statistics

Table 2 describes our primary patient and physician samples. The first column (moving left to right) shows the entire sample of diabetic patients (N = 27,220) in the third quarter of 2013, the first quarter of our "baselin" analysis period, before outcome-based rewards were in place. The second column describes the sample once we limit the analysis to patients who are present in the data each quarter between July 2013-June 2015 (N = 19,228). The third column further limits the sample to patients for whom we have an A1c test value in the first quarter of each reference year (N = 5,656).[5] The final column describes our primary analysis sample (N = 3,248), which further limits the sample to patients who have at least one prescription drug claim per year in 2013-2015 (some patients have HMSA health insurance but not HMSA drug insurance).

Most of the population characteristics we present in the table are similar across these different samples. Our primary analysis sample is 55% male, has a mean age of 57.7, and 13% rural. Their mean recorded A1c level is 7.36 (SD = 1.45) and approximately 13% have A1c levels above 9, the threshold for the outcomes-based rewards we study. Patients in the sample see, on aggregate, 367 distinct primary physicians who are incentivized to improve their A1c outcomes. Patients spend, on average, $151 per quarter with their primary care physician, $2058 per quarter total, and $480

[5]To fulfill this condition, patients must have had at least one HbA1c test in each July-September of 2013 and in July-September of 2014. We must also have that test value in our data. While around 92% of the patients in the full balanced panel had an A1c test within the last year according to claims data, we only have recorded A1c values for around 71% of those patients.

per quarter out of pocket.

Sample Restrictions and Descriptive Statistics Q3 2013				
	Full Sample	*Balanced Panel Q3 2013-Q2 2015*	*Has A1c Value in Q3 or Q4 in Data (Balanced)*	*Recent A1c and RX Value in Data (Balanced)*
Total Patients	27,220	19,228	5,656	3,248
Patients w A1c $\leq$ 9	3,077	2,196	702	422
Physicians	592	497	392	367
Patients/Physician	46	38.7	14.4	8.9
A1c	7.37	7.39	7.33	7.36
	(1.56)	(1.52)	(1.42)	(1.45)
Recent A1c Test (12m) in Claims	0.91	0.92	0.98	1.00
Male	0.53	0.54	0.55	0.55
Age	58.9	58.8	58.5	57.7
	(10)	(9.4)	(8.7)	(8.6)
Rural	0.16	0.16	0.17	0.13
Total Spending at PCP in Quarter	109	116	145	151
	(144)	(142)	(146)	(137)
Total Spending in Quarter	2461	2230	1964	2058
	(11825)	(10726)	(6828)	(7003)
Total OOP Spending in Quarter	1021	912	499	480
	(7283)	(6947)	(2641)	(2746)

Table 2: This table describes sample restrictions and basic summary statistics.

Table 3 presents a breakdown of the distribution of A1c scores over time for our primary analysis sample, including the within-individual and across-individual variance in these scores. For the period 2014-2015, after the introduction of outcomes-based rewards, mean patient A1c for their first test in a year is 7.57, the standard deviation of this first test is 1.51, indicating the level of across-individual variation in this score. The standard deviation within a patient's scores during a year is 0.43, indicating that there is some meaningful room for change in the A1c score within a calendar year. Figure 16 in the Appendix presents the entire distribution of A1c test scores in the primary analysis sample, across the different years we study.

A1c Test Distribution			
	First A1c Test	*Within-Ind. μ / σ*	*Total A1c Test*
July 2013 - June 2014	7.387	7.434	3.27
	(1.507)	(0.443)	
July 2014 - June 2015	7.567	7.526	3.72
	(1.508)	(0.431)	

Table 3: This table describes the breakdown of the distribution of A1c scores. The first column gives the mean and standard deviation in the population for patients' first A1c test in each year. The second column gives the population mean of the individual-level test mean over the year, and the population mean of the individual-level standard deviation of tests within the year.

3.1 Drug Cost and Utilization

In this section we report summary data on total costs, patient costs, and patient purchases for insulin treatments and for oral diabetes medications. This material gives a sense of baseline cost and use for these key treatments in our sample. It also highlights that there were no major changes to patient drug costs during our sample period, removing this as a potential factor underlying the utilization changes we document corresponding to the P4Q program shift to outcomes-based payments.

3.1.1 Insulin

Table 4 breaks down patient purchases, total cost, and patient cost for the different common insulin treatments purchased during our study period. Most insulin treatments have unit cost in the $20-30 range and patient cost-sharing percentages between 3-5%. Both total drug costs and patient drug costs are quite stable over the two years 2014 and 2015 presented in the table. One outlier is an insulin glargine treatment Toujeo, which is new to HMSA patients in 2015. That treatment averages $69 per unit and has average patient cost-sharing of 11.4% ($102 per filled prescription). This treatment is used relatively infrequently, as is Humulin (human insulin) which has low patient cost-sharing but high unit costs ($56 in 2015).

Pcode Ref	Proprietary Name	Non Proprietary Name	Year	Avg Unit Cost	Avg Total Cost	Avg Pat Share	Avg Pat Cost	Count	Unique Pats	Avg Fills Per Pat
D	Lantus Solostar	Insulin Glargine	2014	21	837	4.2	27	9,701	3,571	1.7
D	Lantus Solostar	Insulin Glargine	2015	24	1,032	4.7	38	8,933	3,996	1.9
I	Novolog	Insulin Aspart	2014	22	829	3.8	22	3,485	1,432	1.5
I	Novolog	Insulin Aspart	2015	26	1,078	4.4	33	3,403	1,654	1.8
A	Humalog	Insulin Lispro	2014	22	965	3.4	28	2,767	1,029	1.6
A	Humalog	Insulin Lispro	2015	25	1,206	3.7	34	2,205	1,039	1.8
H	Levemir	Insulin Detemir	2014	23	873	3.8	19	798	681	1.4
H	Levemir	Insulin Detemir	2015	24	1,091	4.6	33	2,719	1,238	1.7
E	Lantus	Insulin Glargine	2014	20	849	4.3	33	1,393	582	2
E	Lantus	Insulin Glargine	2015	24	1,167	4.5	41	1,208	577	2.3
G	Novolog Mix 70/30	Insulin Aspart	2014	23	1,290	3.2	29	1,255	459	1.9
G	Novolog Mix 70/30	Insulin Aspart	2015	26	1,697	3.9	45	1,177	541	2.2
F	Apidra Solostar	Insulin Glulisine	2014	22	802	10.3	64	407	169	1.5
F	Apidra Solostar	Insulin Glulisine	2015	26	1,132	9.5	77	327	165	1.8
C	Toujeo	Insulin Glargine	2015	69	1,319	11.4	102	185	150	1.6
B	Humulin	Insulin Human	2014	48	2,679	0.9	20	92	30	2
B	Humulin	Insulin Human	2015	56	2,877	1.4	30	68	29	1.8

Table 4: This table describes the prescribing behavior for Insulin drugs.

Figure 6 presents boxplots of total drug costs and patient drug costs, to shed more light on variance in these quantities across prescriptions filled at HMSA during our study period. The left part of the figure, presenting the distribution of total drug costs per fill by drug, illustrates that total drug costs are concentrated in a narrow band, indicating that HMSA pays similar total costs for a given drug across its insurance products and across our sample period. The same is true for patient cost-sharing, presented in the right part of this figure, though patient cost-sharing has a long tail of higher cost-sharing percentages (relative to low mean cost-sharing), indicative of

patients in high-deductible health plans or patients with copayments that are high relative to total drug costs for a given drug.

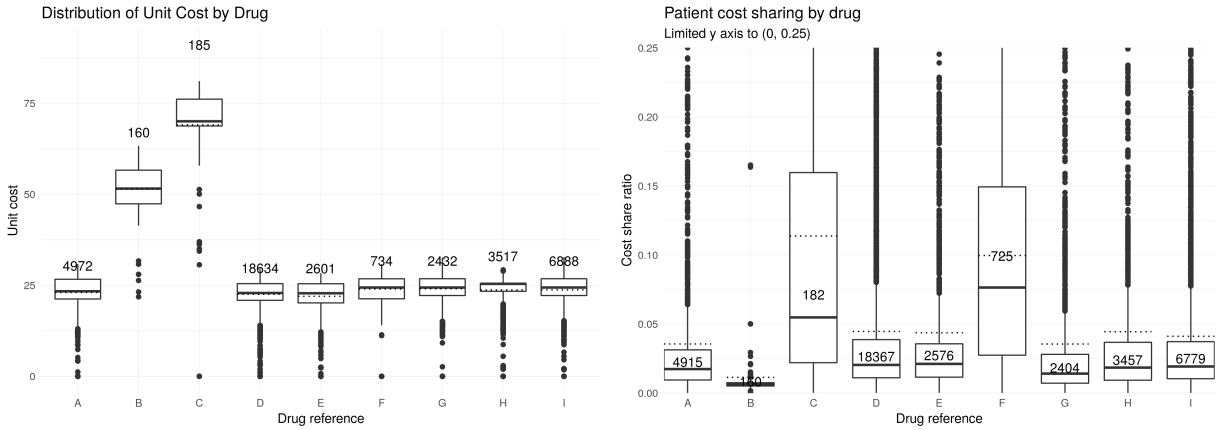

Figure 6: This figure presents the distribution of insulin costs and cost sharing.

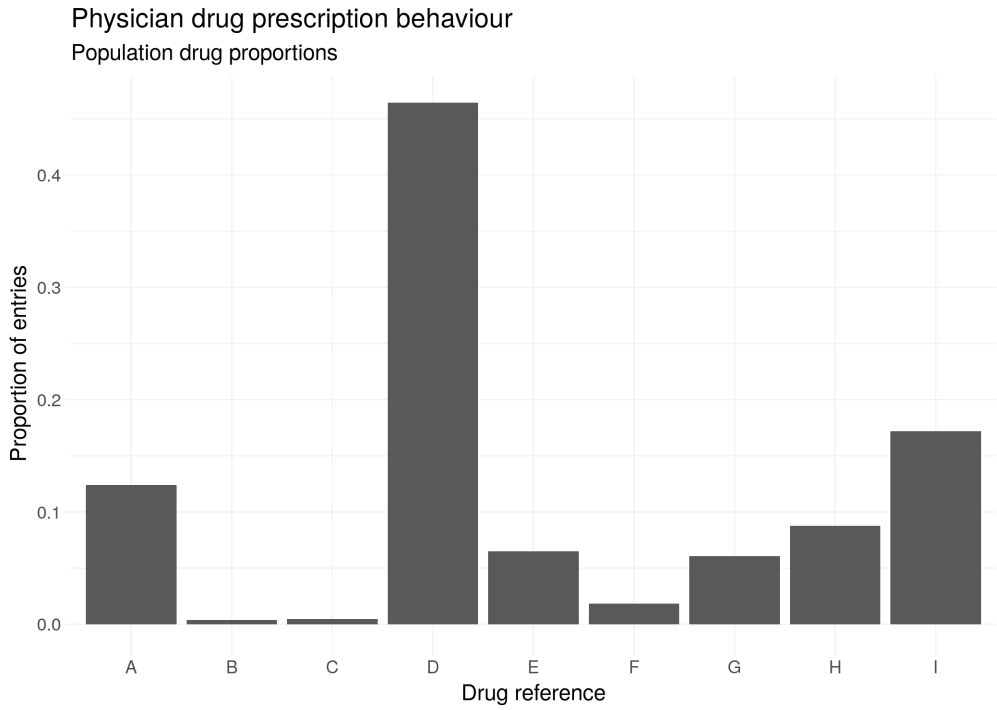

Figure 7: This figure presents the distribution of prescribed insulin drugs.

3.2 Oral Diabetes Medications

Table 5 presents average total costs, average patient costs, and prescription fills for the most commonly prescribed oral diabetes medications in our data. The distribution of total costs for these drugs is bimodal: most have unit costs (e.g. cost per pill) below $1 and total prescription costs of less than $30. However, there are several drugs with high unit costs: (i) Januvia ($10 / unit, $1,011

per prescription in 2015) (ii) Victoza ($67 / unit, $1,495 per prescription in 2015) (iii) Janumet ($6 / unit, $912 per prescription in 2015) and (v) Tradjenta ($10.1 per unit, $1,042 per prescription in 2015). These more expensive oral medications began to be marketed in 2007, 2010, 2007, and 2011 respectively and are on patent until 2022, 2023, 2022/2026[6], and 2023/2030[7] Importantly, some of these higher cost medications also have high utilization levels. Januvia had 8,547 prescription fills in 2015, Victoza had 3,554, Janumet had 5,409, and Tradjenta had 1,512. These medications are thus commonly used for patients with diabetes and also quite costly for insurers. Patient cost-sharing is also meaningful higher, as an absolute amount, for these medications, ranging from $65.8 per fill (Victoza) to $108.9 per fill (Tradjenta).

