Effects of Changes in Diabetes Pay-for-Performance Incentive Designs on Patient Risk Selection

*Hui-Min Hsieh, Shu-Ling Tsai, Lih-Wen Mau, and
Herng-Chia Chiu*

Objective. Taiwan's National Health Insurance (NHI) Program implemented a Diabetes Pay-for-Performance Program (P4P) based on process-of-care measures in 2001. In late 2006, that P4P program was reformed to also include achievement of intermediate health outcomes. This study examined how the change in design affected patient risk selection.

Designs/Study Populations. Study populations were identified from a 2002 to 2003 period (Phase 1) and a 2007 to 2008 period (Phase 2), spanning pre- and postimplementation of reforms in the P4P incentive design. Phase 1 had 74,529 newly enrolled P4P patients and 215,572 non-P4P patients, and Phase 2 had 76,901 newly enrolled P4P patients and 299,573 non-P4P patients. Logistic regression models were used to estimate the effect of changes in design on P4P patient selection.

Principal Findings. Patients with greater disease severity and comorbidity were more likely to be excluded from the P4P program in both phases. Furthermore, the additional financial incentive for patients' intermediate outcomes moderately worsened patient risk selection.

Conclusions. Policy makers need to carefully monitor the care of the diabetes patients with more severe and complex disease statuses after the changes of P4P financial incentive design.

Key Words. Diabetes pay-for-performance, incentive design, unintended consequences, patient risk selection

Pay-for-Performance (P4P) or value-based purchasing, a payment model in which health care providers are rewarded if their patients reach specified health goals or benchmark statuses, has been implemented in many countries with the goal of promoting quality of health care (Petersen et al. 2006; Cromwell and Smith 2011; Korda and Eldridge 2011; Blumenthal and Dixon 2012; Berenson, Pronovost, and Krumholz 2013; Ryan and Damberg 2013; White

2013). Whether pay-for-performance programs improve care by changing provider behavior is still unclear (Mehrotra, Sorbero, and Damberg 2010; Jha et al. 2012; Eijkenaar et al. 2013). One great concern regarding these incentives is that, while some providers could work harder to provide better care, others may select only healthier patients to more easily achieve these health targets and receive their rewards (Shen 2003; Doran et al. 2008; Chen et al. 2011; Chang, Lin, and Aron 2012). Such behavior has been found in a study by Shen (2003) which reported that providers treating substance abuse patients in a performance-based contracting program in Maine tended to treat less severely addicted patients. Hofer et al. (1999) also suggested that physicians treating patients with diabetes in a physician profiling program could increase their rewards by deselecting patients with histories of high medical costs, poor adherence, or poor response to treatment.

Taiwan's National Health Insurance (NHI) initiated the Diabetes P4P program near the end of 2001. Initially, the program was designed as a "pay-for-participation" program, meaning that it rewarded providers if they provided their P4P enrollees with predefined process-of-care services, for example, the provision of HbA1c or LDL tests. Near the end of 2006, that program was modified to also include "pay-for-improvement," meaning the program added financial incentives for improvements in patients' intermediate outcomes, for example, a HbA1c <7 percent (Chen et al. 2011). Two studies, one performed during the first phase of Taiwan's Diabetes P4P program and one during the second phase, have reported on physician risk selection behavior (Chen et al. 2011; Chang, Lin, and Aron 2012). The two studies did not, however, compare changes in physician patient-selecting behaviors across phases (Chen and Lai 2012). Therefore, it is not known whether changes in incentive designs in the P4P program mitigated or exacerbated the problem of patient risk selection.

To find out, we used two large nationwide population databases to investigate the effect of these programs on providers' selection of patients to enroll as P4P patients. Study populations were identified from a 2002 to 2003 period

Address correspondence to Herng-Chia Chiu, Ph.D., Department of Healthcare Administration and Medical Informatics, Kaohsiung Medical University, 100 Shin-Chuan 1st Road, Kaohsiung, Taiwan 80708; e-mail: chiu@kmu.edu.tw. Hui-Min Hsieh, Ph.D., is with the Department of Public Health, Kaohsiung Medical University, Kaohsiung, Taiwan. Shu-Ling Tsai, Ph.D., is with the National Health Insurance Administration, Ministry of Health and Welfare, Taipei, Taiwan. Lih-Wen Mau, Ph.D., is with the Patient & Health Professional Services, National Marrow Donor Organization, Edina, MN. Herng-Chia Chiu, Ph.D., is with the Department of Healthcare Administration and Medical Informatics, Kaohsiung Medical University, Kaohsiung, Taiwan.

(Phase 1) and a 2007 to 2008 period (Phase 2), spanning pre- and postimplementation reforms in incentive design of the P4P program. A diabetes-specific complexity measure and a generic measure were used to characterize disease severity of the patients in our study population. Our findings, in conjunction with two previous earlier studies, provide a more complete picture of the association between incentives using process or intermediate outcome measures and potential unintended negative effects as a result of patient risk selection in Taiwan's diabetes P4P program.

METHODS

Study Setting

Like many countries, Taiwan has undergone substantial financial stress from increasing health care expenditures. In an effort to bring these rapidly rising costs under control, Taiwan's NHI implemented a global budget policy in 2001. Meanwhile, to avoid the possible negative consequences that might arise from implementing a strict cost-containment policy, the NHI initiated several experimental P4P programs aimed at maintaining and improving the care of patients. The Diabetes P4P program that started at the end of 2001 was one of the earliest initiative P4P programs.

