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

Presenter: Harold Luft, PhD[1,2]
Collaborators:
Sukyung Chung, PhD[1,2]
Latha Palaniappan, MD, MS[1]
Haya Rubin, MD, PhD
Laurel Trujillo, MD[3]

[1]Palo Alto Medical Foundation Research Institute
[2] Phillip R Lee Institute for Health Policy Studies, UCSF
[3]Palo Alto Medical Foundation

Supported by AHRQ Task Order HHSA290200600023

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 some of the physicians' patients were eligible for incentives
  - Based on claims data
    - Limited physician-level information; no ability to investigate 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

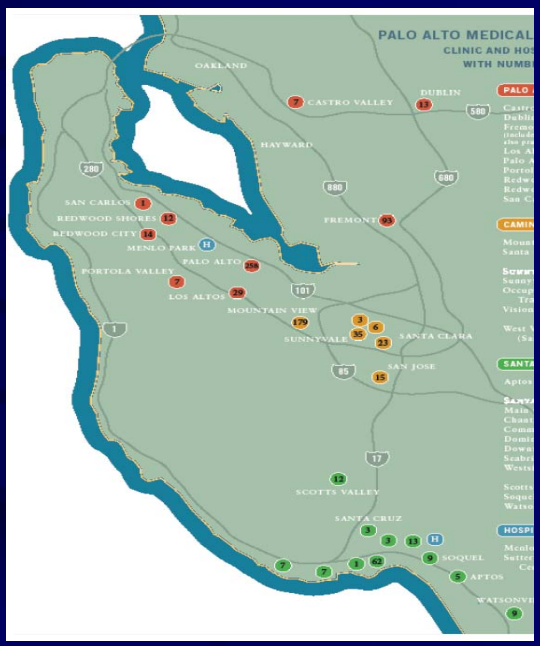
Research Questions

- 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?
- Do physician characteristics explain variations in scores and changes over time?

Study Setting

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

PAMF Clinics



The Incentive Program

- Physician-specific incentives based on own performance
- Comprehensive
  - All the primary care physicians (N = 179) and all their patients regardless of specific insurance plan
  - Family Medicine, Internal Medicine, Pediatrics
- Physician participation
  - In determining performance measures and incentive formula

The Incentive Program (cont.)

- Frequency and amount of bonus payment
  - Random 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 (but a two month delay for the first quarter reporting & payment)

The Incentive Program (cont.)

- Quarterly performance reporting
  - Quarterly email alert with an electronic link to quality workbook (2004+)
- Funding of the incentive program
  - IHA P4P incentives were supplemented by the PAMF organizational fund
  - Allowed application to all patients, not just those in IHA plans

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

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.

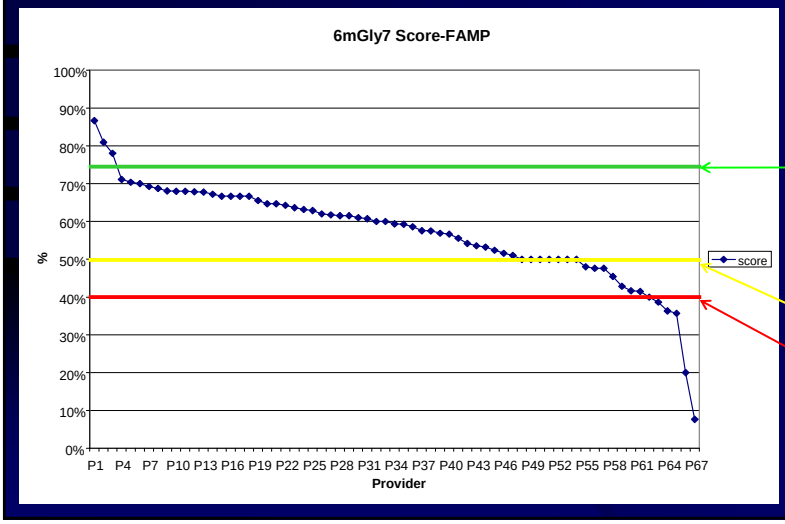
†These measures apply to some pediatrics patients.

