

Effect of Physician Pay-for-Performance (P4P) Incentives in a Large Primary Care Group Practice

Presenter: Laurel Trujillo, MD¹

Collaborators:
Harold Luft, PhD^{2,3}
Sukyung Chung, PhD^{2,3}
Latha Palaniappan, MD, MS²
Haya Rubin, MD, PhD²

¹Palo Alto Medical Foundation

²Palo Alto Medical Foundation Research Institute

³Phillip R. Lee Institute for Health Policy Studies, UCSF

Supported by AHRQ Task Order HHS A290200600023

Empirical Evidence of P4P

- Recent studies of P4P show modest effects
 - Group level incentives
 - Rosenthal et al. (2005): increase in cervical cancer screening, but no effect on mammography and HbA1c testing
 - Roski et al. (2003): better documentation of tobacco use, but no change in provision of quitting advice
 - Physician-specific (vs. no) financial incentives
 - Levin-Scherz et al. (2006): increased diabetes screening, but no effect on asthma controller prescription
 - Beaulieu & Horrigan (2005): improvement in most of the process and outcome measures of diabetes care
 - Gilmore et al. (2007): improvement in most process measures (e.g. cancer screening, diabetes care)
 - Financial incentives were generally accompanied by other quality improvement efforts such as performance reporting

Empirical Evidence of P4P (cont.)

- Limitations of previous studies:
 - Payer-driven initiatives
 - Quality measures and incentive schemes were given to, rather than chosen by, physicians or physician groups
 - Only part of the physicians' patients were eligible for incentives
 - Based on claims data
 - Limited physician-level information; no opportunity to investigate specific physician characteristics associated with incentives
 - Incentives paid annually or at the end of the study
 - Effect of timing of receipt of payment, in addition to the provision of performance reporting, is unknown

3

Research Questions

- Does a P4P program with physician-specific incentives implemented in a large primary care group practice improve quality of care provided?
- Does the frequency of payment (quarterly vs. year-end) make a difference in performance?
- What are the physician characteristics explaining variations in scores over time?

4

Study Setting

- Palo Alto Medical Foundation (PAMF)
 - Non-profit organization
 - Contracts with 3 physician groups in Northern California
- Palo Alto Division (PAMF/PAD)
 - 5 sites at Bay Area: Palo Alto, Los Altos, Fremont, Redwood City, Redwood Shores
 - Physician payment is based on relative value units of service
 - Electronic health records since 2000
 - Implemented *physician-specific* financial incentives in 2007

5

The Incentive Program

- Physician-specific incentives based on own performance
- Comprehensive
 - All the primary care physicians (N = 179) and all their patients regardless of insurance type
 - Family Medicine, Internal Medicine, Pediatrics
- Physician participation
 - In determining performance measures and incentive formula
- Frequency and amount of bonus payment
 - Randomly assignment to “quarterly” or “year-end” bonus
 - Maximum bonus: \$1250/qtr or \$5000/yr (~2-3% of salary)
 - Payment delivered about 6 weeks following the evaluation quarter (with two months delay for the first quarter reporting & payment)

6

The Incentive Program (cont.)

- Various quality measures
 - Both outcome and process measures
 - 10 were existing measures reported to physicians (2004+)
 - 5 new pediatrics-specific measures were selected based on guidelines and some were further modified during the year; These pediatric measures are excluded in our analyses
- Quarterly performance reporting
 - Quarterly email alert with an electronic link to quality workbook (2004+)
- Funds
 - IHA P4P incentives were supplemented by the organizational fund
 - Allowed application to all patients, not just those in IHA plans