The program consists of several important features. First, only physicians who specialized in metabolic disorders or endocrinology or those who had participated in a training program for diabetes shared care were eligible to participate in and voluntarily enroll patients into this special P4P for diabetes care (Lee et al. 2010; Chen et al. 2011; Chen, Lee, and Kuo 2012). Second, they and their medical care staff members at the various hospitals and clinics were expected to work as a coordinated physician-led multidisciplinary team adhering to clinical guidelines established for the care of diabetes patients (Chen, Lee, and Kuo 2012). Third, all participating providers were mandated to submit follow-up reports and claims data to Taiwan's NHI for reimbursement.

The financial incentive schemes in this program were revised at one point. This created two periods marked by different quality incentive metrics designs (Lee et al. 2010; Chen et al. 2011). From 2001 to the end of 2006 (Phase 1), the P4P program was a "pay-for-participation" program, which financially rewarded P4P physicians if they provided predefined process outcomes (e.g., documented provisions of HbA1c or LDL tests) to their P4P enrollees. The provision of care included medical history assessment, physical examination, laboratory evaluation, management plan evaluation, and

diabetes self-management health education (Chen et al. 2011; Chen, Lee, and Kuo 2012; Cheng, Lee, and Chen 2012). The participating P4P physicians received an extra TWD \$450 (USD\$15) in addition to regular physician fees, TWD \$1,845 (USD\$60) per initial enrollment visit, TWD \$875 (USD\$30) per follow-up visit, and TWD \$2,245 (USD\$75) per annual evaluation visit. This quality incentive metrics design remained unchanged until the end of 2006. In Phase 2, to encourage providers to focus on clinical outcome, “pay-for-improvement” was added as extra financial incentive for improvements in patients’ intermediate outcome (e.g., an increase in the percentage of patients with controlled HbA1c). A composite score for each participating physician is calculated based on the average of these intermediate outcome indicators using equal weighting among his or her patients. Clinical outcome measures for calculating the composite scores include two negative outcomes (i.e., percentage of HbA1c >9.5 percent and percentage of LDL >130 mg/dL) and one positive outcome indicator (percentage of HbA1c <7 percent). Physicians receive quality rewards of TWD\$1,000 (USD\$30) per each enrollee who completes the annual follow-up if his or her performance ranked within the top 25 percent of the composite scores found for all participating physicians (Lee et al. 2010; Chen et al. 2011). Such a revision in P4P program provides a real-world opportunity to understand the effect of the changes in financial incentive metrics designs on the unintended negative effects.

Data Source

We conducted a cross-sectional study design using data from two population-based databases from the NHI Administration in Taiwan. One was a national diabetes P4P registry database for which we could precisely identify whether patients were newly enrolled in the P4P program. Because health care providers (both clinics and hospitals) who participated in the P4P program are required to report patient data to receive P4P follow-up payments for the care of enrolled patients, this database is complete and reliable. The other was Taiwan’s NHI administrative claims database from which we could obtain patient demographic data and comorbidities as well as health provider characteristics.

Study Population

Using the nationwide NHI claims data, we first included all type 2 diabetes patients with primary diagnoses of diabetes (ICD-9-CM codes with 250.xx or A-code 181, excluding 250.x1 or 250.x3) in at least two outpatient visits or in

at least one inpatient hospitalization for each year between 2002 and 2003 (Phase 1) and between 2007 and 2008 (Phase 2). Garnering information from the diabetes P4P registry database, we then identified newly enrolled P4P patients as intervention groups for each phase and identified non-P4P patients as comparison groups. To be included as non-P4P patients, a patient could not be enrolled in the P4P program during the patient identification periods but had to be treated by physicians eligible to participate in the P4P program. The index date for each patient was defined as the date that they were diagnosed with diabetes during patient identification periods. The same inclusion and exclusion criteria were used for both Phase 1 and Phase 2 cohorts. In both groups, we excluded patients who were not assigned to a primary physician (described below), who had type 1 diabetes because they made up less than 1 percent of total diabetes patients, who were below 18 years old at the index date, or who had certain problems related to the data available on them (e.g., gender variable was missing or data were erroneous). Because type 2 diabetes is a growing problem among the youth, we conducted a sensitivity analysis to include patients age <18. The overall findings were identical to the results when the youth diabetes population was excluded. More information about study inclusion and exclusion criteria can be seen in Figure 1.

To address the issue of patients with multiple outpatient visits to different hospitals or providers accountably, we used a plurality provider algorithm to assign a non-P4P patient to the most frequently seen physician (or practice), which we determined based on the physician billing for the greatest number of care visits during the identification period in each phase (Pham et al. 2007; Chen et al. 2011). We directly assigned P4P patients to their frequently seen P4P enrolling physician. In the end, we studied 1,025 eligible physicians with 74,529 newly enrolled P4P patients and 215,572 non-P4P patients for Phase 1. For Phase 2, we studied 1,698 eligible physicians with 76,901 newly enrolled P4P patients and 299,573 non-P4P patients. We further examined the characteristics for eligible participating physicians between Phase 1 and 2, and found no statistically significant difference among physicians in both phases. More details regarding inclusion and subject characteristics can be found in the supplementary Table S1.

Variable Definitions

The major outcome of interest was the effect of the change in policy on that a patient would be excluded from the P4P program. We also included a dummy variable to capture the effect of the policy, with Phase 1 as the reference group.

Figure 1: Flowchart of Inclusion and Exclusion Criteria for P4P and Non-P4P Patients in Phase 1 and Phase 2

We used two ICD-9-CM coding adaptations to capture patient disease severity level during the 1 year prior to index date. One indicator was the diabetes complication severity index (DCSI), which has been commonly used to

capture diabetes severity in the literature (Young et al. 2008). The DCSI score takes into consideration seven categories of complications: cardiovascular complications, nephropathy, retinopathy, peripheral vascular disease, stroke, neuropathy, and metabolic disorders. DCSI has a total score of 13 points; the higher the score, the more severe the disease state. In both phases, DCSI was categorized into four categories (i.e., a score 0, 1, 2 or ≥ 3) based on the quartiles of DCSI scores in Phase 1. The other indicator of severity we used adapted from the ICD-9-CM code was the Deyo–Charlson Comorbidity Index (CCI), a measurement of a patient's overall clinical condition by classifying or weighting comorbid conditions (comorbidities). This index has also been widely utilized by health researchers to measure burden of disease and case mix, with low scores representing lowest risk (Deyo, Cherkin, and Ciol 1992; Halanych et al. 2007; Quan et al. 2011). Similarly, CCI was also categorized into four categories (i.e., a score 0–1, 2, 3, or ≥ 4) based on the quartiles in Phase 1.