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

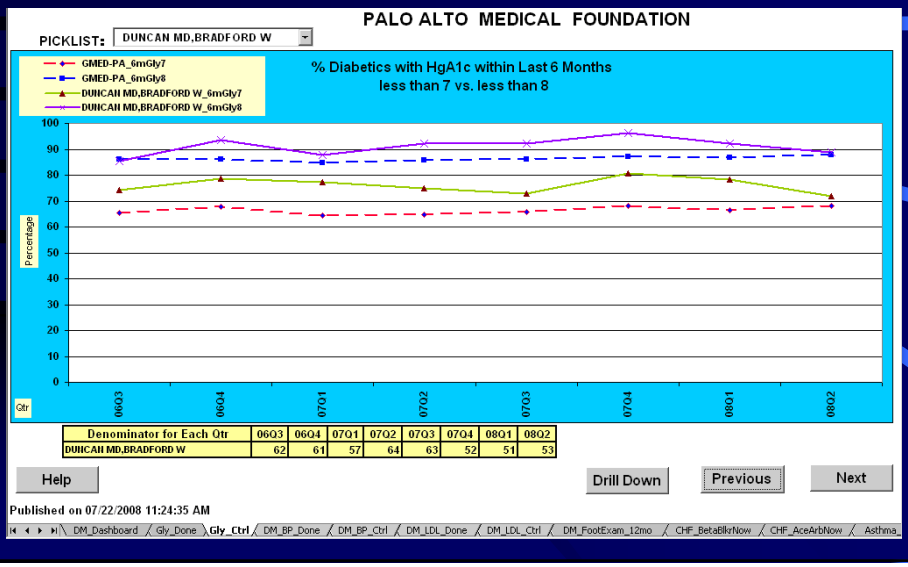
*These were not incentivized, but were reported in the quality workbook.

Example: Quality Workbook for "Diabetes HbA1c Control"



Example: Quality Workbook (cont.)

Individual Physician's vs. Department's Score



Incentive Formula

- Incentive payment =
  composite score * maximum amount {=$1250/quarter}
- Composite score =
  ∑ achieved points / ∑ 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 received a bonus for the quarter

Physician Characteristics

Variables	Mean (SD) [min, max]	Frequency % (n)
Female		65% (107)
Age	42 (9) [29, 70]	
Years at PAMF	8 (7) [0, 40]	
Years of practice not at PAMF	7 (6) [0, 45]	
Department		
Family practice		42% (69)
General internal medicine		34% (56)
Pediatrics		25% (41)
Practice site		
Palo Alto		42% (78)
Los Altos		15% (24)
Fremont		26% (43)
Redwood city		5% (9)
Redwood shores		7% (12)

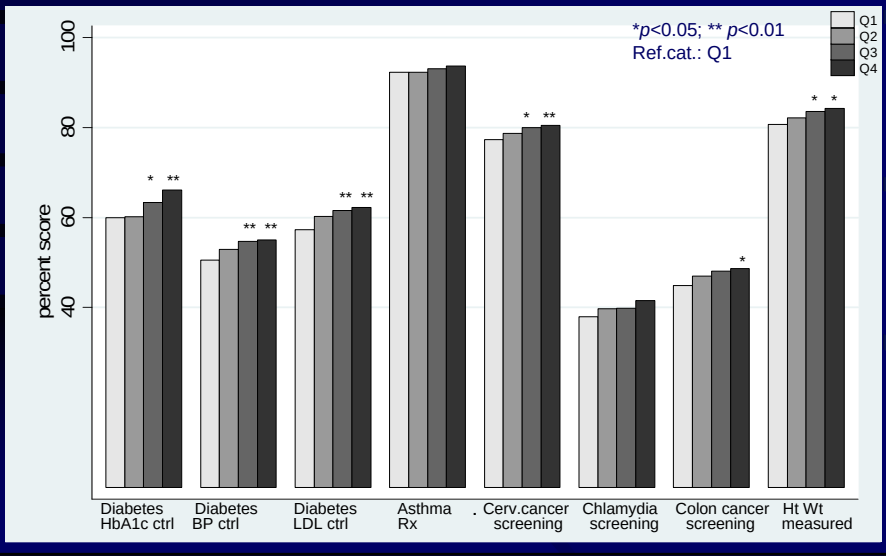
N=167; Among the initial sample (n=179), 12 physicians did not participate in the program due to various reasons (e.g. lack of number of patients, medical/sabbatical leave, etc.).