7

Incentivized Quality Measures

| Measure | Description | Category |
|---------------------------|---|----------|
| Diabetes HbA1c control* | HbA1c <=7 (diabetes patients) | Outcome |
| Diabetes BP control | Blood pressure <=130/80 (diabetes patients) | Outcome |
| Diabetes LDL control* | LDL <=100 (diabetes patients) | Outcome |
| Asthma Rx*† | Long-term controller prescribed (asthma patients) | Process |
| Ht & Wt measured | Height and weight measured for BMI calculation | Process |
| Chlamydia screening**† | Chlamydia testing done (eligible women) | Process |
| Colon cancer screening | Colon cancer screening complete (adults age 50+) | Process |
| Cervical cancer screening | Pap smear done (eligible women) | Process |
| Tobacco Hx entered† | History of tobacco use was asked and recorded | Process |

Percent score = [numerator (i.e. patients who met the guideline) / denominator (i.e. patients who were eligible for the recommended care)] X100

*Similar measures (with different targets and population) were included in the IHA P4P program.

†These measures apply to some pediatrics patients.

8

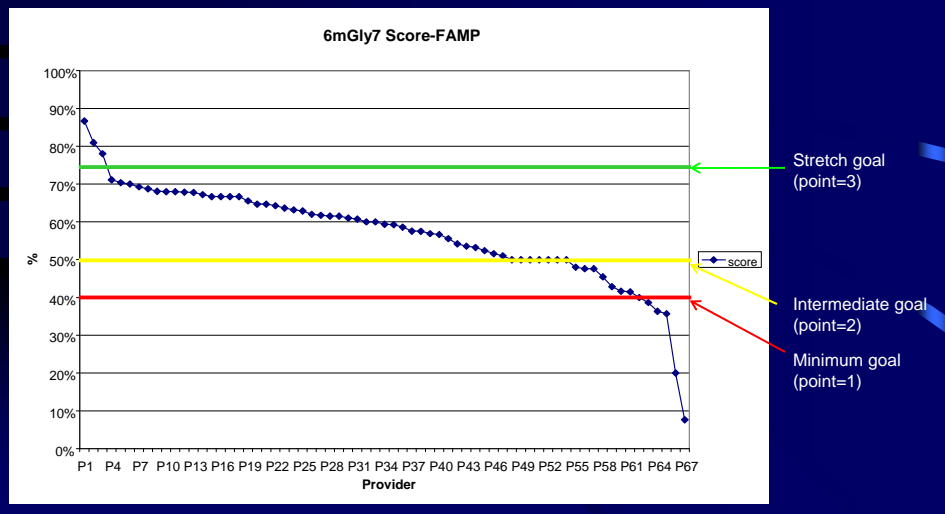
Other Quality Measures: Examples*

| Measure | Description | Category |
|-------------------------|---|----------|
| Diabetes HbA1c control* | HbA1c <=8 (diabetes patients) | Outcome |
| Diabetes BP control | Blood pressure <=140/90 (diabetes patients) | Outcome |
| Diabetes LDL control* | LDL <=130 (diabetes patients) | Outcome |
| Hypertension BP control | Blood pressure <=140/90 (hypertension patients) | Outcome |
| Diabetes HbA1c check | HbA1c was measured within the past 6 months | Process |
| Diabetes BP check | BP was measured within the past 12 months | Process |
| Diabetes LDL check | LDL was measured within the past 12 months | Process |
| Hypertension BP check | BP was measured within the past 12 months | Process |
| Alcohol Hx entered | History of alcohol use was asked and recorded | Process |