Patients' demographic characteristics we included were age and gender. The physician characteristics we included for our analysis were physician specialty (i.e., metabolism and endocrinology, nephrology, cardiology, internal medicine, family medicine, and others) and volume of diabetes patients they treated per year. Physician service volume was categorized into low, median, and high volume based on the thresholds of 33th and 67th percentile of numbers of diabetes in Phase 1. Health care institution characteristics included accreditation level (medical center, regional hospital, local hospital, or clinic), ownership type (public, not-for-profit, or for-profit), teaching status, and geographic location (Taipei, northern, central, southern, Kao-Ping, or eastern area).

Statistical Analysis

To control for variations between two phases from patient characteristics and health care provider factors, we pooled the two periods of data and used patient-level logistic regression models to estimate the probability of a patient being excluded. A heteroskedasticity-robust standard error adjustment was used, and two-way clustering of standard errors was also used for clustering patients within physicians and health care institutions to control for unequal error variances across physicians and health institutions. In addition, to put the magnitude of these responses into perspective, we examined the patient severities in each phase to determine probability of patient risk selection for that phase. We generated the adjusted predicted probabilities (APPs) and

marginal effects (MEs) of the probabilities (Williams 2012). APPs and MEs were estimated and adjusted by patient demographic, clinical characteristics, and health care institutional characteristics in the logistic regressions. Bootstrapping with 300 times replications with sample size equivalent to the original was used to obtain 95 percent confidence interval for the APPs and MEs (Weintraub and Cohen 2009; Cromwell and Smith 2011). All statistical operations were performed using SAS version 9.3 (SAS institute, Cary, NC, USA) and Stata version SE 12. A p -value $<.05$ was considered significant.

RESULTS

Table 1 is a summary of baseline patient and health care provider characteristics for the two phases. In Phase 1, P4P patients were 47.34 percent male, had a mean age of 59.40 years, a mean DCSI of 1.00, and a mean CCI index of 2.18. Non-P4P patients in Phase 1 were 49.84 percent male and had a mean age of 59.87 years, a mean DCSI score of 1.02, and a mean CCI index of 2.27. The majority of the P4P and non-P4P patients in Phase 1 were cared for by physicians who were treating a high volume of diabetes and whose professional specialties were internal medicine or metabolism-endocrinology. In Phase 2, P4P patients were slightly younger and less sick than P4P patients in Phase 1.

Table 2 shows the results of two logistic regression models. Model 1 included the DCSI variable and Model 2 the CCI to determine whether patients with more severe diabetes were more likely to be excluded from P4P programs by health care providers. Both models showed consistently that the odds of being excluded were greater in Phase 2 than in Phase 1 (OR = 1.344 and OR = 1.347, $p < .001$). In general, patients who were male, older than 75 years, and those with DCSI scores or CCI index scores larger than 2 were more likely to be excluded from enrollment ($p < .001$). Nevertheless, patient exclusions were less likely to occur among patients being treated by participating physicians whose professional specialties were metabolism and endocrinology ($p < .05$) and those treating a high volume of diabetes patients per year ($p < .01$), those serving in health care institutions that were not-for-profit facilities ($p < .01$) and regional hospitals, local hospitals, or clinics ($p < .001$).

Table 3 summarizes the results of our analysis of mean adjusted probability that patients with different levels of severity would be excluded and also the ME of those predicted probabilities in Phase 1 and 2. In Phase 1, the APPs of patients being excluded the severity levels were accessed for DCSI at 0, 1, 2, and ≥ 3 were 73.82, 74.14, 74.84, and 77.78 percent, respectively. In Phase 2,

Table 1: Descriptive Results of Baseline Characteristics for P4P and Non-P4P in Phase 1 and Phase 2