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)
Outcomes			
Diabetes HbA1c control	122	39	60%
Diabetes BP control	122	49	51%
Diabetes LDL control	122	43	57%
Process			
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%

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

Improvement in Scores over the Four Quarters of 2007



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 HbA1ccontrol (<=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%	**	**	**

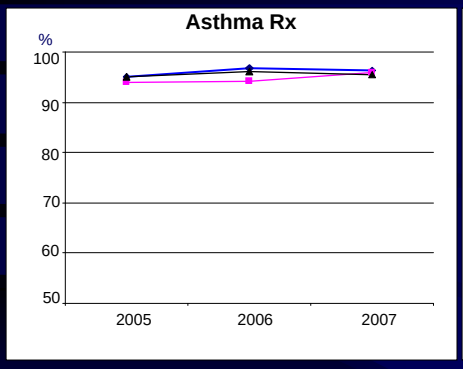
*p<0.05; **:p<0.01
†Statistics based on the results from the multilevel mixed-effects linear regression (z-statistics).
‡Parentheses are used when the difference ((p2007 – p2006) or (p2006 – p2005)) is negative.

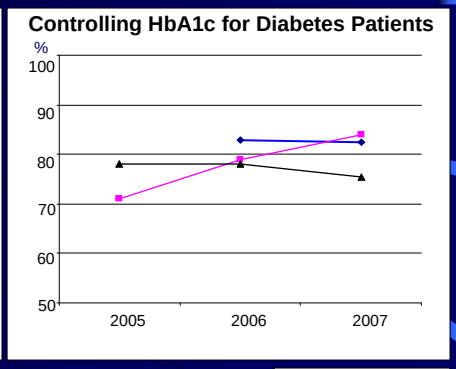
Comparison of 2006-7 Change to 2005-6 Change: Non-P4P Measures

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.

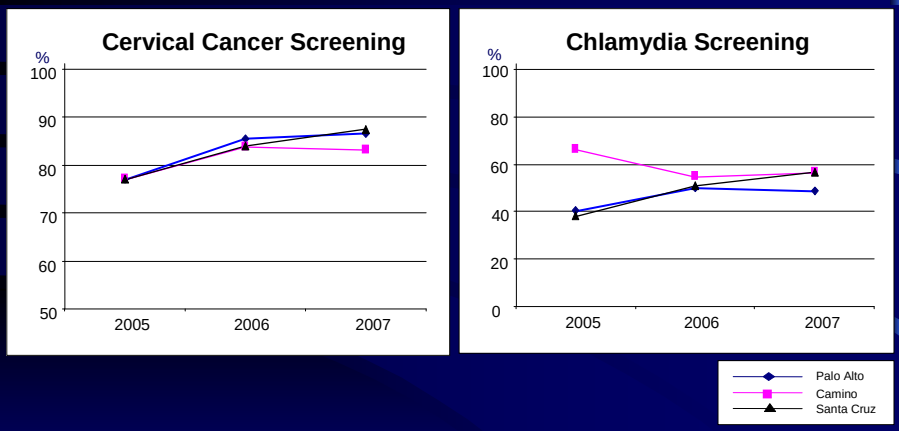
Comparisons Across the Three PAMF Groups (2005-2007)





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.

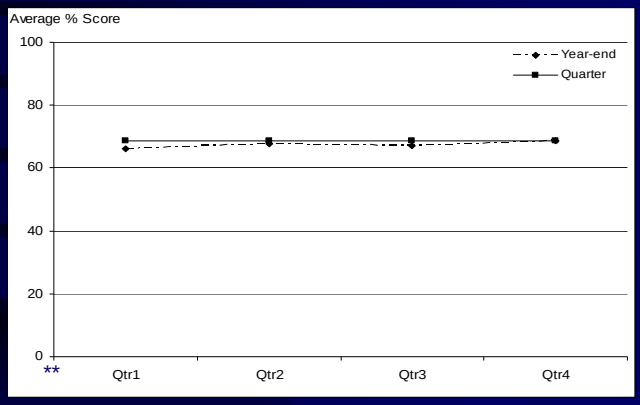
Comparisons Across the Three PAMF Groups (2005-2007)



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.