Pcode Ref	Proprietary Name	Non Proprietary Name	Year	Avg Unit Cost	Avg Total Cost	Avg Pat Share	Avg Pat Cost	Count	Unique Pats	Avg Fills Per Pat
C	Metformin Hydrochloride	Metformin Hydrochloride	2014	0	28	86.7	9	50,204	18,353	1.7
C	Metformin Hydrochloride	Metformin Hydrochloride	2015	0	32	77.7	8	46,224	19,426	1.6
I	Glipizide	Glipizide	2014	0	22	69.7	11	11,271	4,336	1.8
I	Glipizide	Glipizide	2015	0	25	60	11	11,152	4,846	1.8
B	Januvia	Sitagliptin	2014	9	808	8.3	66	8,772	3,304	1.9
B	Januvia	Sitagliptin	2015	10	1,011	7.7	73	8,547	3,589	2
A	Janumet	Sitagliptin & Metformin Hydrochloride	2014	5	717	9.3	61	6,108	2,212	1.8
A	Janumet	Sitagliptin & Metformin Hydrochloride	2015	6	912	8.5	69	5,409	2,334	2
D	Glimepiride	Glimepiride	2014	0	10	88.5	9	5,738	2,178	1.9
D	Glimepiride	Glimepiride	2015	0	16	66.7	10	5,174	2,257	1.9
F	Victoza	Liraglutide (Rdna Origin) Injection	2014	61	1,229	5.4	60	3,832	1,372	1.8
F	Victoza	Liraglutide (Rdna Origin) Injection	2015	67	1,495	5	66	3,554	1,570	1.9
K	Pioglitazone Hydrochloride	Pioglitazone Hydrochloride	2014	2	162	18.5	14	2,998	1,139	1.8
K	Pioglitazone Hydrochloride	Pioglitazone Hydrochloride	2015	1	139	17.6	12	2,9	1,261	1.7
H	Kombiglyze	Saxagliptin & Metformin Hydrochloride	2014	5	514	10.5	25	10	10	1.3
H	Kombiglyze	Saxagliptin & Metformin Hydrochloride	2015	7	820	7.9	59	1,847	812	1.8
E	Glyburide	Glyburide	2014	0	41	42.6	13	1,835	727	1.6
E	Glyburide	Glyburide	2015	0	43	37.1	12	1,375	634	1.6
J	Tradjenta	Linagliptin	2014	9	818	12	89	1,587	690	2.2
J	Tradjenta	Linagliptin	2015	10	1,042	11.8	109	1,512	737	2.6
G	Onglyza	Saxagliptin	2015	10	838	10	64	1,192	555	1.8
N	Bydureon	Exenatide	2014	105	1,120	6.1	69	838	312	2.2
N	Bydureon	Exenatide	2015	112	1,285	6.7	82	220	172	2.3
M	Glyburide & Metformin Hydrochloride	Glyburide & Metformin Hydrochloride	2014	0	31	49.6	13	790	301	1.8
M	Glyburide & Metformin Hydrochloride	Glyburide & Metformin Hydrochloride	2015	0	36	38.4	12	479	231	1.7
L	Nateglinide	Nateglinide	2014	1	223	6.8	13	559	214	1.9
L	Nateglinide	Nateglinide	2015	1	260	5.2	11	517	246	2

Table 5: This table describes the prescribing behavior for oral diabetes drugs.

[6]The basic components of Janumet expire in 2022, but the salt of the monophosphate version which is marketed expires in 2026.

[7]Different components of Tradjenta have varying patent end periods, some expired as early as 2017, some will only expire in 2030.

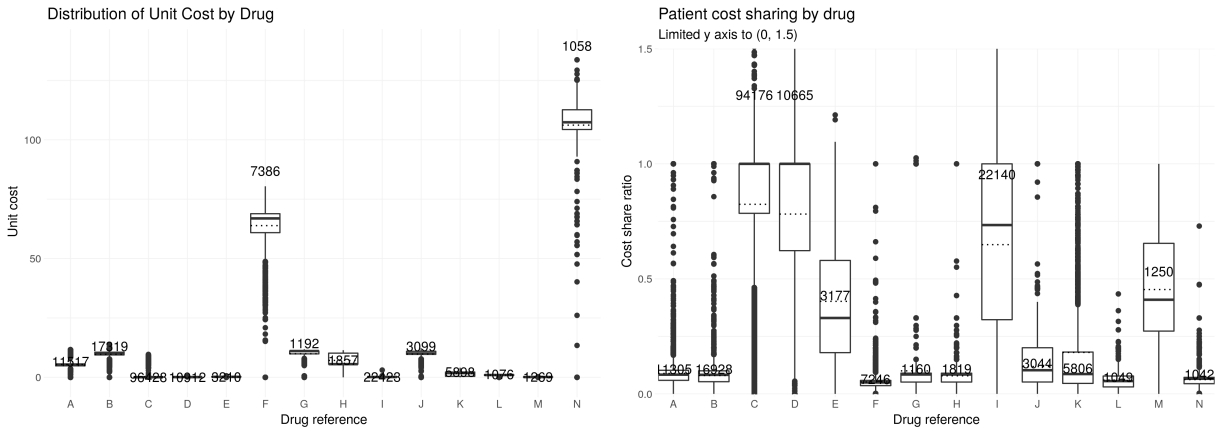

Figure 8: This figure presents the distribution of oral diabetes drug costs and cost sharing.

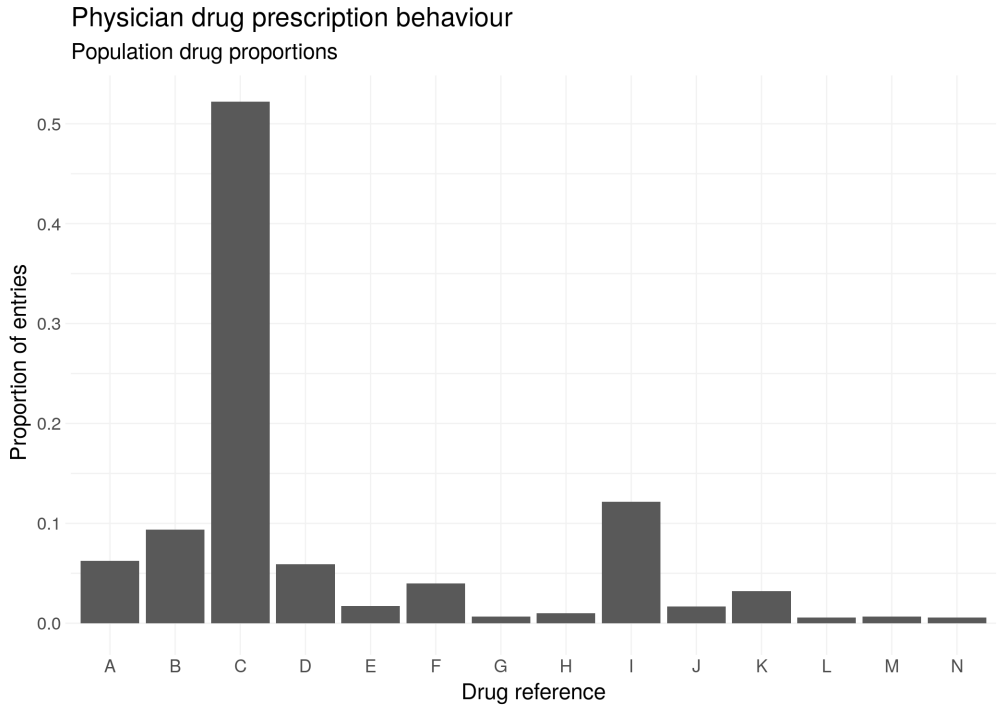

Figure 9: This figure presents the distribution of prescribed oral drugs.

4 Impact of Outcomes-Based Incentives

We now turn to our analysis assessing the impact of the shift to outcomes-based incentives in July 2014. We study the impacts of this shift on spending, health care, and health outcomes. We also investigate physician response heterogeneity to better understand the determinants of the changes that we document on each of these dimensions.

4.1 Impact on Health Care

We begin by investigating the impact of the switch to outcomes-based payments on diabetic-related treatments. Table 6 shows the change over time in insulin prescription rates, as a function of patient A1c level. The left column compares the year before outcomes-based payments went into effect (July 2013-June 2014) to the year when they were first in effect (July 2014-June 2015). There are meaningful increases from the pre-period to the post-period in insulin prescription rates for those above the A1c outcomes-based reward cutoff of 9. From the period beginning in 2013 to that beginning in 2014, for patients with HbA1c level 11-12.9, there is a 15.4 percentage point increase in insulin prescription rates from 2013 to 2014, equivalent to a 66% increase in insulin usage for these patients (significant at $\alpha = 0.05$). There is also a 14.5% increase for patients in the 9-10.9 range (note that very few patients fall into the 13-14.9 range, which accounts for the large standard errors for that bin). As a placebo check, the right column compares prescription rates during the year before outcomes-based payments went into effect (July 2013-June 2014) and the preceding year (July 2012-June 2013). There is minimal change in insulin prescription rate moving from across these two years for those with A1c above 9. This illustrates that there is no increasing trend in insulin rates for high A1c patients prior to the onset of outcomes-based physician rewards.

Insulin PCP Prescription Rates			
A1c	*Time*	*Marginal Effect* *Pre: July '13 - June '14* *Post: July '14 - June '15*	 *Pre: July '12 - June '13* *Post: July '13 - June '14*
5 - 6.9	*Pre*	0.0330 (0)	0.0331 (0)
	Post	0.0338 (0)	0.0347 (0)
7 - 8.9	*Pre*	0.1397 (0.0001)	0.1092 (0.0001)
	Post	0.1331 (0.0001)	0.1370 (0.0001)
9 - 10.9	*Pre*	0.2461 (0.0005)	0.2662 (0.0005)
	Post	0.2819 (0.0004)	0.2514 (0.0006)
11 - 12.9	*Pre*	0.2323 (0.0015)	0.3019 (0.0018)
	Post	0.3867 (0.0019)	0.2632 (0.0015)
13 - 14.9	*Pre*	0.3738 (0.0083)	0.2384 (0.0040)
	Post	0.3062 (0.0081)	0.1878 (0.0089)

Table 6: These tables and figures shows the change over time in insulin prescription rates, as a function of patient HbA1c level.

Based on this pre-post design, we estimate a difference-in-differences regression model to com-

pare the change in prescription rates from before to after the implementation of outcomes-based payments for patients who are below and above the cutoff A1c test value of 9. This controls for a range of physician-specific and patient-specific factors to document the change in health care received for patients above the A1c threshold relative to those below this threshold. We also perform this regression analysis for the placebo period in the two years prior to the start of outcomes-based rewards. We estimate the following equation using OLS for continuous outcomes and probit for binary outcomes, clustering standard errors at the patient level:

$$y_{qi} = \beta_1 + \beta_2 I(A1c \geq 9)_{qi} * Post_q + \beta_3 Post_q + \beta_4 I(A1c \geq 9)_{qi} + Z_i + Z_q + \epsilon_{qi} \quad (2)$$

Here, q denotes a physician and i a patient. Y_{qi} are health care variables, described in more depth momentarily, that reflect the care received for a given patient i at a given physician q. β_2 is our primary coefficient of interest, reflecting the change in the outcome variable in the post-period for patients with A1c levels above 9, relative to the same change over time for patients with A1c levels below 9. A1c value is measured from the first test of the first 3 months of the reference year (July-September of 2013 for the pre-period and 2014 for the post period), rather than the separately by quarter, to limit concerns of changing test results which are endogenous to the policy change. Outcomes are measured in each of the four quarters of the reference year.[8] Z_i and Z_q are, respectively, patient controls and physician controls. In our primary specification we include quarter of year fixed effects, as well as patient controls including insurance plan type, age, sex, distance from primary care physician office, rural zipcode, island, and family size. The physician controls we include are physician specialty, age and sex.

We report the results in Table 7. The table reports the coefficients β_2, transformed to probabilities. The rows describe the different Y_i variables we study related to diabetic drug utilization. Overall, we find a 4.5 percentage point (14%) increase in insulin prescription rates for patients with an A1c above 9 from 2013 to 2014 relative to patients below 9. In the placebo DiD for the two years prior to the outcomes-based rewards change, we find slight decreases in prescription rates for this group. Results in the table also show that (i) all of the relative increase in insulin prescriptions is coming from patients' primary care physicians and (ii) that most of the increase is coming from patients who have previously used insulin using it again– there is an increase in new insulin prescription rates as well but it accounts for only around 20% of the total treatment effect. The table also reports DiD estimates for different types of insulin, finding meaningful relative increases in fills for long acting insulin and Novo Nordisk brand insulin.