	Phase 1			Phase 2		
	P4P	Non-P4P	p-value	P4P	Non-P4P	p-value
<i>Patients' demographic characteristics</i>						
N	74,529	215,572		76,901	299,573	
<i>Gender</i>						
Male (N, %)	35,282 (47.34)	107,450 (49.84)	<.001	38,478 (50.04)	154,145 (51.45)	<.001
Female (N, %)	39,247 (52.66)	108,122 (50.16)		38,423 (49.96)	145,428 (48.55)	
Age years (mean ± SD)	59.40 (± 11.65)	59.87 (± 13.15)	<.001	58.64 (± 12.33)	60.64 (± 13.48)	<.001
<i>Age categories (N, %)</i>						
<45	8,498 (11.40)	28,903 (13.41)	<.001	10,142 (13.19)	36,762 (12.27)	<.001
45–54	18,291 (24.54)	50,615 (23.48)		19,592 (25.48)	67,306 (22.47)	
55–64	21,962 (29.47)	54,348 (25.21)		22,462 (29.21)	78,714 (26.28)	
65–74	19,598 (26.30)	53,831 (24.97)		17,131 (22.28)	67,664 (22.59)	
75+	6,180 (8.29)	27,875 (12.93)		7,574 (9.85)	49,127 (16.40)	
<i>Patients' severity characteristics</i>						
DCSI score (mean ± SD)	1.00 (± 1.26)	1.02 (± 1.36)	<.001	.76 (± 1.17)	.98 (± 1.36)	<.001
<i>DCSI categories (N, %)</i>						
1st quartile (DCSI: 0)	35,028 (47.00)	105,314 (48.85)	<.001	45,265 (58.86)	154,344 (51.52)	<.001
2nd quartile (DCSI: 1)	19,533 (26.21)	51,440 (23.86)		16,095 (20.93)	67,204 (22.43)	
3rd quartile (DCSI: 2)	11,278 (15.13)	31,061 (14.41)		9,167 (11.92)	41,351 (13.80)	
4th quartile (DCSI: ≥3)	8,690 (11.66)	27,757 (12.88)		6,374 (8.29)	36,674 (12.24)	
CCI index (mean ± SD)	2.18 (± 1.61)	2.27 (± 1.70)	<.001	1.99 (± 1.46)	2.21 (± 1.64)	<.001
<i>CCI categories (N, %)</i>						
1st quartile (CCI: 0–1)	31,106 (41.74)	87,760 (40.71)	<.001	37,371 (48.60)	129,866 (43.35)	<.001
2nd quartile (CCI: 2)	17,283 (23.19)	50,458 (23.41)		18,716 (24.34)	70,484 (23.53)	
3rd quartile (CCI: 3)	12,707 (17.05)	35,645 (16.54)		10,772 (14.01)	47,645 (15.90)	
4th quartile (CCI: ≥4)	13,433 (18.02)	41,709 (19.35)		10,042 (13.06)	51,578 (17.22)	

Continued

Table 1. Continued

Phase 1			Phase 2			
	P4P	Non-P4P	p-value	P4P	Non-P4P	p-value
Health care providers' characteristics						
Physician level characteristics						
Physician specialty (N, %)						
Metabolism and endocrinology	15,745 (21.13)	49,125 (22.79)	<.001	19,887 (25.86)	67,465 (22.52)	<.001
Nephrology	4,561 (6.12)	10,414 (4.83)		2,802 (3.64)	11,562 (3.86)	
Cardiology	3,269 (4.39)	5,402 (2.51)		1,721 (2.24)	8,683 (2.90)	
Internal medicine	37,401 (50.18)	114,336 (53.04)		38,361 (49.88)	154,226 (51.48)	
Family medicine	11,118 (14.92)	28,313 (13.13)		10,422 (13.55)	44,112 (14.72)	
Others	2,435 (3.27)	7,982 (3.70)		3,708 (4.82)	13,525 (4.51)	
Categories of physician volumes for caring diabetes patients (N, %) [†]						
Low volume	3,320 (4.45)	9,994 (4.64)	<.001	3,660 (4.76)	14,031 (4.68)	<.001
Median volume	11,815 (15.85)	36,379 (16.88)		13,769 (17.90)	54,034 (18.04)	
High volume	59,394 (79.69)	169,199 (78.49)		59,472 (77.34)	231,508 (77.28)	
Health care institution level characteristics						
Accreditation level (N, %)						
Medical center	12,339 (16.56)	66,118 (30.67)	<.001	12,650 (16.45)	91,688 (30.61)	<.001
Regional hospital	33,983 (45.60)	77,632 (36.01)		26,701 (34.72)	98,834 (32.99)	
Local hospital	17,481 (23.46)	38,006 (17.63)		15,376 (19.99)	55,632 (18.57)	
Clinics	10,726 (14.39)	33,816 (15.69)		22,174 (28.83)	53,419 (17.83)	
Ownership type (N, %)						
Public	19,244 (25.82)	69,798 (32.38)	<.001	17,459 (22.70)	91,127 (30.42)	<.001
Not-for-profit	28,173 (37.80)	88,137 (40.89)		30,336 (39.45)	123,066 (41.08)	
For-profit	27,112 (36.38)	57,637 (26.74)		29,106 (37.85)	85,380 (28.50)	

Continued

Table 1. Continued

	Phase 1			Phase 2		
	P4P	Non-P4P	p-value	P4P	Non-P4P	p-value
Location (N, %)						
Taipei	18,237 (24.47)	70,921 (32.90)	<.001	25,721 (33.45)	98,331 (32.82)	<.001
Northern	10,315 (13.84)	29,544 (13.70)		7,491 (9.74)	48,307 (16.13)	
Central	24,936 (33.46)	42,955 (19.93)		19,164 (24.92)	55,256 (18.44)	
Southern	9,093 (12.20)	31,858 (14.78)		10,385 (13.50)	47,205 (15.76)	
Kao-Ping	10,224 (13.72)	33,564 (15.57)		11,768 (15.30)	42,357 (14.14)	
Eastern	1,724 (2.31)	6,730 (3.12)		2,372 (3.08)	8,117 (2.71)	

p-value here was to compare patient demographic and clinical characteristics between P4P and Non-P4P patients in Phase 1 and Phase 2; chi-square tests were used for categorical variables and *t*-tests were used for continuous variables.
*Physician volume of diabetes patients was categorized into low, median, and high volume using 2002–2003 mean volume of 33 percentile (164) and 67 percentile (407) as thresholds.
CCI, Charlson comorbidity index; DCSI, diabetes complications severity index.