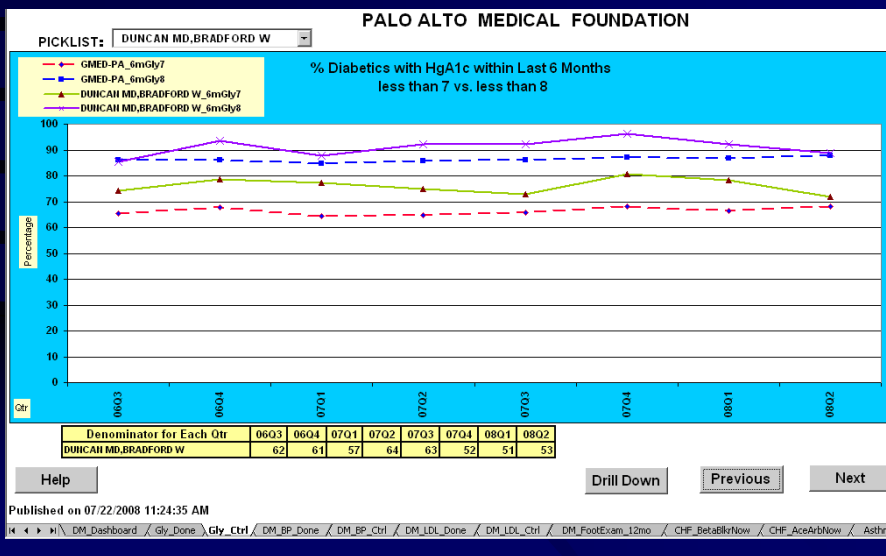
*These were not incentivized, but were reported in the quality workbook.

9

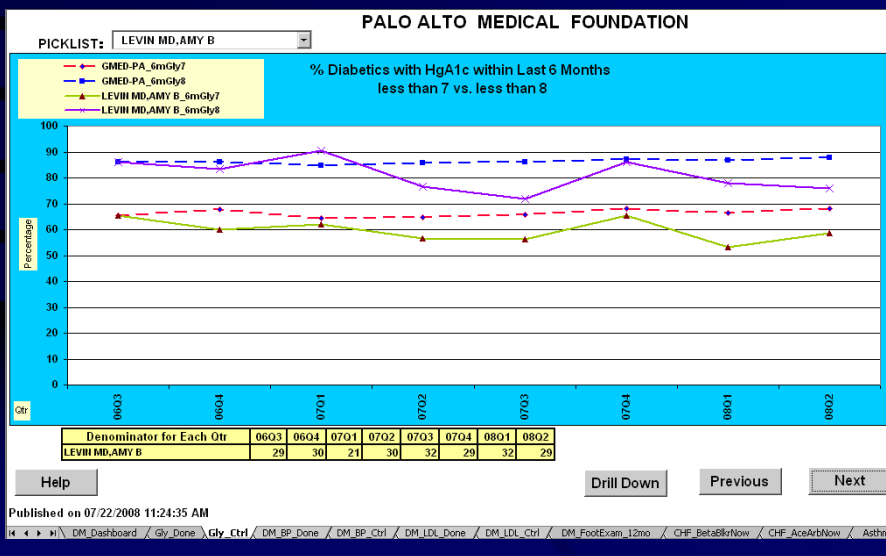
Example: Quality Workbook for "Diabetes HbA1c Control"



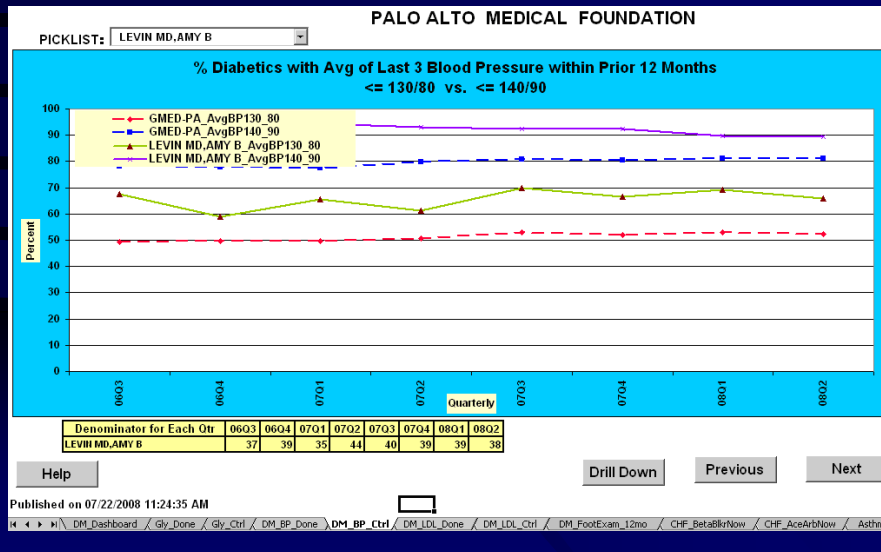
Example: Quality Workbook (cont.) Individual Physician's vs. Department's Score