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

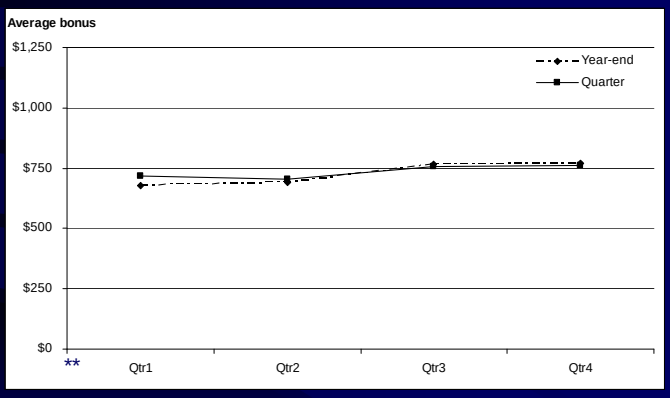
No Effect of Frequency of Payment on Scores*



* 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 physician characteristics explain variations in scores and changes in scores over time?

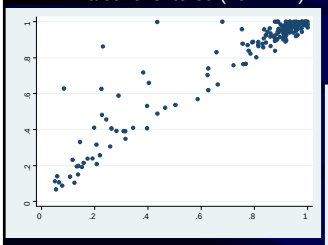
Effects of Physician Characteristics

Dependent variable: Average score (0-100) in 2007 for each measure (n=1182)			
Average score in 2006 (0-100)			0.81**
			(0.01)
Female	3.52**	3.10**	0.07
	(0.90)	(0.90)	(0.45)
Years at PAMF (0-40)	0.25**		
	(0.06)		
Years at PAMF: 6-10 (ref: 0-5years)		2.62**	-1.69**
		(0.99)	(0.49)
Years at PAMF: 11-20 (ref: 0-5 years)		3.87**	-1.23*
		(1.04)	(0.52)
Years at PAMF: 21-40 (ref: 0-5 years)		4.01*	-1.62
		(1.88)	(0.92)
Years of practice not at PAMF (0-45)	-0.01	-0.03	0.01
	(0.07)	(0.07)	(0.03)
EPIC “power user”	2.08	1.87	-0.35
	(1.50)	(1.50)	(0.74)
R-squared	0.674	0.674	0.921

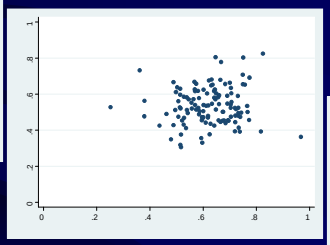
* p<0.05; ** p<0.01
Linear regression; unit of observation: physician-measure
Other covariates included are indicators of each measure, department and practice site.

Correlation in Scores Across Measures (within physicians)

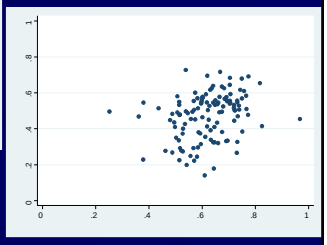
Y: Hx tobacco entered (P4P)
X: Hx alcohol entered (non-P4P)



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



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



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
  - Physician scores for a measure are consistent over time
  - No strong correlation across measures

Discussion

Confounders

- (Lack of) improvement with P4P may be due to simultaneous ongoing QI efforts
- For example,
  - Palo Alto Division
    - Bronze/Silver program since 2006
  - Camino
    - Efforts focused on IHA measures/populations
  - Santa Cruz
    - Departmental level incentives (?)