Table 2: Factors Associated with Patient Selection Logistic Regression Model Results of Being Excluded by Different Severity Measures

	Model 1 OR (95% CI)	Model 2 OR (95% CI)
<i>Main policy effect</i>		
P4P period (Ref = Phase 1)		
Phase 2	1.344*** (1.151, 1.569)	1.347*** (1.153, 1.572)
<i>Patients' severity</i>		
DCSI categories (Ref. group: 1st quartile (DCSI: 0))		
2nd quartile (DCSI: 1)	1.017 (0.975, 1.062)	—
3rd quartile (DCSI: 2)	1.058** (1.013, 1.104)	—
4th quartile (DCSI: ≥3)	1.255*** (1.186, 1.329)	
CCI index (Ref. group: 1st quartile (CCI: 0–1))		
2nd quartile (CCI: 2)	—	1.038* (1.002, 1.075)
3rd quartile (CCI: 3)	—	1.079** (1.030, 1.131)
4th quartile (CCI: ≥4)		1.192*** (1.123, 1.265)
<i>Patients' demographic characteristics</i>		
Gender (Ref. = Female)		
Male	1.070*** (1.048, 1.091)	1.067*** (1.045, 1.089)
Age (Ref. <45)		
45–54	0.877*** (0.850, 0.904)	0.876*** (0.849, 0.903)
55–64	0.829*** (0.795, 0.864)	0.829*** (0.794, 0.865)
65–74	0.911*** (0.866, 0.958)	0.912** (0.866, 0.961)
75+	1.454*** (1.371, 1.542)	1.460*** (1.375, 1.549)
Health care providers' characteristics		
<i>Physician level characteristics</i>		
Physician specialty (Ref. = others)		
Metabolism and endocrinology	0.630* (0.429, 0.927)	0.628* (0.427, 0.924)
Nephrology	0.775 (0.475, 1.265)	0.775 (0.475, 1.266)
Cardiology	0.908 (0.570, 1.448)	0.911 (0.572, 1.452)
Internal medicine	0.918 (0.681, 1.239)	0.914 (0.677, 1.234)
Family medicine	1.064 (0.770, 1.470)	1.059 (0.766, 1.464)
Categories of physician volumes (Ref. = low volume)		
Median volume	0.873 (0.690, 1.104)	0.872 (0.689, 1.104)
High volume	0.645** (0.472, 0.882)	0.644** (0.471, 0.882)
<i>Health care institution level characteristics</i>		
Accreditation level (Ref. = Medical center)		
Regional hospital	0.390*** (0.289, 0.526)	0.389*** (0.289, 0.524)
Local hospital	0.358*** (0.262, 0.490)	0.357*** (0.261, 0.488)
Clinics	0.259*** (0.176, 0.382)	0.258*** (0.175, 0.379)
Ownership type (Ref. = Public)		
Not-for-profit	0.687** (0.520, 0.906)	0.688** (0.522, 0.907)
For-profit	0.809 (0.641, 1.021)	0.812 (0.643, 1.025)
Location (Ref. = Eastern)		
Taipei	1.072 (0.602, 1.909)	1.066 (0.598, 1.900)
Northern	1.352 (0.720, 2.539)	1.346 (0.716, 2.530)

Continued

Table 2. Continued

	Model 1 OR (95% CI)	Model 2 OR (95% CI)
Central	0.561 (0.312, 1.010)	0.561 (0.312, 1.010)
Southern	1.230 (0.683, 2.216)	1.225 (0.680, 2.209)
Kao-Ping	0.973 (0.525, 1.803)	0.970 (0.523, 1.797)

Notes. Model 1 was the logistic regression model that controlled for patients' severity using DCSI variables specifically along with patients' baseline demographic and health care institutional characteristics, while Model 2 controlled for patients' severity using CCI variables specifically.
** $p < .01$, *** $p < .001$.
CCI, Charlson comorbidity index; DCSI, diabetes complications severity index; Ref. = reference group.

they were 78.89, 79.16, 79.77, and 82.29 percent, respectively. The MEs of the predicted probabilities for DCSI in Phase 1 were 0.31 percent at DCSI = 1 ($p < .05$), 1.02 percent at DCSI = 2, and 3.96 percent at DCSI ≥ 3 , compared with DCSI = 0 ($p < .001$), indicating the higher disease severity, the greater likelihood of exclusion. The MEs of the predicted probabilities for DCSI in Phase 2 were slightly lower than Phase 1, which were 0.27 percent at DCSI = 1 ($p < .001$), 0.88 percent at DCSI = 2 ($p < .001$), and 3.40 percent at DCSI ≥ 3 when compared to DCSI = 0. Similar patterns were found for the predicted probabilities of patients being excluded when analyzing data by CCI scores at 0–1, 2, 3, and ≥ 4 .

DISCUSSION

This study examined the effect of two different financial incentive designs used by a P4P program in Taiwan on unintended negative effects. One diabetes-specific complexity measure (i.e., DCSI) and one generic measure (i.e., CCI) were used to characterize disease severity of the patients in our study population. We then analyzed the risk that patients at different disease severity levels would be deselected. Consistent with prior studies (Chen et al. 2011; Chang, Lin, and Aron 2012), we found that health care providers responded to outcome metrics design with observable measurable differences in patient risk selection. Patients with more severe diabetes and higher levels of comorbidity were more likely to be excluded from the P4P program. In addition, patients who were older age (75+) were more likely to be excluded from the P4P program in both phases. The incentive design with additional financial incentive for patients' intermediate outcomes worsened the magnitude of patient risk selection.