Example: Quality Workbook (cont.) Individual Physician's vs. Department's Score



Example: Quality Workbook (cont.) Individual Physician's vs. Department's Score



Incentive Formula

- Incentive payment =
composite score X maximum amount {=\$1250/quarter}
- Composite score =
$$\frac{\sum \text{achieved points}}{\sum \text{maximum achievable points}}$$
- Required number of patients and measures for a bonus
 - Measures with <6 eligible patients for a physician in a quarter were not counted as a qualifying measure
 - Physicians with <4 qualifying measures in a quarter did not receive a bonus for the quarter

Number of Participating Physicians

| Category | N=167* | Frequency (%) |
|---------------------------|--------|---------------|
| Incentive frequency | | |
| Quarterly | 77 | 46.1 |
| Year-end | 90 | 53.9 |
| Location | | |
| Fremont | 44 | 26.4 |
| Los Altos | 26 | 15.6 |
| Palo Alto | 76 | 45.5 |
| Redwood City | 9 | 5.4 |
| Redwood Shores | 12 | 7.2 |
| Department | | |
| Family medicine | 68 | 40.7 |
| General internal medicine | 56 | 33.5 |
| Pediatrics | 43 | 25.8 |

*Among the initial sample (n=179), 12 physicians did not participate in the program due to various reasons (e.g. lack of number of patients, medical/sabbatical leave, etc.).

15

Average Number Patients and Scores at Quarter I, 2007

| Quality measure | #physicians with 6+ eligible patients at Q1 (N=167) | Average # eligible patients/ physician (denominator) | Average % Score = (numerator / denominator) x 100) |
|---------------------------|---|--|--|
| <i>Outcomes</i> | | | |
| Diabetes HbA1c control | 122 | 39 | 60% |
| Diabetes BP control | 122 | 49 | 51% |
| Diabetes LDL control | 122 | 43 | 57% |
| <i>Process</i> | | | |
| Cervical cancer screening | 123 | 529 | 77% |
| Chlamydia screening | 138 | 41 | 36% |
| Colon cancer screening | 122 | 315 | 45% |
| Asthma Rx | 136 | 21 | 92% |
| Ht & Wt measured | 152 | 926 | 71% |
| Tobacco Hx entered | 161 | 328 | 77% |

16

Does a P4P program with physician-specific incentives implemented in a large primary care group practice improve quality of care provided?

Comparison of 2006-7 Change to 2005-6 Change: P4P Measures

| Measures | Average % Score | | | Diff. [06-05] †‡ | Diff. [07-06] †‡ | Diff-in-diff [07-06]-[06-05] ‡ |
|--------------------------------|-----------------|------|------|------------------------|------------------------|--------------------------------------|
| | 2005 | 2006 | 2007 | | | |
| Diabetes HbA1c control (<=7) | 58% | 60% | 62% | | ** | |
| Diabetes BP control (<=130/80) | 47% | 49% | 53% | ** | ** | ** |
| Diabetes LDL control (<=100) | 60% | 63% | 60% | ** | ** | (**) |
| Cervical cancer screening | 75% | 77% | 79% | ** | ** | |
| Chlamydia screening | 36% | 37% | 38% | | * | |
| Colon cancer screening | 38% | 40% | 47% | ** | ** | ** |
| Asthma Rx | 91% | 92% | 92% | | * | |
| Ht & Wt measured | 68% | 70% | 73% | ** | ** | |
| Tobacco Hx entered | 72% | 75% | 79% | ** | ** | ** |

*p<0.05; **p<0.01

†Statistics based on the results from the multilevel mixed-effects linear regression (z-statistics).