Table 8 presents the results of the switch to outcomes-based incentives on other health care quantities. There are meaningful percentage increases in whether a patient visited (i) an ED provider (27%) (ii) a hospitalist (47%) (iii) a nephrologist (44%) (iv) a diabetes education provider (21%) (v) a cardiologist (34%) and (vi) any specialist (9%). Of these increases, only the cardiologist

[8] We also run a version of this analysis in which we only analyze diagnoses and claims in the final 6 months of the reference period (January-June), to ensure that we capture only those occurring after the reference A1c test and to avoid the results being impacted by potential missing data in the final quarter of 2013. These results can be found in Appendix Table 14

DID Estimates: Impact of Outcomes Based Payments on Patients with A1c ≤ 9 (Insulin)								
	Test Period (July 2013-June 2015)				Placebo Period (July 2012-June 2014)			
	N	*Baseline for* ≥ 9	*Baseline for* < 9	*DID Coeff*	*N*	*Baseline for* ≥ 9	*Baseline for* < 9	*DID Coeff*
				Extensive Margin				
Insulin RX	3260	0.316	0.108	0.045** (0.023)	3277	0.328	0.096	-0.027 (0.022)
Insulin RX, PCP	3260	0.248	0.077	0.050** (0.020)	3277	0.272	0.068	-0.029 (0.020)
New Insulin RX	3260	0.013	0.003	0.009** (0.004)	3277	0.019	0.003	-0.004 (0.004)
Long Acting Insulin	3260	0.245	0.084	0.070*** (0.021)	3277	0.255	0.073	-0.018 (0.020)
Rapid Acting Insulin	3260	0.145	0.052	-0.002 (0.017)	3277	0.151	0.051	-0.013 (0.016)
Eli Lilly Brand Insulin	3260	0.045	0.018	-0.005 (0.010)	3277	0.063	0.019	-0.013 (0.010)
Novo Nordisk Brand Insulin	3260	0.096	0.033	0.037** (0.015)	3277	0.085	0.032	0.004 (0.012)
Sanofi Brand Insulin	3260	0.245	0.083	0.023 (0.020)	3277	0.256	0.073	-0.019 (0.020)
				Quantity				
Insulin RX	3260	2180.1	617.0	58.9 (247.9)	3277	2070.9	651.4	179.6 (237.6)
Long Acting Insulin	3260	1331.6	347.7	235.1 (169.0)	3277	1223.4	358.7	96.7 (140.4))
Rapid Acting Insulin	3260	796.0	269.3	-138.9 (128.0)	3277	847.5	292.7	16.8 (125.4)
Eli Lilly Brand Insulin	3260	373.8	92.0	-153.8 (95.2)	3277	347.3	115.8	62.3 (118.9)
Novo Nordisk Brand Insulin	3260	494.5	172.5	151.0 (118.8)	3277	485.3	165.3	26.2 (82.5)
Sanofi Brand Insulin	3260	1311.8	352.5	61.7 (163.5)	3277	1238.4	370.4	91.1 (142.6)
				Intensive Margin				
Insulin RX	516	5758.6	4842.8	-755.4 (661.2)	467	5445.1	6024.5	1549.8** (677.1)
Long Acting Insulin	401	4340.5	3492.4	-309.9 (523.9)	367	4019.7	4188.6	1080.1** (478.9)
Rapid Acting Insulin	261	4243.3	4044.5	-1022.2 (633.9)	240	4408.5	4925.3	1073.6 (653.6)
Eli Lilly Brand Insulin	87	6525.0	3912.3	-2806.0** (1351.6)	88	4828.1	5584.4	2190.6 (1414.3)
Novo Nordisk Brand Insulin	168	3924.5	4140.7	-882.4 (830.5)	147	4074.0	4354.3	715.3 (631.5)
Sanofi Brand Insulin	397	4271.2	3572.7	-4.5 (534.5)	366	4065.3	4309.0	1056.9** (495.4)
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 7: This table reports difference-in-differences estimates of the change in prescription rates for patients who are above the cutoff HbA1c test value of 9 and take Insulin

provider visit increase is statistically significant on its own, since, given the lower likelihood of these types of care, we lack power to detect even large effects for these outcomes. Table 9 also presents the impact of the change to outcomes-based payments on oral diabetes medications. We find limited impacts overall on oral diabetes medication prescriptions. However, we see a relative decrease of 3.5 percentage points in extensive margin usage of the high-cost oral drug Januvia, and relative increase of 1.9 percentage points in usage of another high-cost oral drug, Tradjenta.

DID Estimates: Impact of Outcomes Based Payments on Patients with A1c ≤ 9 (Health Care)

	Test Period (July 2013-June 2015)				Placebo Period (July 2012-June 2014)			
	N	*Baseline for* ≥ 9	*Baseline for* < 9	*DID Coeff*	*N*	*Baseline for* ≥ 9	*Baseline for* < 9	*DID Coeff*
Visited PCP	3260	0.7448	0.7666	0.0179 (0.0155)	3277	0.8035	0.8035	-0.0095 (0.0154)
Visited Specialist	3260	0.2197	0.2041	0.0204 (0.0196)	3277	0.1991	0.2085	0.0427** (0.0178)
Emergency	3260	0.0536	0.0446	0.0147 (0.0092)	3277	0.0578	0.0446	0.0079 (0.0094)
Hospitalist	3260	0.0233	0.0271	0.0110 (0.0073)	3277	0.0278	0.0279	0.0045 (0.0074)
Endocrinologist	3260	0.0833	0.0631	-0.0142 (0.0145)	3277	0.0616	0.0635	0.0278** (0.0115)
Nephrologist	3260	0.0175	0.0281	0.0077 (0.0084)	3277	0.0193	0.0305	0.0046 (0.0073)
Cardiologist	3260	0.0763	0.0881	0.0260** (0.0131)	3277	0.0771	0.0933	0.0059 (0.0116)
Diabetes Educator	3260	0.0157	0.0116	0.0034 (0.0050)	3277	0.0214	0.0134	0.0004 (0.0062)
Claim for Diabetes Education	3260	0.0163	0.0121	0.0041 (0.0053)	3277	0.0219	0.0135	-0.0002 (0.0063)
Total Spending	3260	1683.39	2051.07	307.15 (330.27)	3277	2459.00	2123.70	-189.33 (514.38)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 8: This table reports difference-in-differences estimates of the change in health care for patients who are above the cutoff HbA1c test value of 9

4.2 Impact on Health Care Outcomes

In this section we ask whether the switch to outcomes-based payments improved health care outcomes, in addition to impacting health care received. Table 10 below shows the results of these regressions, again reporting the coefficients β_2, transformed to probabilities. We study 14 relevant health care outcomes including 9 outcomes related to whether patients were diagnosed with complications that were (i) ketoacidosis (.15 percentage point increase from a baseline of 0) (ii) hyperosmolarity (88% increase) (iii) renal (39% increase) (iv) ophthamalic (15% increase) (v) neurological (15% increase) (vi) peripheral (6% increase) (vii) other (-3% decrease) and (viii) unspecified (10% increase). Of these eight specific complication-related outcomes, only the increase in renal complications is statistically significant on its own at the 5% level. Our ninth complications-related

DID Estimates: Impact of Outcomes Based Payments on Patients with A1c ≤ 9 (Oral & Other Meds)								
	Test Period (July 2013-June 2015)				Placebo Period (July 2012-June 2014)			
	N	*Baseline for* ≥ 9	*Baseline for* < 9	*DID Coeff*	*N*	*Baseline for* ≥ 9	*Baseline for* < 9	*DID Coeff*
				Extensive Margin				
Oral Diabetes RX	3260	0.7296	0.6836	-0.0168 (0.0196)	3277	0.7334	0.6925	0.0125 (0.0195)
Oral Diabetes RX, PCP	3260	0.6538	0.6194	-0.0032 (0.0223)	3277	0.6793	0.6341	-0.0139 (0.0214)
Oral Diabetes RX, New	3260	0.0052	0.0046	-0.0019 (0.0025)	3277	0.0118	0.0079	-0.0051* (0.0029)
High Cost Oral Diabetes RX	3260	0.2506	0.1898	-0.0296 (0.0217)	3277	0.2650	0.1908	-0.0076 (0.0218)
Januvia RX	3260	0.1084	0.0814	-0.0347** (0.0159)	3277	0.1204	0.0894	-0.0065 (0.0170)
Victoza RX	3260	0.0816	0.0489	-0.0074 (0.0142)	3277	0.0707	0.0422	-0.0072 (0.0117)
Janumet RX	3260	0.0524	0.0458	-0.0078 (0.0099)	3277	0.0567	0.0498	0.0104 (0.0114)
Tradjenta RX	3260	0.0111	0.0180	0.0189*** (0.0066)	3277	0.0177	0.0132	0.0002 (0.0078)
Any RX, New	3260	0.6399	0.5994	0.0182 (0.0181)	3277	0.6740	0.6219	0.0009 (0.0185)
				Quantity				
Oral Diabetes RX	3260	197.171	156.609	-8.105 (10.010)	3277	194.986	166.231	4.869 (9.627)
High Cost Oral Diabetes RX	3260	21.393	17.637	-3.740 (2.478)	3277	22.249	18.727	2.557 (2.746)
Januvia RX	3260	9.371	7.205	-2.471 (1.503)	3277	10.439	8.298	-0.256 (1.581)
Victoza RX	3260	1.848	1.067	-0.233 (0.380)	3277	1.339	0.956	0.039 (0.258)
Janumet RX	3260	9.196	7.802	-2.530 (2.029)	3277	8.978	8.334	2.627 (2.285)
Tradjenta RX	3260	0.979	1.562	1.495** (0.62)	3277	1.494	1.139	0.147 (0.864)
				Intensive Margin				
Oral Diabetes RX	2541	237.265	200.556	-10.565 (10.928)	2470	237.497	218.640	4.890 (10.702)
High Cost Oral Diabetes RX	711	69.417	77.206	-7.536 (7.155)	681	71.601	86.406	7.630 (7.392)
Januvia RX	306	74.830	71.517	-10.130 (7.224)	285	80.313	83.360	-11.596* (6.833)
Victoza RX	199	18.900	17.904	-3.464 (2.733)	169	14.392	19.097	3.653* (2.142)
Janumet RX	163	142.826	145.036	-33.781* (19.605)	171	132.054	149.580	20.878 (18.308)
Tradjenta RX	42	63.000	71.959	0.831 (19.791)	53	67.500	63.920	8.605 (20.676)
Note:								*p<0.1; **p<0.05; ***p<0.01

Table 9: This table reports difference-in-differences estimates of the change in prescription rates for patients who are above the cutoff HbA1c test value of 9 and are taking oral or other non-Insulin medications

outcome here groups all of these complications together and finds a 15% increase in overall diagnosed complications, significant at the 10% level. The other 5 outcomes we investigate are (i) a hypoglycemia diagnosis (+2%) (ii) a kidney disease diagnosis (+31%) (iii) a lower limb cellulitis diagnosis (+14%) (iv) a heart attack diagnosis (+35%) and (v) whether a patient had uncontrolled diabetes (+9%). Of these five, only kidney disease diagnosis is statistically significant, at a 10% level.

These results all point in one direction: a higher level of visits to specialists and a greater degree of diagnosed complications or other major health problems associated with diabetes. At first glance, this seems counter-intuitive: people are using more intensive medication as a result of the shift to outcomes-based payments, but having what looks like worse health outcomes. This pattern can be rationalized by noting that, one way to help control diabetes is via PCP referrals to specialists. In doing so, the overall level of major health issues diagnosed may increase as a result of increased specialist attention, rather than due to an increase in actual health issues (see e.g, Song et al. (2010)). In future work, with a longer-run panel of health claims and health outcomes for these patients, it would be interesting to study whether there is a positive effect on long-run health outcomes as a result of increased interaction with the medical system due to the shift to outcomes-based incentives.

One way to deal with this issue is to directly study the impact of the shift to outcomes-based payments on an actual health outcome, HbA1c scores. To this end, we look directly at whether or not beginning to use (or restarting) insulin is associated with lower HbA1c levels. Figure 10 presents an event study of HbA1c level relative to the event of starting or restarting use of insulin. The figure clearly shows that starting or restarting use of insulin is associated with meaningful drops in average HbA1c levels. This is even more true when we condition on those who have an A1c value above 9 when starting insulin, as shown in figure 11. Figures 17-19 in the appendix show these event studies separately by year, with similar results. In both 2013-14, patients start insulin when their A1c value is around 9, on average. In contrast, patients start insulin at slightly higher A1c values in 2012, closer to 9.5.