Table 3: Results of Adjusted Predicted Probabilities and Marginal Effects of the Association between Patient Severity and Probability of Patients Being Excluded in Phase 1 and 2

	Phase 1 Adjusted Predicted Probability (95% CI)	Phase 2 Adjusted Predicted Probability (95% CI)	Marginal Effect (ME) in Phase 1 (95% CI)	Marginal Effect (ME) in Phase 2 (95% CI)
DCSI				
1st quartile (DCSI: 0)	73.82 (73.64, 74.01)	78.89 (78.74, 79.04)	—	—
2nd quartile (DCSI: 1)	74.14 (73.88, 74.39)	79.16 (78.94, 79.38)	0.31* (0.03, 0.59)	0.27* (0.03, 0.51)
3rd quartile (DCSI: 2)	74.84 (74.53, 75.15)	79.77 (79.50, 80.04)	1.02*** (0.69, 1.34)	0.88*** (0.60, 1.16)
4th quartile (DCSI: ≥3)	77.78 (77.46, 78.11)	82.29 (82.01, 82.57)	3.96*** (3.60, 4.32)	3.40*** (3.09, 3.71)
CCI				
1st quartile (CCI: 0–1)	73.56 (73.38, 73.75)	78.70 (78.54, 78.85)	—	—
2nd quartile (CCI: 2)	74.24 (74.01, 74.47)	79.28 (79.09, 79.48)	0.68*** (0.43, 0.92)	0.59*** (0.37, 0.80)
3rd quartile (CCI: 3)	74.94 (74.65, 75.24)	79.89 (79.65, 80.13)	1.38*** (1.08, 1.68)	1.19*** (0.93, 1.45)
4th quartile (CCI: ≥4)	76.68 (76.41, 76.95)	81.38 (81.14, 81.61)	3.11*** (2.82, 3.41)	2.68*** (2.43, 2.93)

Note. Bootstrapping with 300 replications with sample size equivalent to the original was used to obtain 95% confidence interval for the adjusted predicted probability and marginal effects. Covariates controlled were listed in the Table 2 Model 1 and Model 2. Bootstrapping confidence intervals were in the parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$.
CCI, Charlson comorbidity index; DCSI, diabetes complications severity index.

Ideally, when predefined process measures are rewarded in P4P programs, physicians are encouraged to order essential tests for their patients. These process measures of quality are appropriate for all diabetes regardless of disease severity, comorbidity, and critical to managing their conditions. It is possible that disenrollment practices under such a program would mainly be aimed at not selecting “noncompliant” or “unwilling to attend” patients (McDonald and Roland 2009; Gucciardi et al. 2011; Chang, Lin, and Aron 2012). However, when intermediate outcome measures such as the achievement of target blood glucose and lipid levels are used as a basis of reward in a P4P program, the results depend to a greater degree on the patients’ condition. In such circumstances, clinicians tend to focus their efforts on patients who can more easily meet targets (Shen 2003; Petersen et al. 2006; Chen et al. 2011; Chang, Lin, and Aron 2012; Eijkenaar 2013; Eijkenaar et al. 2013). Therefore, outcome incentives may exacerbate the potential for “gaming” under such P4P incentive schemes.

Previous studies have addressed several common approaches used to prevent adverse selection. For example, in one P4P program in the United Kingdom, physicians were not allowed to select the patients to include in the program, but they could use a range of criteria to exclude individual patients from the quality calculations that determine their pay, which is also called “exception reporting” (Doran et al. 2008; Gravelle, Sutton, and Ma 2010). However, insufficient funds from the NHI in Taiwan make it difficult to mandate participation and cover the additional incentives for quality improvements for all diabetes patients (Chen et al. 2011). Another approach to preventing adverse selection is to adjust patient risk statistically for outcome measures (Milgate and Cheng 2006; Rosenthal and Frank 2006; Chen et al. 2011). Appropriate risk adjustment is important and even more critical when dollars are attached to those outcome comparisons. To achieve fair comparisons, the burden of data collection must be balanced with the level of accuracy of the risk adjustment (Milgate and Cheng 2006; Chen and Lai 2012). In addition, appropriate risk adjustment must take into account generic- or disease-specific illness indexes are being used as risk adjustment indicators. In the present study, both disease-specific index (DCSI) as well as generic illness index (CCI) sensitively predicted the likelihood of patient risk selection, so the predictive ability of these two indices might make them useful in monitoring and preventing potential risk selection. Nonetheless, these approaches are relatively passive approaches to eliminating potential selection for more complex patients. In other words, physicians would be encouraged to enroll

patients into the program and then these post hoc approaches can be used to access and reduce the patient risk selection.

P4P incentives could be designed much proactively to reward providers for including patients with multiple chronic conditions and encourage diabetes care teams to include patients who might require more effort and time to achieve targeted outcomes (Pogach et al. 2010; Chang, Lin, and Aron 2012). P4P incentives can also be designed to promote provider integration between both primary care physician and specialists to improve the care of diabetic patients (Mechanic and Altman 2009). Primary care physicians can provide the initial diagnosis, treatment in less complex cases, and ensure overall continuity of care, while specialists can be more involved in the care of more complex cases that do not respond to first-(and second-) line treatment (Katon et al. 2001; O'Donnell et al. 2013). Therefore, the opportunity exists to enhance practitioner roles in diabetes care and to promote an integrated care health system under the P4P system (Katon et al. 2001; O'Donnell et al. 2013). Specifically, our study observed that physicians with specialties in metabolism/endocrinology and those with higher volume were less likely to exclude patients from the program (more detailed information can be found in the supplementary Table S2). We also found that patients treated in health care institutions other than medical centers were more likely to be enrolled in the program. These results also show this system generally operated like an "integrated care" care model, in which the routine management of uncomplicated type 2 diabetes tends to shift to primary care while the care of those with complicated type 2 diabetes is managed by the secondary care (Greenfield et al. 1995; Katon et al. 2001; O'Donnell et al. 2013). Based on our findings, Taiwan's Diabetes P4P program may want to further revise its incentives matrix to reward the inclusion of patients with more complex diseases and also design incentives to promote integrated care under a P4P program.