‡Parentheses are used when the difference ((p2007 – p2006) or (p2006 – p2005)) is negative.

Comparison of 2006-7 Change to 2005-6 Change: Non-P4P Measures

| Measures | Average % Score | | | Diff. [06-05] †‡ | Diff. [07-06] †‡ | Diff-in-diff [07-06]-[06-05] ‡ |
|--------------------------------|-----------------|------|------|------------------------|------------------------|--------------------------------------|
| | 2005 | 2006 | 2007 | | | |
| Diabetes HbA1c control (<=8) | 81% | 81% | 83% | * | ** | |
| Diabetes BP control (<=140/90) | 77% | 78% | 81% | ** | ** | |
| Diabetes LDL control (<=130) | 86% | 88% | 87% | ** | | (**) |
| Hypertension BP ctl (<=140/90) | 64% | 67% | 72% | ** | ** | ** |
| Hypertension BP check | 90% | 90% | 90% | | | |
| Alcohol Hx entered | 67% | 69% | 73% | ** | ** | ** |

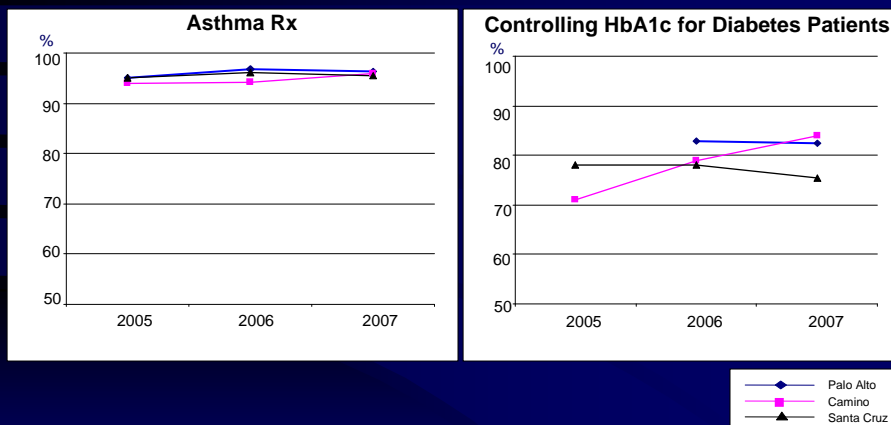
*p<0.05; **p<0.01

†Statistics based on the results from the multilevel mixed-effects linear regression (z-statistics).

‡Parentheses are used when the difference is negative.

19

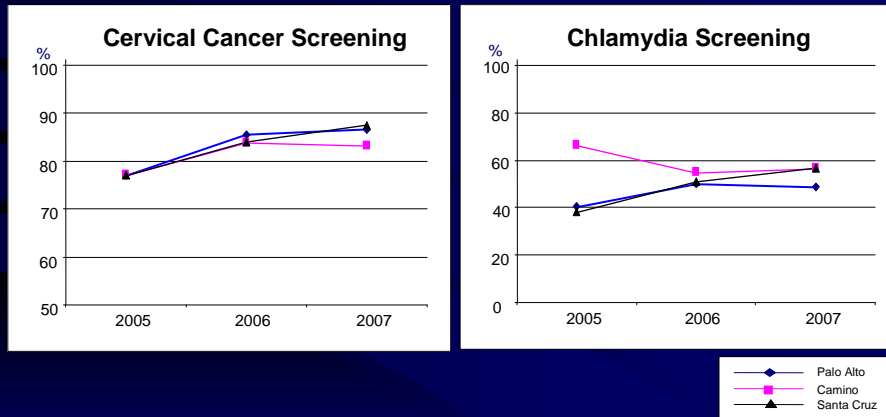
Comparison to Other Groups' Scores (2005-2007)



These are IHA P4P measure scores. Definitions of the measures were similar to those incentivized at PAD, but the eligible patients for the IHA measures are limited to HMO patients.