DID Estimates: Impact of Outcomes Based Payments on Patients with A1c ≤ 9 (Health Outcomes)								
	Test Period (July 2013-June 2015)				Placebo Period (July 2012-June 2014)			
	N	*Baseline for* ≥ 9	*Baseline for* < 9	*DID Coeff*	*N*	*Baseline for* ≥ 9	*Baseline for* < 9	*DID Coeff*
Indicated Uncontrolled	3260	0.3998	0.2025	0.0340 (0.0216)	3277	0.3913	0.1975	0.0339 (0.0212)
Has Diabetes Complications	3260	0.2150	0.1692	0.0329* (0.0199)	3277	0.1997	0.1265	-0.0012 (0.0195)
Ketoacidosis	3093	0.0000	0.0004	0.0015 (0.0012)	3274	0.0000	0.0004	0.0033* (0.0018)
Hyperosmolarity	2595	0.0017	0.0030	0.0015 (0.0013)	3134	0.0043	0.0053	-0.0027 (0.0026)
Renal	3260	0.0769	0.0765	0.0301** (0.0152)	3277	0.0551	0.0438	-0.0002 (0.0137)
Ophthalmic	3260	0.0763	0.0517	0.0114 (0.0121)	3277	0.0803	0.0394	0.0002 (0.0115)
Neurological	3260	0.0507	0.0300	0.0078 (0.0107)	3277	0.0385	0.0204	0.0174* (0.0091)
Peripheral	3260	0.0082	0.0073	0.0005 (0.0048)	3134	0.0091	0.0062	-0.0015 (0.0049)
Other	3260	0.0117	0.0121	-0.0004 (0.0053)	3277	0.0150	0.0098	0.0042 (0.0062)
Unspecified	3260	0.0332	0.0211	0.0034 (0.0070)	3277	0.0369	0.0233	0.0017 (0.0091)
Diagnosis: hypoglycemia	3260	0.0140	0.0150	0.0003 (0.0059)	3277	0.0171	0.0120	0.0041 (0.0065)
Diagnosis: kidney disease	3260	0.0804	0.0872	0.0252* (0.0152)	3277	0.0755	0.0780	-0.0144 (0.0141)
Diagnosis: lower limb cellulitis	3260	0.0058	0.0040	0.0008 (0.0026)	3277	0.0091	0.0061	0.0012 (0.0030)
Diagnosis: heart attack	3260	0.0023	0.0026	0.0008 (0.0026)	3277	0.0027	0.0030	0.0001 (0.0024)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 10: This table reports difference-in-differences estimates of the change in health outcomes for patients who are above the cutoff HbA1c test value of 9

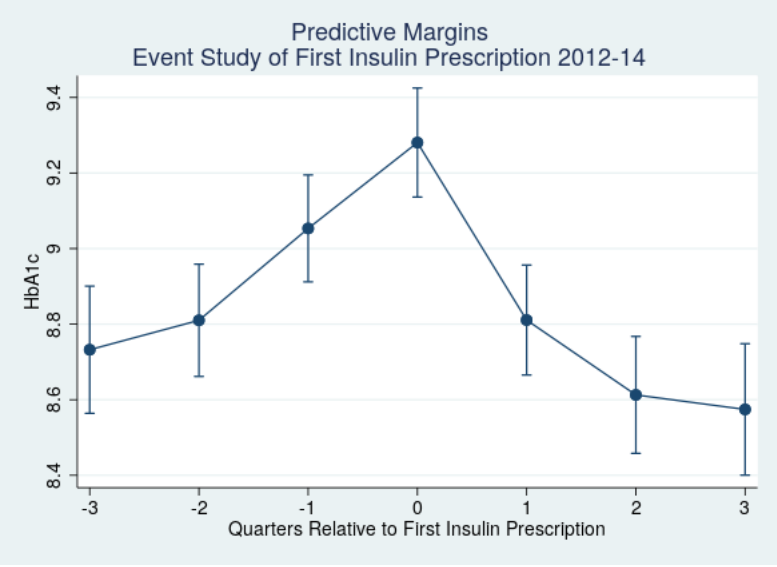

Figure 10: Event study of A1c response to starting insulin

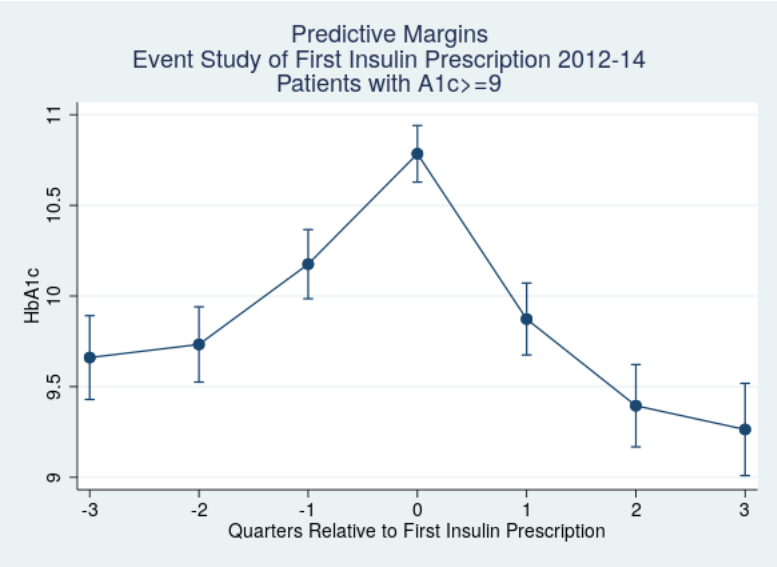

Figure 11: Event study of A1c response to starting insulin, conditional on starting A1c value ≥ 9

Figure 20 in the appendix shows the change in HbA1c level by baseline test value over the year just prior to the shift to outcomes-based rewards, compared to the year following the shift to outcomes-based rewards. The figure shows that test results increase on average across the distribution. The increase is slightly lower during the post-outcomes based incentive period, though these differences are not statistically significant. This indicates that though increasing insulin prescriptions for high A1c patients likely leads to greater A1c control for this sub-population, this doesn't translate to statistically significant reductions in the overall distribution of A1c results for patients with diabetes. Note that Figure 21 in the appendix, which shows changes in A1c patterns over the two pre-outcomes based incentive years, shows a different pattern of A1c changes in 2012, two years before the shift to outcomes based payments. Here, we see more disparity in outcomes between the initially uncontrolled high A1c patients and those who start out with a lower A1c test value. By 2013, the year prior to the start of outcomes based payments, we see relatively larger increases for those starting out low, and relatively smaller increases for those with A1c values already above 9.

4.3 Physician Heterogeneity

In addition to assessing the impact of outcomes-based payments on physician behavior overall, we also want to assess the heterogeneous impacts of this payment shift in order to determine if the effect is driven by certain types of physicians changing their behavior. First, we investigate a modified version of our baseline DiD specification:

$$y_{qi} = \beta_1 + \beta_2 I(A1c \geq 9)_{qi} * Post_q + \beta_3 Post_q + \beta_4 I(A1c \geq 9)_{qi} + \beta_5 I(A1c \geq 9)_{qi} * Post_q * Z_q + Z_i + Z_q + \epsilon_{qi} \quad (3)$$

Here, the β_5 coefficients measure the differential DiD coefficient from the baseline regression

as a function of the demographic variables Z_q, including age, gender, and primary care provider specialty. A relative change in prescription rates for patients above 9 supports physician response to the rewards, which is indicated by the variable $I(A1c \geq 9)_{qi} * Post_q$. Interactions of this variable with physician characteristics indicate a differential response by type of physician. Table 11 presents the results of the model, run separately for each dimension of heterogeneity. Specifically focusing on Extensive Margin Insulin Fills, we find that male physicians, physicians above the median age, and physicians who are internal medicine specialists have larger positive treatment effects than other physicians in our sample (though these differences are not statistically significant). We also analyze whether belonging to a physician group is associated with differential physician behavior change. We find that physicians working in physician groups have an 11.3 percentage point larger impact than solo practitioners, implying that physicians in groups are responsible for essentially all of the documented increases in insulin prescribing. This is consistent with the hypotheses that it is easier to disseminate guidelines within an organization and easier to make payment changes salient. Furthermore, we see statistically significant differences in oral diabetes medication prescribing behavior among group vs non-group physicians. The difference-in-differences estimate of changes in oral diabetes prescriptions for patients with A1c above 9 is about 8 percentage points lower for group physicians. This could indicate a possible substitution of insulin for oral diabetes medication among group physicians. Note that the table also includes placebo results for the two years prior to the shift to outcomes-based payments, finding no effect for group vs. non-group.

Next, we estimate physician heterogeneity in a more granular way. We start by estimating the following regression on the sample of patients used for analysis above. In this case, Z_q and Z_i are quarter of year patient demographic variables only (insurance type, patient age, sex, distance to physician, location, and family size).

$$y_{qi} = Z_i + Z_q + \epsilon_{qi} \tag{4}$$

Residuals are obtained, and the sample is limited to patients of physicians who have at least two patients falling below 9 A1c, and two falling equal to or above 9 (according to the first A1c test value in the reference year), in both the pre and post outcomes based payments period (N=81). Then, mean residuals are calculated by physician within each quadrant (above/below 9 X pre/post outcome payments), and a by-physician difference-in-differences estimate is produced. These are then plotted in the histogram below. This method risk-adjusts for patient characteristics and investigates changes in behavior conditional on patient risk-adjustment.

We can see there is significant heterogeneity in how physicians change their insulin prescribing behavior over the study period– only around 1/3 of physicians keep insulin RX rates for high and low A1c patients at roughly the same relative rate (+-5 percentage point change) within this period. Around 20% of this sample of physicians increase prescription rates for high A1c patients by more than 35 percentage points relative to their low A1c patients. The estimates show that there is meaningful heterogeneity across physicians in the impact that the shift to outcomes-based payments has on their prescribing behavior.

DID Estimates: Physician Heterogeneous Impacts of Outcomes Based Payments on Patients with A1c ≤ 9								
	Test Period (July 2013-June 2015)				Placebo Period (July 2012-June 2014)			
	N	*Baseline for* ≥ 9	*Baseline for* < 9	*DID Coeff*	*N*	*Baseline for* ≥ 9	*Baseline for* < 9	*DID Coeff*
Physician in a Group vs. Independent								
Any Visit to PCP	3260	0.7448	0.7666	0.0039	3277	0.8035	0.8035	0.0143
				(0.0311)				(0.0306)
Any Oral Diabetes RX Fill	3260	0.7296	0.6836	-0.0829**	3277	0.7334	0.6925	0.0082
				(0.0403)				(0.0396)
Any Oral Diabetes RX Fill, PCP	3260	0.6538	0.6194	-0.0925**	3277	0.6793	0.6341	0.0511
				(0.0452)				(0.0431)
Any Insulin RX Fill	3260	0.3159	0.1078	0.1126**	3277	0.3282	0.0957	-0.0273
				(0.0455)				(0.0453)
Any Insulin RX Fill, PCP	3260	0.2483	0.0770	0.0915**	3277	0.2725	0.0681	-0.0027
				(0.0418)				(0.0426)
Physician in top 1/3 of 2013 Cozeva Users vs. Bottom 2/3								
Any Visit to PCP	3260	0.7448	0.7666	0.0127	3277	0.8035	0.8035	-0.0043
				(0.0327)				(0.0329)
Any Oral Diabetes RX Fill	3260	0.7296	0.6836	0.0309	3277	0.7334	0.6925	-0.0504
				(0.0445)				(0.0409)
Any Oral Diabetes RX Fill, PCP	3260	0.6538	0.6194	0.0392	3277	0.6793	0.6341	-0.0814*
				(0.0506)				(0.0465)
Any Insulin RX Fill	3260	0.3159	0.1078	-0.0856*	3277	0.3282	0.0957	0.0354
				(0.0497)				(0.0432)
Any Insulin RX Fill, PCP	3260	0.2483	0.0770	-0.0283	3277	0.2725	0.0681	0.0007
				(0.0439)				(0.0377)
Physician Internal Medicine vs. Other Specialities								
Any Visit to PCP	3260	0.7448	0.7666	-0.0458	3277	0.8035	0.8035	-0.0467
				(0.0419)				(0.0393)
Any Oral Diabetes RX Fill	3260	0.7296	0.6836	-0.0337	3277	0.7334	0.6925	-0.0003
				(0.0445)				(0.0467)
Any Oral Diabetes RX Fill, PCP	3260	0.6538	0.6194	-0.0692	3277	0.6793	0.6341	-0.0126
				(0.0536)				(0.0512)
Any Insulin RX Fill	3260	0.3159	0.1078	0.0468	3277	0.3282	0.0957	-0.0467
				(0.0596)				(0.0513)
Any Insulin RX Fill, PCP	3260	0.2483	0.0770	0.0388	3277	0.2725	0.0681	-0.0400
				(0.0574)				(0.0487)
Physician Male vs. Female								
Any Visit to PCP	3260	0.7448	0.7666	0.0286	3277	0.8035	0.8035	-0.0259
				(0.0391)				(0.0421)
Any Oral Diabetes RX Fill	3260	0.7296	0.6836	0.0390	3277	0.7334	0.6925	-0.0619
				(0.0525)				(0.0504)
Any Oral Diabetes RX Fill, PCP	3260	0.6538	0.6194	-0.0258	3277	0.6793	0.6341	-0.0269
				(0.0586)				(0.0563)
Any Insulin RX Fill	3260	0.3159	0.1078	0.0342	3277	0.3282	0.0957	-0.0386
				(0.0602)				(0.0608)
Any Insulin RX Fill, PCP	3260	0.2483	0.0770	0.0198	3277	0.2725	0.0681	-0.0457
				(0.0557)				(0.0574)
Physician Below Median Provider Age vs Above								
Any Visit to PCP	3260	0.7448	0.7666	0.0027	3277	0.8035	0.8035	0.0176
				(0.0325)				(0.0324)
Any Oral Diabetes RX Fill	3260	0.7296	0.6836	-0.0396	3277	0.7334	0.6925	0.0065
				(0.0415)				(0.0411)
Any Oral Diabetes RX Fill, PCP	3260	0.6538	0.6194	-0.0101	3277	0.6793	0.6341	-0.0571
				(0.0480)				(0.0457)
Any Insulin RX Fill	3260	0.3159	0.1078	-0.0437	3277	0.3282	0.0957	0.0545
				(0.0487)				(0.0471)
Any Insulin RX Fill, PCP	3260	0.2483	0.0770	-0.0318	3277	0.2725	0.0681	0.0517
				(0.0442)				(0.0422)
Note:				*p<0.1; **p<0.05; ***p<0.01				

Table 11: This table reports difference-in-differences estimates of the change in health care for patients who are above the cutoff HbA1c test value of 9, assuming heterogeneity on the physiscian level; 5 different heterogeneity indicators are interacted with the main specification in separate regressions.