This study has several limitations. First, we conducted a cross-sectional design in two phases. Patients' characteristics in two phases may vary. To be able to compare two phases, we pooled two periods of data together and adjusted for patients' baseline demographic and clinical comorbidities and health providers' characteristics to statistically minimize potential unobservable confounding effects. Second, our study only compared the probability of not being enrolled in the program between newly enrolled P4P and non-P4P patients under the care by eligible P4P physicians at two cross-sectional time periods. The number of existing enrolled P4P patients in Phase 2 may have been greater than Phase 1, because in Phase 1, the P4P program had just started. In Phase 2, the P4P program had been already implemented for nearly

5 years. The number of existing enrolled P4P patients may affect the overall capacity of the services physicians provide for both newly enrolled P4P and non-P4P patients. However, this study lacked certain data on the hospital/clinics (e.g., bed size, numbers of full time employees or nurse case managers) which would have allowed us to capture available service capacity among health care providers. Third, because P4P incentive designs differ among countries and because our study is of diabetes patients only, our results may not be generalized to other P4P programs in other countries.

In conclusion, our analysis of over 10 years' experience with P4P based on incentives for "pay-for-participation" to "pay-for-improvement" in Taiwan found an association between disease severity and comorbidity and the likelihood of exclusion from the P4P program. We found an association between incentive design-based intermediate outcomes and moderately worsened unintended negative effects as a result of patient risk selection. In light of these findings, policy makers in Taiwan may want to carefully monitor the care of the patients with more severe and complex disease states after such changes of P4P financial incentive designs.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: Our manuscript has not been submitted nor has it been published in whole or in part elsewhere. This project was funded by DOH Grant (DOH101-NH-9018): "Evaluation of Pay-for-Performance and Establishment of Prospective Payment for Integrated Care," awarded by the National Health Insurance Administration of the Ministry of Health and Welfare in Taiwan. We declare that none of the authors have a conflict of interest with regard to this manuscript and all authors are willing to take public responsibility for the work and satisfy the requirements of authorship.

Disclosures: None.

Disclaimers: None.

REFERENCES

- Berenson, R., P. J. Pronovost, and H. M. Krumholz. 2013. "Achieving the Potential of Health Care Performance Measures." *Timely Analysis of Immediate Health Policy Issues* [accessed on March 3, 2014]. Available at http://www.urban.org/UploadedPDF/412823-Achieving-the-Potential-of-Health-Care-Performance-Measures.pdf?RSSFeed=UI_HealthPolicyCenter.xml

- Blumenthal, D., and J. Dixon. 2012. "Health-Care Reforms in the USA and England: Areas for Useful Learning." *Lancet* 380 (9850): 1352–7.
- Chang, R. E., S. P. Lin, and D. C. Aron. 2012. "A Pay-for-Performance Program in Taiwan Improved Care for Some Diabetes Patients, But Doctors May Have Excluded Sicker Ones." *Health Affairs (Millwood)* 31 (1): 93–102.
- Chen, T. T., and M. S. Lai. 2012. "The Unintended Consequences of Pay-for-Performance." *Health Affairs (Millwood)* 31 (5): 1127; author reply 28.
- Chen, P. C., Y. C. Lee, and R. N. Kuo. 2012. "Differences in Patient Reports on the Quality of Care in a Diabetes Pay-for-Performance Program between 1 year Enrolled and Newly Enrolled Patients." *International Journal for Quality in Health Care* 24 (2): 189–96.
- Chen, T. T., K. P. Chung, I. Lin, and M. S. Lai. 2011. "The Unintended Consequence of Diabetes Mellitus Pay-for-Performance (P4P) Program in Taiwan: Are Patients with More Comorbidities or More Severe Conditions Likely to Be Excluded from the P4P Program?" *Health Services Research* 46 (1 Pt 1): 47–60.
- Cheng, S. H., T. T. Lee, and C. C. Chen. "A Longitudinal Examination of a Pay-for-Performance Program for Diabetes Care: Evidence from a Natural Experiment." *Medical Care* 50 (2): 109.
- Cromwell, J., and K. W. Smith. 2011. "Evaluating Pay for Performance Interventions." In *Pay for Performance in Health Care: Methods and Approaches*, edited by J. Cromwell, M. G. Trisolini, G. C. Pope, J. B. Mitchell, and L. M. Greenwald. No. BK-0002-1103. Research Triangle Park, NC: RTI Press [accessed on May 15, 2013]. Available at <http://www.rti.org/rtipress>
- Deyo, R. A., D. C. Cherkin, and M. A. Ciol. 1992. "Adapting a Clinical Comorbidity Index for Use with ICD-9-CM Administrative Databases." *Journal of Clinical Epidemiology* 45 (6): 613–9.
- Doran, T., C. Fullwood, D. Reeves, H. Gravelle, and M. Roland. 2008. "Exclusion of Patients from Pay-for-Performance Targets by English Physicians." *New England Journal of Medicine* 359 (3): 274–84.
- Eijkenaar, F. 2013. "Key Issues in the Design of Pay for Performance Programs." *The European Journal of Health Economics* 14 (1): 117–31.
- Eijkenaar, F., M. Emmert, M. Scheppach, and O. Schoffski. 2013. "Effects of Pay for Performance in Health Care: A Systematic Review of Systematic Reviews." *Health Policy* 110 (2–3): 115–30.
- Gravelle, H., M. Sutton, and A. Ma. 2010. "Doctor for Behaviour under a Pay for Performance Contract: Treating, Cheating, and Case Finding?" *The Economic Journal* 120 (542): F129–56.
- Greenfield, S., W. Rogers, M. Mangotich, M. F. Carney, and A. R. Tarlov. 1995. "Outcomes of Patients with Hypertension and Non-Insulin Dependent Diabetes Mellitus Treated by Different Systems and Specialties: Results from the Medical Outcomes Study." *Journal of the American Medical Association* 274 (18): 1436–44.
- Gucciardi, E., V. W.-S. Chan, M. Fortugno, S. Khan, S. Horodezny, and S. J. Swartzack. 2011. "Primary Care Physician Referral Patterns to Diabetes Education Programs in Southern Ontario, Canada." *Canadian Journal of Diabetes* 35 (3): 262–8.