Generalizabilty

- Established measures
  - Regular audit/feedback on individual physicians' quality on these measures for several years
- High quality organization
  - Already high performing for the measures assessed
- Information technology
  - Allowed for easy tracking of target patients and individual physician's performance
- Patient population
  - Relatively high education and wealth status

"Bonus"

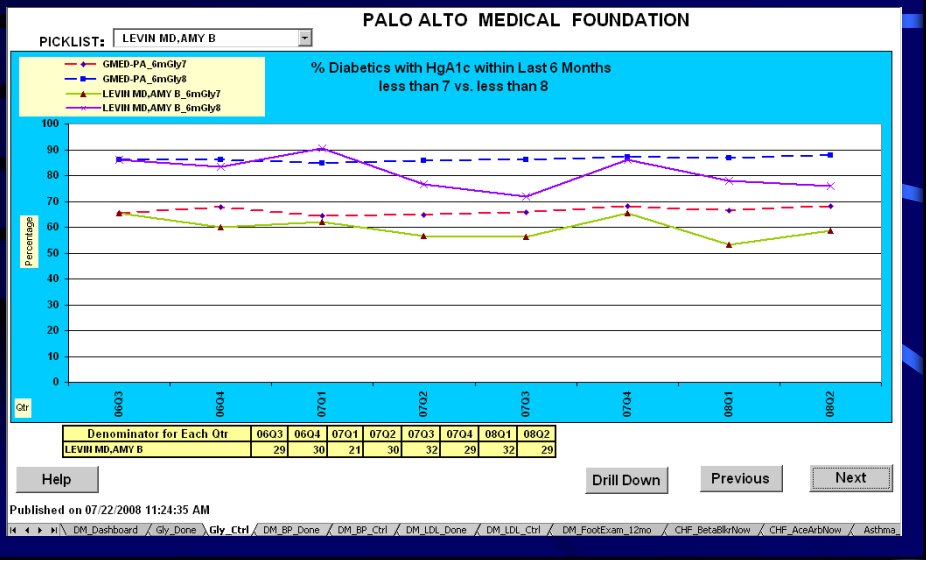
- Is maximum $5000/year too much or too little?
- Once a year vs. more frequent payment?
- Other forms of bonus payment?
- Does the bonus really matter?

Potential other use of the funds to improve quality?

- Increasing coverage for staff hours dedicated to QI
- Information technology to easily track target patients
- Other ideas?

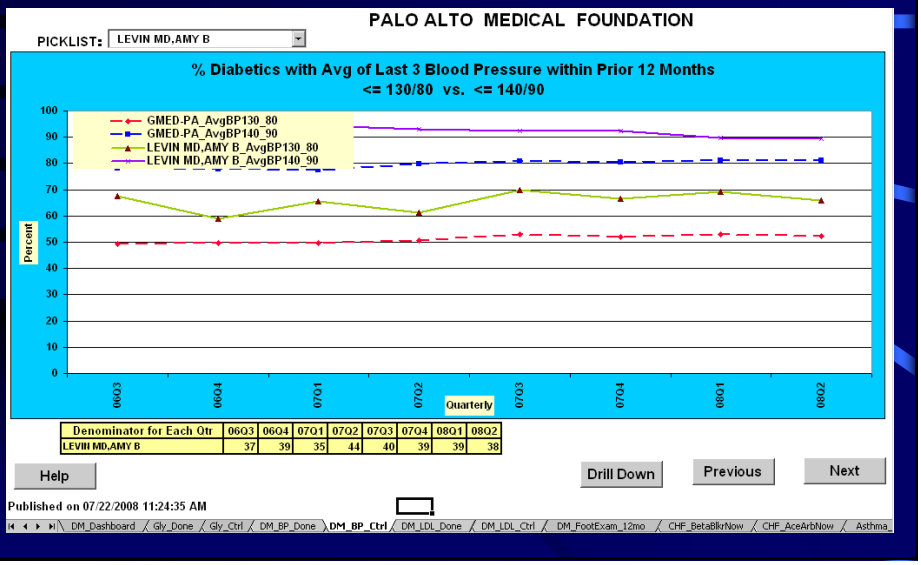
Example: Quality Workbook (cont.)

Individual Physician's vs. Department's Score



Example: Quality Workbook (cont.)

Individual Physician's vs. Department's Score



PAMF/PAD Catchment Area