Heterogeneity Samples		
Indicator	*N Patients*	*N Physicians*
In a Group	1458	187
Top Cozeva User in 2013	2287	202
Internal Medicine Specialist	2534	238
Male	2595	265
Below Median Age	1432	183

Table 12: This indicates the sample breakdown for the heterogeneity indicators used in the analysis above. Total sample is 3260 patients and 367 physicians.

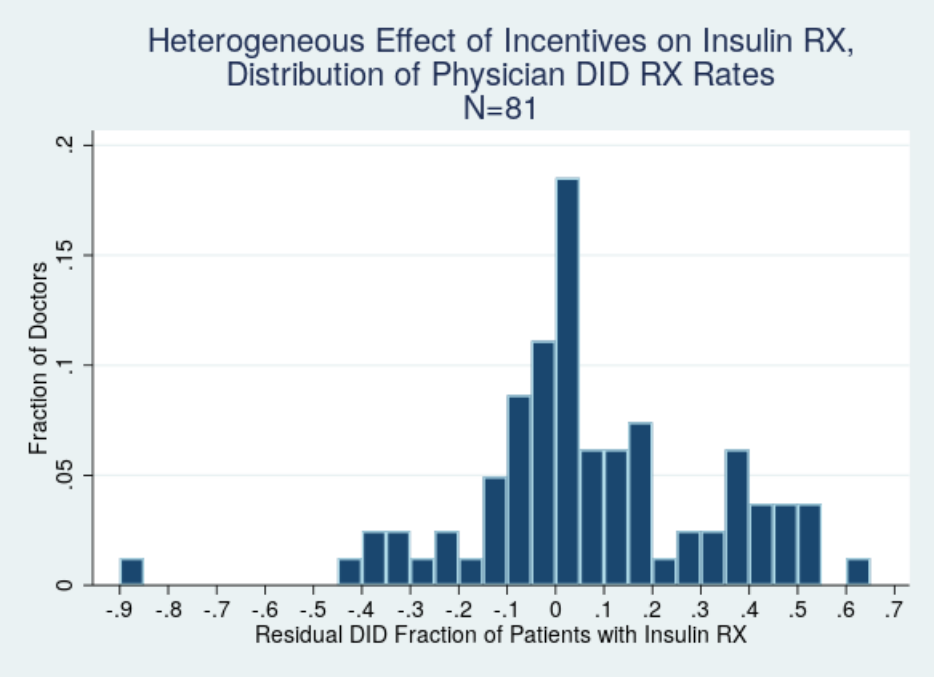

Figure 12: Shows distribution of physician level changes in insulin prescribing behavior for low and high A1c patients before and after the introduction of outcomes based payments.

Figure 13 analyzes how the difference in insulin prescription rates changes after outcomes incentives are introduced for high IT users vs. low IT users (defined by activity in IT system in Cozeva system during 2013). The figure shows the difference in average physician-level insulin prescription rates for patients with $A1c \geq 9$ vs. those with $A1c < 9$ by quarter, separately for the top 1/3 and the bottom 2/3 of IT users (note that this breakdown is based on all physicians in 2013). The figure shows that, for high IT users, insulin prescription rates were already relatively high before the introduction of outcomes based payments. These prescription rates then increased relatively little after the implementation of outcomes-based payments for these physicians, suggesting that high IT use (and/or other factors associated with high IT use) could serve as a substitute for outcomes-based payments. Conversely, low IT users had no meaningful increase in insulin prescription rates prior to the onset of outcomes-based payments, but did meaningfully increase their insulin prescription

rates after the shift to outcomes-based payments. This is consistent with the heterogeneity results reported in Table 11, which show a negative impact on insulin RX rates for high A1c patients of physicians who were in the top third of Cozeva users relative to those of physicians in the bottom two thirds. One possible explanation for these results is that high IT use physicians leverage their use of IT to improve patient care in the absence of outcomes-based incentives, perhaps due to altruistic motivations, while low IT use physicians are directly incentivized by the outcomes-based program to improve care.

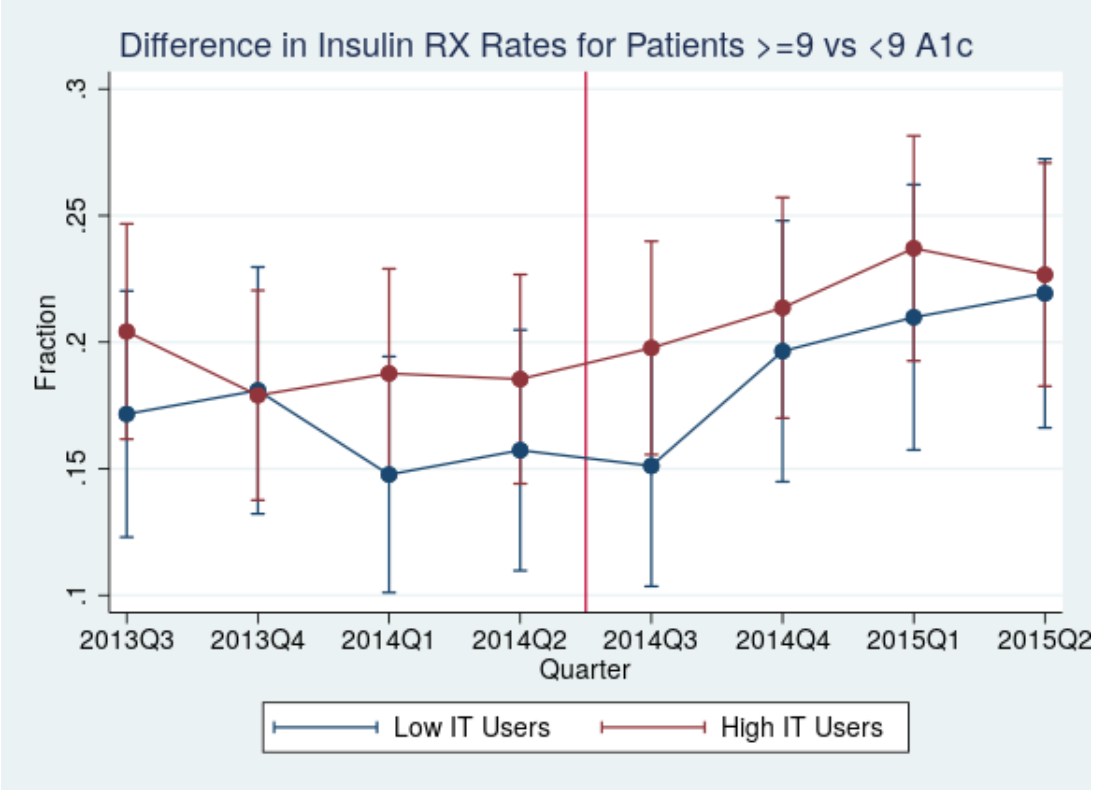

Figure 13: Shows Difference in Average Doc-Level Insulin Prescription Rate for Patients $A1c \geq 9$ vs $A1c < 9$ by Quarter, separately for top 1/3 and bottom 2/3 of IT users

5 Conclusion

We study how providers respond to a shift from process to outcome-based payments in the treatment of patients with diabetes. We find significant changes in how patients with lack of blood glucose control are treated, including increased rates of insulin prescription and more visits to specialists. We document a 14% increase in insulin prescription rates for patients with A1c test values above 9. We also examine downstream impacts on patient health. We find that patients with uncontrolled diabetes are more likely to be diagnosed with complications as a result of the shift to outcomes-based incentives, which we attribute to endogenous increases in detection of these diseases stemming from more intensive care. Though we are unable to detect statistically significant changes in the overall distribution of patient A1c test values, we link starting or re-starting insulin prescriptions to

clinically and statistically significant reductions in A1c, particularly for patients starting out with very high A1c values.

We also investigate the predictors and mechanisms of physician response to outcomes based incentives, finding significant heterogeneity. We find that providers who are part of physician groups respond more intensively, increasing insulin prescription rates for uncontrolled patients by 11.3 percentage points more than solo practitioners, meaning that these physicians are responsible for the entirety of the impact we observe. We also find higher responses among physicians with relatively limited use of IT, whose use of insulin to treat uncontrolled patients increases to match that of high-IT users only after the incentive changes.

Though blood sugar control is only one of many relevant health indicators for patients with diabetes, there is substantial evidence that lower A1c levels are associated with significant decreases in the probability of complications from diabetes, and reductions in overall medical costs (see, e.g. Fitch et al. (2013)). Nearly 20% of medical expenditures in the U.S. are attributed to patients with diabetes, and almost one in ten people have this diagnosis. Addressing how to improve health and reduce costs for this population is both a case study in efficiently achieving care quality, and significant to the overall healthcare system. One valuable avenue for future research is to dig further into the mechanisms underlying physician responses to incentive programs (both outcomes and process based) and how those programs impact physician behavior via habit formation vs. directly incentivizing effort. Another valuable avenue for future research is to study study outcome-based incentives in other disease contexts, where physicians and/or patients may have relatively more or less control over health outcomes.

References

Abaluck, Jason, Leila Agha, Chris Kabrhel, Ali Raja, and Arjun Venkatesh, "The Determinants of Productivity in Medical Testing," *American Economic Review*, 2016, *106* (2), 3730–3764.

ADA, "Standards of Medical Care in Diabetes—2015 Abridged for Primary Care Providers," *Clinical Diabetes*, apr 2015, *33* (2), 97–111.

Barro, Jason and Nancy Beaulieu, "Selection and Improvement: Physician Responses to Financial Incentives," 2003. NBER working paper no. 10017.

Brot-Goldberg, Zarek C, Amitabh Chandra, Benjamin R Handel, and Jonathan T Kolstad, "What does a deductible do? The impact of cost-sharing on health care prices, quantities, and spending dynamics," *The Quarterly Journal of Economics*, 2017, *132* (3), 1261–1318.

Chan, David, "Teamwork and Moral Hazard: Evidence from the Emergency Department," *Journal of Political Economy*, 2016, *124* (3), 734–770.

_ , "The Efficiency of Slacking Off: Evidence from the Emergency Department," *Econometrica*, 2018, *86* (3), 997–1030.

Clemens, Jeffrey and Joshua Gottlieb, "Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health?," *American Economic Review*, 2014, *104* (4), 1320–1349.

Devlin, Rose Ann and Sisira Sarma, "Do Physician Reimbursement Schemes Matter? The Case of Canadian Family Physicians," *Journal of Health Economics*, 2008, *27* (5), 1168–1181.

Dumont, Etienne, Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer, "Physicians' Multitasking and Incentives: Empirical Evidence from a Natural Experiment," *Journal of Health Economics*, 2008, *27* (6), 1436–1450.

Einav, Liran, Amy Finkelstein, and Neale Mahoney, "Long-Term Care Hospitals: A Case Study in Waste," 2018. NBER working paper no. 24946.

Eliason, Paul, Paul Grieco, Ryan McDevitt, and James Roberts, "Strategic Patient Discharge: The Care of Long-Term Care Hospitals," *American Economic Review*, 2018, *108* (11), 3232–3265.

Fitch, Kathryn, Bruce S Pyenson, and Kosuke Iwasaki, "Medical claim cost impact of improved diabetes control for medicare and commercially insured patients with type 2 diabetes," *Journal of Managed Care Pharmacy*, 2013, *19* (8), 609–620.