- Halanych, J. H., M. M. Safford, W. C. Keys, S. D. Person, J. M. Shikany, Y. I. Kim, R. M. Centor, and J. J. Allison. 2007. "Burden of Comorbid Medical Conditions and Quality of Diabetes Care." *Diabetes Care* 30 (12): 2999–3004.
- Hofer, T. P., R. A. Hayward, S. Greenfield, E. H. Wagner, S. H. Kaplan, and W. G. Manning. 1999. "The Unreliability of Individual Physician "Report Cards" for Assessing the Costs and Quality of Care of a Chronic Disease." *Journal of the American Medical Association* 281 (22): 2098–105.
- Jha, A. K., K. E. Joynt, E. J. Orav, and A. M. Epstein. 2012. "The Long-Term Effect of Premier Pay for Performance on Patient Outcomes." *New England Journal of Medicine* 366 (17): 1606–15.
- Katon, W., M. Von Korff, E. Lin, and G. Simon. 2001. "Rethinking Practitioner Roles in Chronic Illness: The Specialist, Primary Care Physician, and the Practice Nurse." *General Hospital Psychiatry* 23 (3): 138–44.
- Korda, H., and G. N. Eldridge. 2011. "Payment Incentives and Integrated Care Delivery: Levers for Health System Reform and Cost Containment." *Inquiry* 48 (4): 277–87.
- Lee, T. T., S. H. Cheng, C. C. Chen, and M. S. Lai. 2010. "A Pay-for-Performance Program for Diabetes Care in Taiwan: A Preliminary Assessment." *The American Journal of Managed Care* 16 (1): 65–9.
- McDonald, R., and M. Roland. 2009. "Pay for Performance in Primary Care in England and California: Comparison of Unintended Consequences." *The Annals of Family Medicine* 7 (2): 121–7.
- Mechanic, R. E., and S. H. Altman. 2009. "Payment Reform Options: Episode Payment is a Good Place to Start." *Health Affairs (Millwood)* 28 (2): w262–71.
- Mehrotra, A., M. E. Sorbero, and C. L. Damberg. 2010. "Using the Lessons of Behavioral Economics to Design More Effective Pay-for-Performance Programs." *The American Journal of Managed Care* 16 (7): 497–503.
- Milgate, K., and S. B. Cheng. 2006. "Pay-for-Performance: The MedPAC Perspective." *Health Affairs (Millwood)* 25 (2): 413–9.
- O'Donnell, M., A. de Siun, M. O'Mullane, D. Smith, C. Bradley, F. M. Finucane, and S. F. Dinneen. 2013. "Differences in the Structure of Outpatient Diabetes Care between Endocrinologist-Led and General Physician-Led Services." *BMC Health Services Research* 13: 493.
- Petersen, L. A., L. D. Woodard, T. Urech, C. Daw, and S. Sookanan. 2006. "Does Pay-for-Performance Improve the Quality of Health Care?" *Annals of Internal Medicine* 145 (4): 265–72.
- Pham, H. H., D. Schrag, A. S. O'Malley, B. Wu, and P. B. Bach. 2007. "Care Patterns in Medicare and Their Implications for Pay for Performance." *New England Journal of Medicine* 356 (11): 1130–9.
- Pogach, L. M., M. Rajan, M. Maney, C. L. Tseng, and D. C. Aron. 2010. "Hidden Complexities in Assessment of Glycemic Outcomes: Are Quality Rankings Aligned with Treatment?" *Diabetes Care* 33 (10): 2133–9.
- Quan, H., B. Li, C. M. Couris, K. Fushimi, P. Graham, P. Hider, J. M. Januel, and V. Sundararajan. 2011. "Updating and Validating the Charlson Comorbidity Index

- and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data from 6 Countries." *American Journal of Epidemiology* 173 (6): 676–82.
- Rosenthal, M. B., and R. G. Frank. 2006. "What is the Empirical Basis for Paying for Quality in Health Care?" *Medical Care Research and Review: MCRR* 63 (2): 135–57.
- Ryan, A. M., and C. L. Damberg. 2013. "What Can the Past of Pay-for-Performance Tell Us about the Future of Value-Based Purchasing in Medicare?" *Healthcare* 1 (1–2): 42–9.
- Shen, Y. 2003. "Selection Incentives in A Performance-Based Contracting System." *Health Services Research* 38 (2): 535–52.
- Weintraub, W. S., and D. J. Cohen. 2009. "The Limits of Cost-Effectiveness Analysis." *Circulation: Cardiovascular Quality and Outcomes* 2 (1): 55–8.
- White, J. 2013. "Cost Control after the ACA." Public Administration Review [accessed on March 4, 2014]. Available at <http://policy.case.edu/CostControlAftertheACA.pdf>
- Williams, R. 2012. "Using the Margins Command to Estimate and Interpret Adjusted Predictions and Marginal Effects." *The Stata Journal* 12 (2): 308–31.
- Young, B. A., E. Lin, M. Von Korff, G. Simon, P. Ciechanowski, E. J. Ludman, S. Everson-Stewart, L. Kinder, M. Oliver, E. J. Boyko, and W. J. Katon. 2008. "Diabetes Complications Severity Index and Risk of Mortality, Hospitalization, and Healthcare Utilization." *The American Journal of Managed Care* 14 (1): 15–23.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Table S1. Results of Physician Characteristics between All P4P Participated Physicians in Phase 1 and Phase 2.

Table S2. Results of the Adjusted Predicted Probabilities and Marginal Effects of the Association between Physician Characteristics and Probability of Patients Being Excluded.

Table S3. Unadjusted Percentage of Nonenrollments by Physician Specialties and Patient Loads.