20

Comparison to Other Groups' Scores (2005-2007)

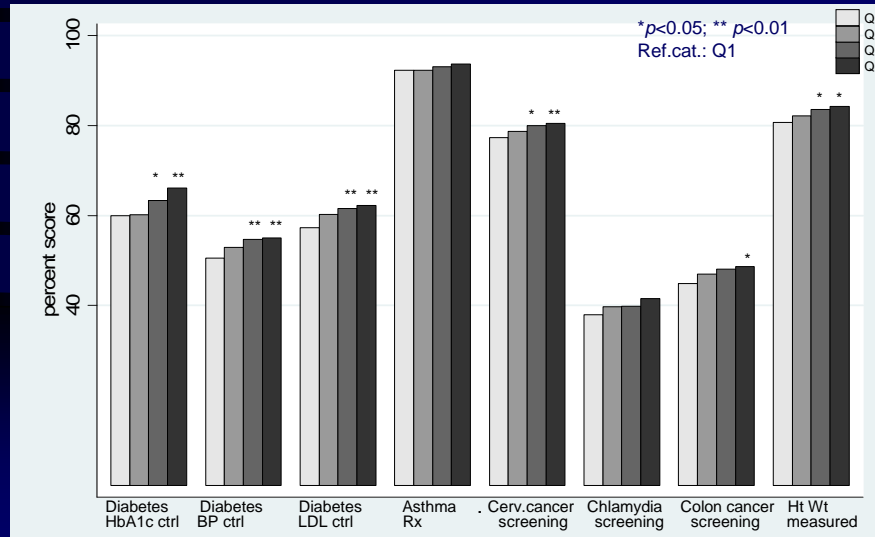


These are IHA P4P measure scores. Definitions of the measures were similar to those incentivized at PAD, but the eligible patients for the IHA measures are limited to HMO patients.

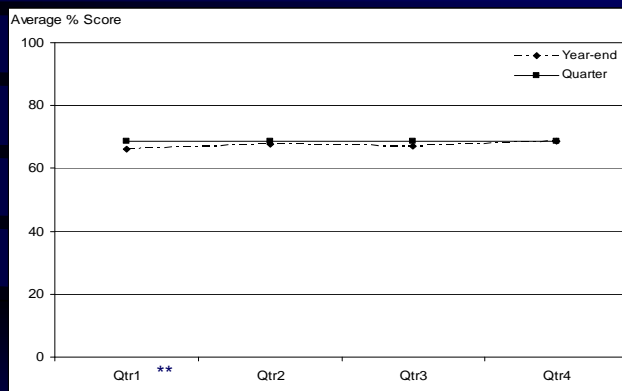
21

Does the frequency of payment (quarterly vs. year-end) make a difference in performance?

Improvement in Scores over the Four Quarters of 2007



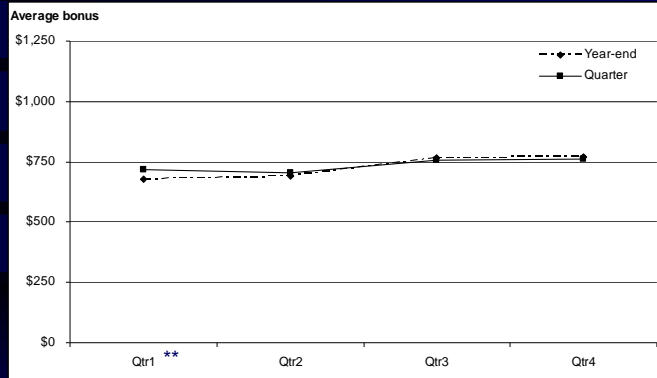
No Effect of Frequency of Payment on Scores*



* No statistical difference in the average score (each quarter) or trend in score (over the year) was detected between two arms, after controlling for indicators of quarter, measure, practice site and department.