Frech, HE, Christopher Whaley, Benjamin R Handel, Liora Bowers, Carol J Simon, and Richard M Scheffler, "Market power, transactions costs, and the entry of accountable care organizations in health care," *Review of Industrial Organization*, 2015, *47* (2), 167–193.

Glied, Sherry and Joshua Graff Zivin, "How Do Doctors Behave When Some (But Not All) of Their Patients are in Managed Care?," *Journal of Health Economics*, 2002, *21* (2), 337–353.

Gruber, Jonathan and Maria Owings, "Physician Financial Incentives and Cesarean Section Delivery," 1994. NBER working paper no. 4933.

Handel, Benjamin, Jonathan Holmes, Jonathan Kolstad, and Kurt Lavetti, "Insurer Innovation: Evidence from Utah," 2019. Berkeley working paper.

Higgins, Victoria, James Piercy, Adam Roughley, Gary Milligan, Andrea Leith, James Siddall, and Mike Benford, "Trends in medication use in patients with type 2 diabetes mellitus: a long-term view of real-world treatment between 2000 and 2015," *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, nov 2016, *Volume 9*, 371–380.

Kolstad, Jonathan T, "Information and quality when motivation is intrinsic: Evidence from surgeon report cards," *The American Economic Review*, 2013, *103* (7), 2875–2910.

Lee, Tai-Ti, Shou-Hsia Cheng, Chi-Chen Chen, and Mei-Shu Lai, "A Pay-for-Performance Program for Diabetes Care in Taiwan: A Preliminary Assessment," *The American Journal of Managed Care*, 2010, *16* (1), 65–69.

Li, Jinhu, Jeremiah Hurley, Philip DeCicca, and Gioia Buckley, "Physician response to pay-for-performance: evidence from a natural experiment," *Health Economics*, 2014, *23* (8), 962–978.

Montvida, Olga, Jonathan Shaw, John J. Atherton, Frances Stringer, and Sanjoy K. Paul, "Long-term Trends in Antidiabetes Drug Usage in the U.S.: Real-world Evidence in Patients Newly Diagnosed With Type 2 Diabetes," *Diabetes Care*, nov 2017, *41* (1), 69–78.

Mullen, Kathleen, Richard Frank, and Meredith Rosenthal, "Can You Get What You Pay For? PayforPerformance and the Quality of Healthcare Providers," *The RAND Journal of Economics*, 2010, *41* (1), 64–91.

Obermeter, Ziad, "INSERT TITLE," 2018. INSERT PAPER NOTE.

Shafrin, Jason, "Operating on Commission: Analyzing How Physician Financial Incentives Affect Surgery Rates," *Health Economics*, 2010, *19* (5), 562–580.

Song, Yunjie, Jonathan Skinner, Julie Byn6um, Jason Sutherland, John Wennberg, and Elliot Fisher, "Regional Variations in Diagnostic Practices," *New England Journal of Medicine*, 2010, *363*, 45–53.

A Appendix: Supplementary Figures and Tables

	Oahu	Hawai'I	Maui	Kaua'I	Moloka'I	Lana'I	Ni'ihau
Population	953,207	185,079	144,444	66,921	7,345	3,135	170
Population Density (per sq mi)	1,468	46	162	106	28	23	1.9

Figure 14: Hawaiian Population and Population Density, by Island

	Oahu	Hawai'I	Maui	Kaua'I	Moloka'I	Lana'I	Ni'ihau
Population	953,207	185,079	144,444	66,921	7,345	3,135	170
Population Density (mile^2)	1,468	46	162	106	28	23	1.9

	Commercial	Medicare	QUEST
Members	506,097	38,829	122,727
Providers	988	524	247

Figure 15: P4Q Program Participation as of Dec, 2012

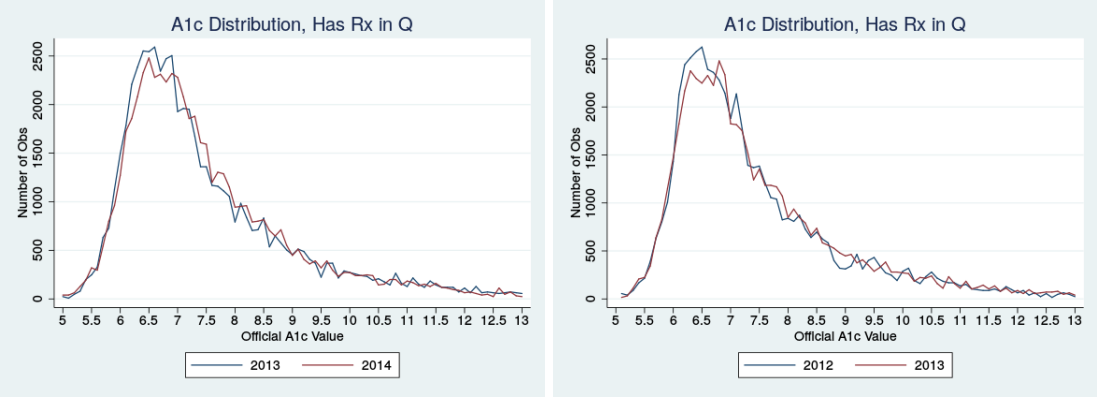

Figure 16: This figure presents the distribution of patient HbA1c over time.

DID Estimates: Impact of Outcomes Based Payments on Patients with A1c ≤ 8.5 (Insulin)								
	Test Period (July 2013-June 2015)				Placebo Period (July 2012-June 2014)			
	N	*Baseline for* ≤ 8.5	*Baseline for* ≥ 8.5	*DID Coeff*	*N*	*Baseline for* ≤ 8.5	*Baseline for* ≥ 8.5	*DID Coeff*
				Extensive Margin				
Insulin RX	3260	0.300	0.099	0.041** (0.018)	3277	0.298	0.089	-0.010 (0.018)
Insulin RX, PCP	3260	0.232	0.070	0.043*** (0.016)	3277	0.241	0.063	-0.016 (0.016)
New Insulin RX	3260	0.014	0.003	0.006* (0.004)	3277	0.016	0.002	-0.003 (0.003)
Long Acting Insulin	3260	0.228	0.078	0.075*** (0.017)	3277	0.233	0.068	-0.008 (0.017)
Rapid Acting Insulin	3260	0.139	0.048	-0.004 (0.013)	3277	0.145	0.047	-0.010 (0.013)
Eli Lilly Brand Insulin	3260	0.044	0.016	-0.010 (0.008)	3277	0.059	0.017	-0.007 (0.009)
Novo Nordisk Brand Insulin	3260	0.092	0.030	0.035*** (0.012)	3277	0.084	0.029	0.000 (0.009)
Sanofi Brand Insulin	3260	0.228	0.077	0.028* (0.016)	3277	0.234	0.068	-0.008 (0.017)
				Quantity				
Insulin RX	3260	1974.9	568.0	230.1 (187.9)	3277	1913.4	603.9	222.4 (202.9)
Long Acting Insulin	3260	1202.6	316.8	367.9*** (122.5)	3277	1132.1	328.7	141.2 (117.4)
Rapid Acting Insulin	3260	734.2	251.2	-110.0 (98.7)	3277	781.2	275.2	34.7 (106.4)
Eli Lilly Brand Insulin	3260	329.4	84.8	-152.4** (70.9)	3277	317.7	108.9	85.8 (98.1)
Novo Nordisk Brand Insulin	3260	457.3	161.3	229.0** (90.6)	3277	452.4	154.0	-0.4 (65.5)
Sanofi Brand Insulin	3260	1188.2	321.9	153.5 (120.9)	3277	1143.3	341.0	136.9 (119.3)
				Intensive Margin				
Insulin RX	516	5453.0	4901.1	-13.8 (581.4)	467	5514.0	6055.8	1597.0** (675.4)
Long Acting Insulin	401	4198.9	3457.5	168.2 (455.0)	367	4089.0	4156.4	1168.2** (460.2)
Rapid Acting Insulin	261	4162.8	4067.4	-981.4* (592.7)	240	4266.3	5130.4	1474.7** (643.6)
Eli Lilly Brand Insulin	87	6015.3	3867.9	-2288.7* (1162.8)	88	4655.5	5879.8	3041.1** (1321.0)
Novo Nordisk Brand Insulin	168	3828.6	4242.4	-308.6 (750.7)	147	3968.4	4486.4	619.6 (627.3)
Sanofi Brand Insulin	397	4144.6	3550.0	365.4 (486.2)	366	4126.2	4295.5	1147.7** (475.5)

Table 13: This table reports difference-in-differences estimates of the change in prescription rates for patients who are above the cutoff HbA1c test value of 8.5

DID Estimates: Impacts of Outcomes Based Payments on Patients with A1c (Healthcare) ≤ 9
Robustness check limited to final two quarters of reference year

	N	*Baseline for* ≥ 9	*Baseline for* < 9	*% Change*	*DID Coeff*
Visited PCP	3260	0.7448	0.7666	+3%	0.0165 (0.0207)
Visited Specialist	3260	0.2197	0.2041	-7%	0.0182 (0.0238)
Emergency	3260	0.0536	0.0446	-17%	0.0146 (0.0136)
Hospitalist	3260	0.0233	0.0271	+16%	0.0052 (0.0103)
Endocrinologist	3260	0.0833	0.0631	-24%	-0.0252 (0.0164)
Nephrologist	3260	0.0175	0.0281	+61%	0.0078 (0.0090)
Cardiologist	3260	0.0763	0.0881	+15%	0.0436*** (0.0163)
Diabetes Educator	3260	0.0157	0.0116	-26%	0.0019 (0.0064)
Claim for Diabetes Education	3260	0.0163	0.0121	-26%	0.0029 (0.0065)
Oral Diabetes RX	3260	0.7296	0.6836	-6%	-0.0395* (0.0226)
Oral Diabetes RX, PCP	3260	0.6538	0.6194	-5%	-0.0271 (0.0255)
Oral Diabetes RX, New	3260	0.0052	0.0046	-12%	0.0019 (0.0028)
Insulin RX	3260	0.3159	0.1078	-66%	0.0507** (0.0249)
Insulin RX, PCP	3260	0.2483	0.0770	-69%	0.0560** (0.0227)
New Insulin RX	3260	0.0134	0.0034	-75%	0.0014 (0.0055)
New Any RX	3260	0.6399	0.5994	-6%	-0.0035 (0.0243)

Table 14: This table reports difference-in-differences estimates of the change in health outcomes for patients who are above the cutoff HbA1c test value of 9, limited to the last two quarters of the reference year to check for robustness.

DID Estimates: Impacts of Outcomes Based Payments on Patients with A1c (Health Outcomes) ≤ 9
Robustness check limited to final two quarters of reference year

	N	*Baseline for* ≥ 9	*Baseline for* < 9	*% Change*	*DID Coeff*
Indicated Uncontrolled	3260	0.3998	0.2025	-49%	0.0313 (0.0254)
Has Diabetes Complications	3260	0.2150	0.1692	-21%	0.0402* (0.0229)
Ketoacidosis	3093	0.0000	0.0004		0.0027 (0.0022)
Hyperosmolarity	1923	0.0017	0.0030	+72%	0.0000 (0.0027)
Renal	3260	0.0769	0.0765	-1%	0.0230 (0.0170)
Ophthalmic	3260	0.0763	0.0517	-32%	0.0212 (0.0146)
Neurological	3260	0.0507	0.0300	-41%	0.0039 (0.0121)
Peripheral	3260	0.0082	0.0073	-10%	0.0024 (0.0046)
Other	3260	0.0117	0.0121	+4%	0.0034 (0.0071)
Unspecified	3260	0.0332	0.0211	-36%	0.0013 (0.0090)
Diagnosis: hypoglycemia	3260	0.0140	0.0150	+7%	0.0035 (0.0078)
Diagnosis: kidney disease	3260	0.0804	0.0872	+8%	0.0196 (0.0171)
Diagnosis: lower limb cellulitis	3260	0.0058	0.0040	-32%	0.0040 (0.0036)
Diagnosis: heart attack	3095	0.0023	0.0026	+10%	0.0052 (0.0042)

Table 15: This table reports difference-in-differences estimates of the change in health outcomes for patients who are above the cutoff HbA1c test value of 9, limited to the last two quarters of the reference year to check for robustness.

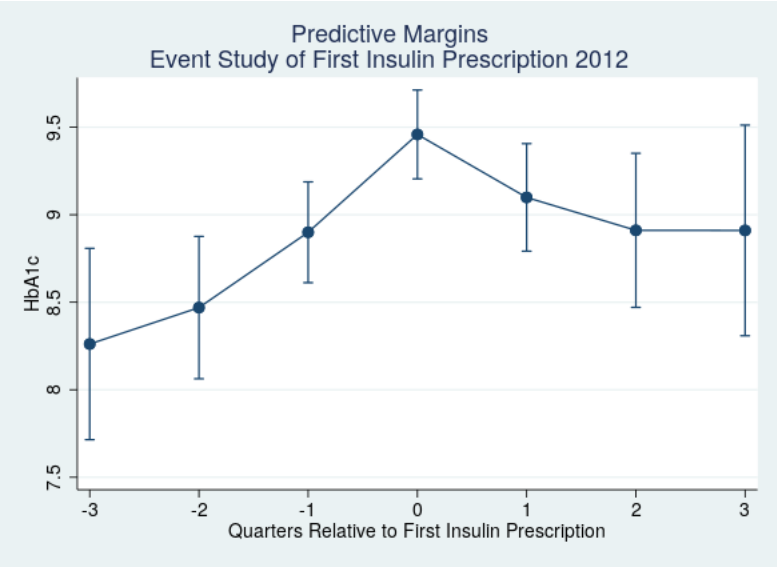

Figure 17: Event study of A1c response to starting insulin in calendar year 2012.

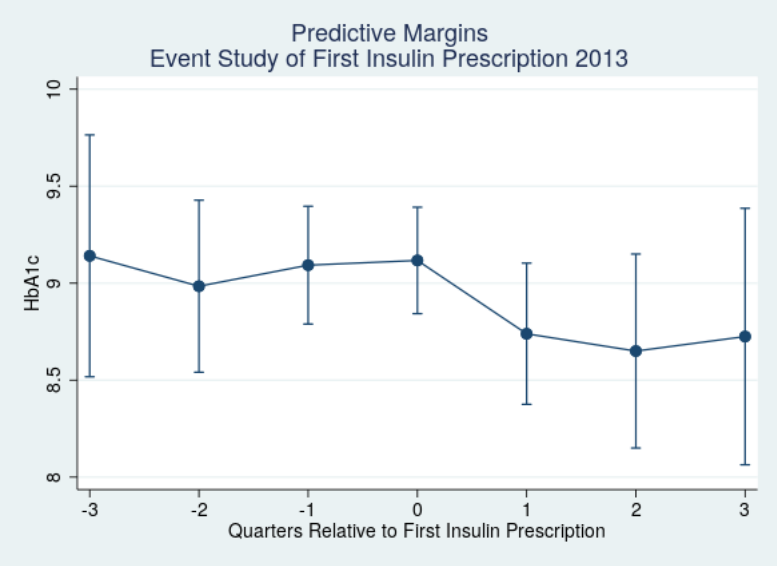

Figure 18: Event study of A1c response to starting insulin in calendar year 2013.

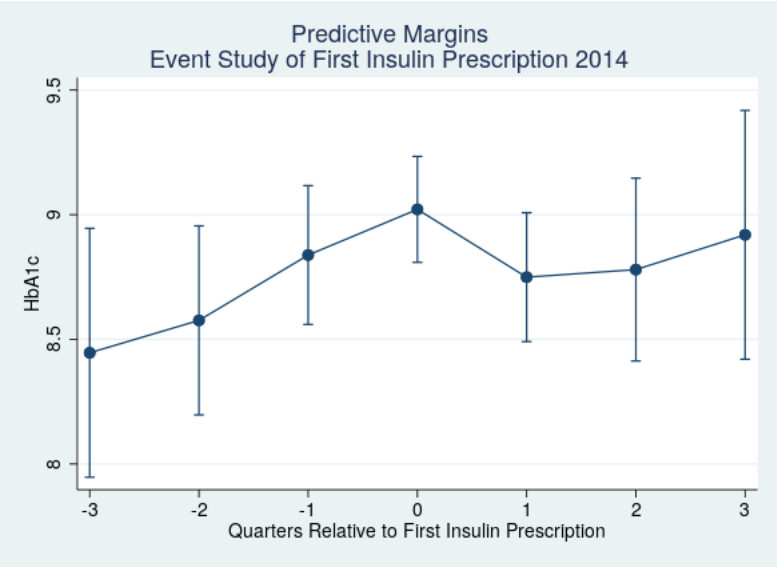

Figure 19: Event study of A1c response to starting insulin in calendar year 2014.

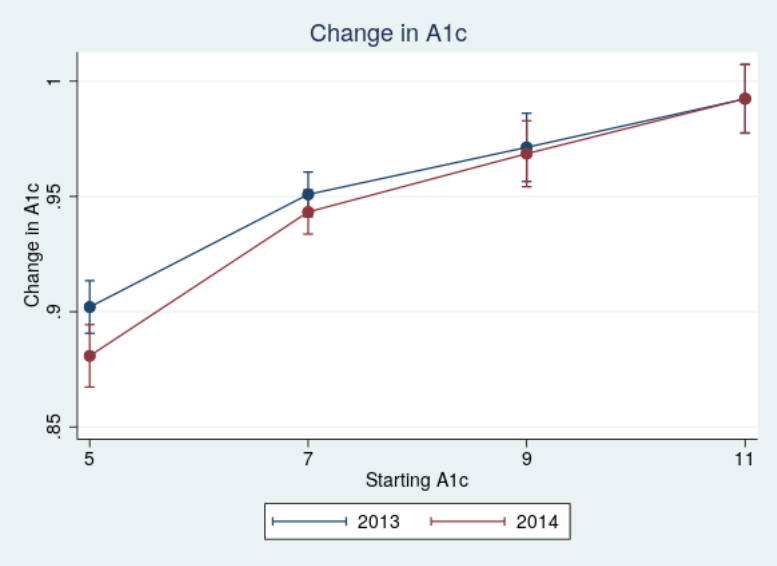

Figure 20: This graph shows change in A1c from starting A1c value (taken from Q3 2013 and Q3 2014) to the final A1c test value in the subsequent three quarters.

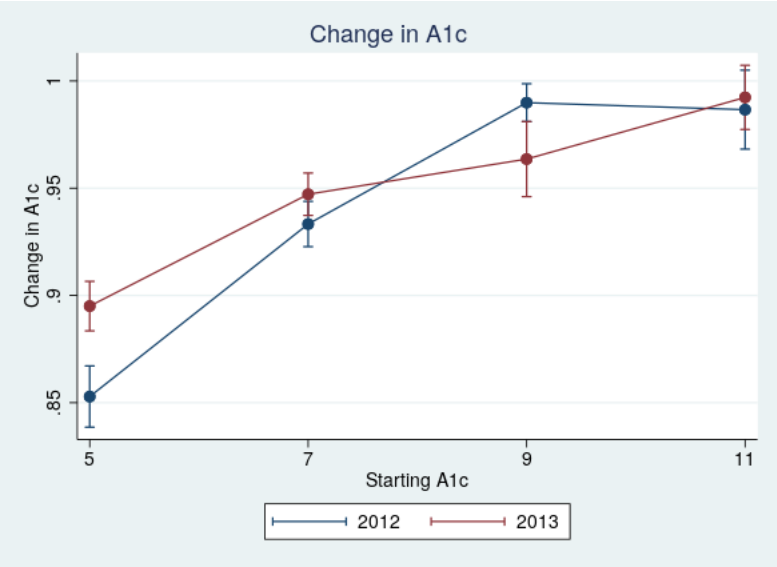

Figure 21: This graph shows change in A1c from starting A1c value (taken from Q3 2012 and Q3 2013) to the final A1c test value in the subsequent three quarters.

B Appendix: Calculation of P4Q Award

As discussed, the P4Q award is ultimately based on a physician's "success rate" for a set of quality measures, and a comparison of this rate to both national and baseline performance rates. However, the P4Q payment formula also makes adjustments with respect to the difficulty of each measure and the size and composition of the physician's practice. We now walk through an example of how a physician's P4Q quarterly bonus is calculated to further clarify the process. We give an example for 2012, where physicians are paid for their past year of performance *once per quarter* (at a rate equal to 1/4th of the yearly P4Q payment money at stake). We note that, as seen in Figure 2, in 2011 physicians only received one payment at the end of the year, for all of 2011 (so not on a rolling quarterly basis).

The award payment for each quarter, for the past year of performance, is calculated in the following steps:

- **Step A:** Calculate the maximum award for each measure
  - First, the total Maximum Payment for the quarter is calculated. Each physician has a certain number of patients attributed to his practice each quarter by HMSA based on claims data. This "primary care patient count" is multiplied by the "per member per month" fee ($4.00 per member per month 2012, $2.00 in 2011) to calculate the total Maximum Payment for the quarter.

    Here, the physician's total Maximum Payment for quarter 4 amounts to $12,928.
  - Next, the panel size for each measure is estimated. This is the number of patients in the physician's practice that are deemed "eligible" for receiving a measure within that quarter. For instance, the panel of patients eligible for the breast cancer measure consists of all women in the primary care patient count that are over 41 in age.
  - To adjust for the composition of the physician's practice, the *patient panel weight factor* for

MONTH	PRIMARY CARE PATIENT COUNT	PMPM AMOUNT	TOTAL MONTHLY POTENTIAL
October	1,095	$4.00	$4,380
November	1,070	$4.00	$4,280
December	1,067	$4.00	$4,268
Quarter 4 Subtotal	3,232		$12,928
Annual Total	**12,792**		**$51,168**

Measurement Period: 4/1/2011 to 03/31/2012
Baseline Period: 4/1/2010 to 03/31/2011

Provider: **Lee, Aloha** Est. Patient Panel Size: 1060 Est. Max Quality Pay*: **$12,720**

MEASURE	I. YOUR ESTIMATED PANEL SIZE	II. YOUR BASE YEAR NUMERATOR COUNT	III. YOUR BASE YEAR RATE	IV. YOUR NATIONAL % RANK	V. NUMBER OF ADDITIONAL PATIENTS TO ACHIEVE 90TH NATIONAL %	VI. ESTIMATED MAX AWARD	VII. ESTIMATED SHARE OF MAX QUALITY PAY
Avoidance of antibiotic treatment in adults with acute bronchitis	82	76	92.68%	90th	0	$379.29	2.98%
Breast cancer screening	371	314	84.64%	90th	0	$1,716.04	13.49%
Cervical cancer screening	399	346	86.72%	90th	0	$1,845.56	14.51%

each measure is calculated. This divides the single measure's panel size by the sum of patient panels over all measures.

For example, Dr. Lee's breast cancer screening weight is 371/2,290 = .162.

– In addition to the practice composition, the measures are also weighted with respect to their importance and effort required. These weights are constant across physicians. For example, the breast cancer importance weight is .05.

ADULT PRIMARY CARE	IMPORTANCE WEIGHT	ADJUSTMENT FACTOR
1. Avoidance of antibiotic treatment in adults with acute bronchitis	1	5.00%
2. Breast cancer screening	1	5.00%
3. Cervical cancer screening	1	5.00%
4. Cholesterol management for patients with cardiovascular conditions - LDL-C screening	1	5.00%
5. Colorectal cancer screening	1	5.00%
6. Comprehensive diabetes care - eye exam	1	5.00%
7. Comprehensive diabetes care - HbA1C testing	2	10.00%

– The importance/effort weight and patient panel weight are combined and normalized.

For example, Dr. Lee's normalized combined breast cancer weight is (.162x.05)/.06=.135.

– Finally, the Maximum Award for each measure is calculated. This is done by multiplying the

normalized weight factor for each measure by the Maximum Payment derived in step 1.

- **Step B:** Calculate Performance and Improvement Points Earned
  - First, determine the physician's prior year's national percentile ranking for each measure. This is based on the physicians "success rate" during their baseline period.

    As mentioned earlier, the "success rate" is given by the $\frac{Number of Procedures Performed}{Number of Eligible Patients in Panel}$.
  - The baseline period is currently based on a rolling year, rather than a static year. For example, for the calculation of a Quarter 1 award in 2012, performance from 4/1/2011 to 3/31/2012 will be compared to performance from 4/1/2010 to 3/31/2011. Similarly, for the calculation of a Quarter 2 award in 2012, performance will be compared from 7/1/2011 - 6/30/2012 to performance in 7/1/2010-6/30/2011.

Threshold Scale Selection

Schedule B: National Percentile Threshold Rates-Adult Primary Care Measures†

MEASURE	10TH	25TH	50TH	75TH	90TH
Avoidance of antibiotic treatment in adults with acute bronchitis	13.60	17.73	21.54	30.17	47.92
Breast cancer screening	62.73	68.52	73.04	77.00	80.98
Cervical cancer screening	70.27	72.79	77.24	79.92	86.15
Cholesterol management for patients with cardiovascular conditions - LDL-C screening	83.33	86.67	89.39	91.64	94.46

  - Depending on the percentile a physician fell into during the baseline period, the corresponding table gives the total performance and improvement points earned based on the current year's performance.

    For example, Dr. Lee's baseline and current year rate for breast cancer screening placed her within the 90th percentile, giving her 12.5 total points. Note that, for a physician who had originally performed quite poorly in 2011, and improved a lot for 2012, the matrix here for performance points would look the same as for Dr. Lee, but the improvement points would be scaled up.