** For the first quarter, there was two months delay in the reporting and payment.

No Effect of Frequency of Payment on Bonus Amount*



* No statistical difference in the average score (each quarter) or trend in score (over the year) was detected between two arms; However, there is increasing trend in bonus amount only in the year-end arm (Q3, Q4 > Q1; $p < 0.01$).

** For the first quarter, there was two months delay in the reporting and payment.

What are the physician characteristics explaining variations in scores over time?

Effects of Physician Characteristics

| Dependent variables: | % score at 2007 Q1 (n=1179) | % score at 2007 Q1 (N=1179) | Change in % score [2007-2006] (N=1142) |
|--|--------------------------------|--------------------------------|---|
| Average score in 2006 (0-100) | | 0.90** (0.01) | -0.19** (0.01) |
| Female | 2.78** (0.95) | -0.48 (0.44) | 0.131 (0.43) |
| Years of practice (4-46) | 0.16** (0.05) | -0.01 (0.02) | -0.03 (0.02) |
| Foreign graduate | 0.80 (2.29) | -1.36 (1.05) | -1.82 (1.02) |
| Internal medicine (vs. Family medicine) | 1.68 (0.94) | -0.15 (0.43) | -0.13 (0.42) |
| Pediatrics (vs. Family medicine) | -31.39** (1.55) | -3.22** (0.85) | -4.55** (0.82) |
| R-squared | 0.64 | 0.92 | 0.26 |

* p<0.05; ** p<0.01

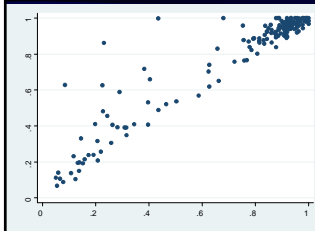
Linear regression; unit of observation: physician-measure

Other covariates included are indicators of each measure and practice site.

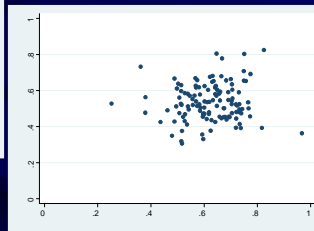
27

Correlation in Scores Across Measures (within physicians)

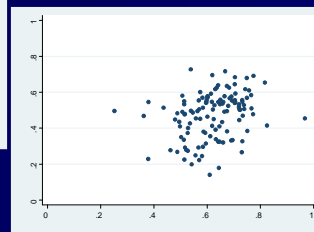
Y: Hx tobacco entered (P4P)
X: Hx alcohol entered (non-P4P)



Y: Diabetes BP control (P4P)
X: Diabetes HbA1c control (P4P)



Y: Colon cancer screening (P4P)
X: Diabetes HbA1c control (P4P)



28

Summary of Findings

- No strong evidence of quality improvement led by physician-specific financial incentives
 - Other simultaneous organizational or regional efforts may have led quality improvement.
- Frequency of incentive payment (quarterly vs. year-end) does not make a difference
 - The effect of frequency of incentive payment may have been mitigated by the quarterly report sent to both arms.
- Within- and across- physician variations
 - For each measure, within-physician scores are consistent over time
 - No strong correlation across measures within a physician

29

Implications

- In the context of other organizational-level quality improvement efforts (e.g. regular audit/feedback on individual physicians' quality; EHR), relatively small financial incentives to individual physicians have limited incremental effects on well-established measures.
- Other types of organizational support (e.g. increasing coverage of staff hours for quality improvement; information technology to easily track target patients) needs to be explored.

30