Table 6: Baseline Period Performance: 90th Percentile

MEASUREMENT PERIOD PERFORMANCE	PERFORMANCE POINTS	SUSTAINED EXCELLENCE	TOTAL POINTS
10th percentile	1.5	0	1.5
25th percentile	2.5	0	2.5
50th percentile	5.0	0	5.0
75th percentile	7.5	0	7.5
90th percentile	10.0	2.5	12.5

- **Step C:** Calculate Actual Payment for Each Measure
  - To calculate how much the physician will be award for each measure, the Maximum Award for each measure is multiplied by the total points earned for each measure and divided by 10.

    For example, Dr. Lee will earn (12.5/10)x$1,716=$2,145 for breast cancer screening.
  - At that point, to determine the total physician quarterly P4Q payment, you just add the award amounts for each measure up over all the possible measures.

Overall, while this algorithm is fairly complex, the basic tenets are simple: a provider gets paid more to do recommended preventive care and chronic disease management for relevant patients, and greater payments are linked to the measures where the provider has more patients who are eligible. However, given the complexity of the underlying incentive scheme, it is unlikely that primary care physicians have full grasp of the micro level incentives.

C Appendix: Cozeva

As part of its initiative to improve population health, in 2012 HMSA began implementing a new IT platform called "Cozeva" for physicians to use in their practices. Cozeva constituted a significant upgrade over previous IT that they could use to measure their performance.[9] Prior to the implementation of Cozeva, physicians used a program called "HBI Online", implemented in 2010, which was a more difficult to access website that physicians could use to check their own performance and patient histories, often using lagged data. Cozeva, on the other hand, is a user-friendly interface that physicians can constantly have open and use in real-time to assess patient histories and needs. Specifically, among its many functions, Cozeva allows physicians to know, in real-time, what P4Q metrics a patient qualifies for, and, whether anything specific needs to be done to help them successfully achieve that metric for a given patient.

More specifically, Cozeva is multi-page, online platform that allows providers and their staff to view claims and supplemental data in real-time, and connect with patients electronically. If used actively, Cozeva has the potential to increase the physician's information set with regards to both their general level of quality care and their level of quality as specifically measured and rewarded by the P4Q program. The information provided to physicians relevant to the P4Q award includes:

- *Patient Panel*: A monthly list of patients attributed to providers by HMSA
- *Care Planning Registry*: Identifies patients who may benefit from additional care as related to P4Q metrics
- *Baseline Quality Report*: Performance measured during the baseline period and compared to national standards
- *Performance Quality Report*: Provides access to a detailed view of each measure, including National Percentile Target Rate and estimated Quality Pay-by-Percentile ranking.

In addition to these key functions, physicians are able to securely message patients, renew prescriptions, schedule appointments, and send individualized preventive care reminders, in hopes of closing gaps in care. The "Member Engagement" tools, through a system of electronic reminders and secure messaging, facilitate communication between physician and patient, and also allow physicians to collaborate with patient family members and friends. Physicians can track medication adherence by identifying prescriptions filled, view gaps in care, and display lab results when available.

Figure 22 shows the primary care physician Cozeva dashboard, which is the main page they view when logging in. The dashboard summarizes their performance on all relevant P4Q measures, highlights places where there are gaps in care, and provides the functionality to dig into performance on specific measures and issues with specific patients. Figure 23 gives an example of the page view for a physician looking in more detail at performance on a specific measure. This page shows more

[9] Cozeva was created by a company called Applied Research Works that works closely with HMSA to effectively roll out this program with physicians.

detailed information of measure performance, and illustrates how Cozeva makes it easy to target patients for whom there are gaps in care.

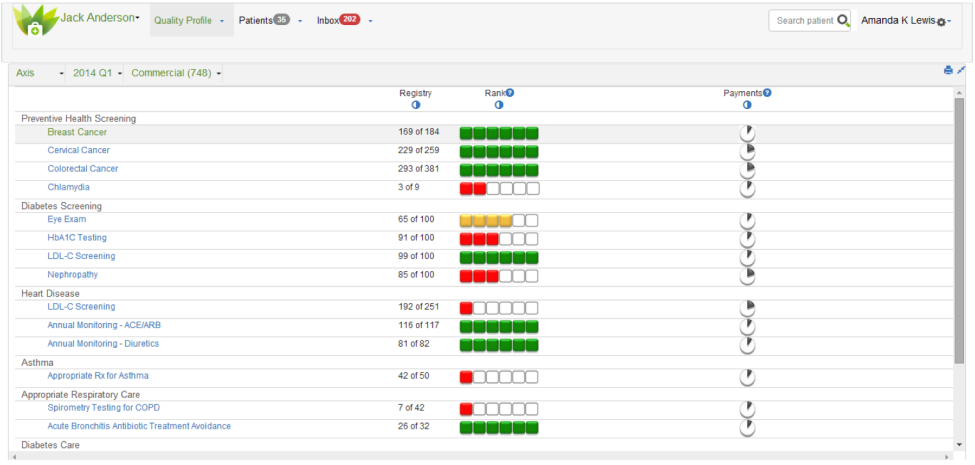

Figure 22: Dashboard view for primary care physician in Cozeva

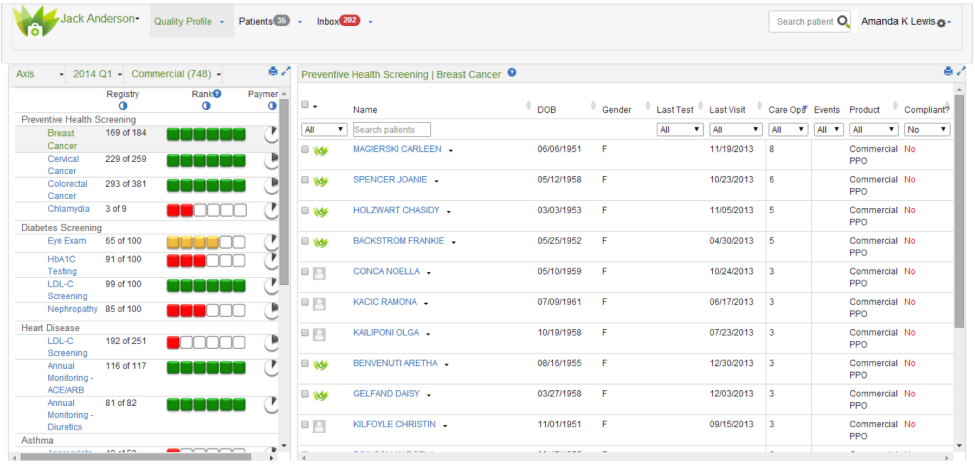

Figure 23: Specific measure view for primary care physician in Cozeva

D Appendix D: Other Results

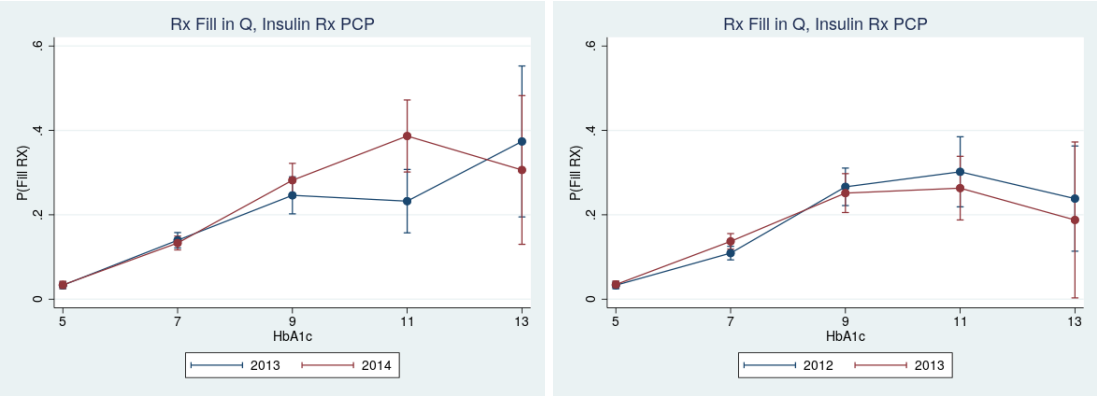

Figure 24: These figures shows the change over time in insulin prescription rates, as a function of patient HbA1c level.

Finally, due to the threshold-based nature of the pay-for-performance program at HMSA, physicians can have quite different marginal values of ensuring patients achieve the quality target for HbA1c. For example, a physician who is one patient below the 90th percentile of the national distribution for proportion of patients with appropriate A1c control has a very high marginal value for getting one more patient below the threshold. This is because he/she gets a big lump sum for passing the threshold but only has to get one more patient to below 9 to get that reward. We characterize marginal payments per closure in a quarter for each physician, and study how marginal incentives for A1c control impacts A1c levels. We find that a $100 increase in the marginal value of a closure is associated with a 2.6 percentage point increase in the fraction of patients that a physician has below the A1c threshold. This is a 20% reduction in the number of patients with A1c levels above 9. Thus, an increase in the marginal dollars at stake for additional closures has a meaningful impact of A1c outcomes.

Lastly, we investigate how physician responses vary with the marginal dollars they have at stake for improving performance. While the P4Q program implemented at HMSA included meaningful money at stake for all primary care physicians, the threshold-based nature of the reward program implies that some physicians faced larger marginal rewards to provide recommended care. For example, a physician just below the 90th percentile threshold of the national performance distribution at the end of a quarter would have received a sizeable financial benefit from treating a few additional patients to reach that threshold. Alternatively, a physician who had just passed the 75th percentile threshold near the end of a quarter would have no short-run benefit of treating a few additional patients this quarter (since he/she would have to treat many to reach the 90th percentile threshold).

There are a range of ways to define physician incentives, reflecting both static and dynamic aspects of the P4Q program and what physicians actually might reasonably perceive / understand about the program and their performance. Here, we define marginal physician payments for achieving higher quality as:

$$V_{i,m,t} = \frac{PayAmt_{k+1,i,m,t} - PayAmt_{k,i,m,t}}{N_{k+1,i,m,t} - N_{k,i,m,t}}$$

Here, $PayAmt_{k+1,i,m,t}$ is the incentive payment physician receives for hitting reward threshold $k+1$ at time t for measure m where k is threshold achieved in prior period, $t-1$. $N_{k+1,i,m,t}$ is the total

number of patient closures needed to reach threshold $k+1$ at t for m. This notion of marginal payments for quality assumes that physicians know how much money they receive for reaching the next quality threshold and know how many more patients they need to close to reach that threshold this quarter. Then, they act as if their marginal value for closing a patient at any point during the quarter is the total money they will receive that quarter for reaching the next threshold, divided by the total number of patients they need to close that quarter to reach that threshold.

We run the following probit regression relating different performance outcomes to this marginal payment value, controlling for physician demographics. Data is at the physician*quarter level, weighted by number of patients per physician, with standard errors clustered by physician.

$$y_{qi} = \beta_1 + \beta_2 MV_{qi} + Z_i + \epsilon_{qi}$$

As Table 16 shows, we study the impact of the shift to outcomes-based payments on the following outcomes: (i) the percentage of patients a physician has with HbA1c below the threshold of 9 (ii) the rate of insulin prescriptions filled for higher HbA1c patients of a physician and (iii) the rate of oral diabetes medication prescriptions filled for higher HbA1c patients of a given physician. We report the coefficients β_2, transformed to probabilities.

Relationship between Marginal Value of Closure ($100) and Performance in Outcomes-Based Period		
	Q4 2014 - Q2 2015	*Q3 2013 - Q2 2015, Pre-Period Marginal Value = 0*
Fraction Patients $\leq$ 9 A1c P4Q Data	0.0264 (0.00435)	
Fraction High A1c Patients on Insulin	-0.00266 (0.00794)	0.008 (0.00513)
Fraction High A1c Patients on Oral Rx	-0.00388 (0.00607)	0.00312 (0.00369)

Table 16: Relationship between marginal payment and outcomes & treatments for patients with diabetes. Mean marginal value of patient closure is $138 (SD= $175), median $56.

Table 16 shows that there is a positive relationship between the marginal payment a physician receives for achieving better outcomes and actually achieving those outcomes. After the shift to outcomes-based rewards, a $100 increase in the marginal value of a closure is associated with a 2.6 percentage point increase in the fraction of patients with an A1c test below 9. Given that the baseline rate of diabetic patients with A1c levels above 9 is 13% (see Table 2) this 2.6 percentage point reduction in patients with an A1c test above 9 is equivalent to a 20% reduction in patients with a poor A1c outcome (as defined by the P4Q guidelines). This suggests both that physicians do respond to the marginal payments at stake, and that they do so in a meaningful way.

Interestingly, while we find a positive relationship between payments and quality outcomes, we do not find a relationship between payments and insulin prescriptions and do not find a relationship between payments and oral diabetes medications. This suggests that, while physicians shift their habits to prescribe more insulin after the switch to outcomes-based payments, they do so without much regard for their specific marginal financial incentives in a given quarter.