

determine whether these contained any additional information not covered by the primary randomized, controlled trial reports.

We excluded articles that did not meet specific criteria in terms of the quality of the research and reporting. These were:

For interventional trials

- Intervention randomized
- Inclusion/exclusion criteria clear and appropriate
- Greater than 75% follow-up
- Note: two criteria usually used to judge the quality of a randomized, controlled trial—provision of placebo to the control group and blinding of the subjects—are not applicable in this situation

For systematic reviews

- Information source appropriate
- Information source adequately searched
- Inclusion/exclusion criteria clear and appropriate
- Data abstraction performed by at least 2 independent reviewers
- Principal measures of effect and the methods of combining results appropriate

Search Strategy

The objective of our search strategy was to identify all published QBP randomized trials and all ongoing research into QBP strategies. For the literature review, we used standard search strategies involving the querying of two online databases (MEDLINE® and Cochrane) using key words, followed by evaluation of the bibliographies of relevant articles, Web sites of relevant organizations (especially of funding agencies providing project summaries and of employer organizations pursuing QBP), and reference lists provided by our Technical Expert Panel (Table 1).

Table 1: Information sources for literature review and catalog of ongoing research

Goal of Search	Databases searched	Relevant Organizations (for Web-based searches)
Identify randomized, controlled trials of quality-based purchasing strategies	MEDLINE® Cochrane	AHRQ Robert Wood Johnson Foundation California HealthCare Foundation Commonwealth Fund National Business Coalition on Health Leapfrog Group

Database Searches

To identify potentially relevant articles in the medical literature, we searched MEDLINE[®] and Cochrane databases and references provided by our Expert Advisors.

MEDLINE[®] search strategies. We searched MEDLINE[®] (January 1980 to December 15, 2003) for English language articles using the search terms described in Table 2. Some citations were reviewed and articles were retrieved in more than one of the searches listed below.

Table 2: MEDLINE[®] searches to identify potentially relevant primary data

Search Terms	Citations reviewed	Articles retrieved
"pay" AND "quality" AND "measurement"	80	1
"incentive" AND "quality" AND "measurement"	195	5
"financial incentive" AND "quality" AND "efficiency"	125	11
"provider supply" AND "incentive"	15	0
"quality" AND "error" AND "safety" AND "cost"	16	0
"pay" AND "performance"	389	2
"pay" AND "incentive" AND "quality"	79	3
"pay" AND "quality" AND "measurement" AND "Randomized Controlled Trial" [Publication Type]	8	1
"incentive" AND "quality" AND "measurement" AND "Randomized Controlled Trial" [Publication Type]	13	2
"financial incentive" AND "quality" AND "efficiency" AND "Randomized Controlled Trial" [Publication Type]	1	1
"provider supply" AND "incentive" AND "Randomized Controlled Trial" [Publication Type]	0	0
"quality" AND "error" AND "safety" AND "cost" AND "Randomized Controlled Trial" [Publication Type]	0	0
"pay" AND "performance" AND "Randomized Controlled Trial" [Publication Type]	6	1
"pay" AND "incentive" AND "quality" AND "Randomized Controlled Trial" [Publication Type]	1	1
"incentive" AND "quality" AND "Randomized Controlled Trial" [Publication Type]	42	2
"pay" AND "quality" AND "Randomized Controlled Trial" [Publication Type]	26	2
"value" AND "incentive" AND "Randomized Controlled Trial" [Publication Type]	49	0
"value" AND "pay" AND "Randomized Controlled Trial" [Publication Type]	10	0
"Insurance, Health, Reimbursement" [MESH] AND "Randomized Controlled Trial" [Publication Type]	72	6
"Medicare Payment Advisory Commission" [MESH] AND "Randomized Controlled Trial" [Publication Type]	0	0
"Physician Payment Review Commission" [MESH] AND "Randomized Controlled Trial" [Publication Type]	0	0
"Prospective Payment Assessment Commission" [MESH] AND "Randomized Controlled Trial" [Publication Type]	1	0
"Prospective Payment System" [MESH] AND "Randomized Controlled Trial" [Publication Type]	28	1
"Salaries and Fringe Benefits" [MESH] AND "Randomized Controlled Trial" [Publication Type]	78	1
"Single-Payer System" [MESH] AND "Randomized Controlled Trial" [Publication Type]	2	0
"Fee-for-Service Plans" [MESH] AND "Randomized Controlled Trial" [Publication Type]	11	1
"Reimbursement Mechanisms" [MESH] AND "Randomized Controlled Trial" [Publication Type]	66	6

Search Terms	Citations reviewed	Articles retrieved
"Reimbursement, Incentive" [MESH] AND "Randomized Controlled Trial" [Publication Type]	10	4
"Cost and Cost Analysis" [MESH] AND "Randomized Controlled Trial" [Publication Type]	2,561	9
"Medical Errors" [MESH] AND "Randomized Controlled Trial" [Publication Type]	678	0
"Medication Errors" [MESH] AND "Randomized Controlled Trial" [Publication Type]	17	0
"Management Quality Circles" [MESH] AND "Randomized Controlled Trial" [Publication Type]	6	0
"Professional Review Organizations" [MESH] AND "Randomized Controlled Trial" [Publication Type]	3	0
"Quality Assurance, Health Care" [MESH] AND "Randomized Controlled Trial" [Publication Type]	586	14
"Quality Control" [MESH] AND "Randomized Controlled Trial" [Publication Type]	161	1
"Quality Indicators, Health Care" [MESH] AND "Randomized Controlled Trial" [Publication Type]	22	0
"Total Quality Management" [MESH] AND "Randomized Controlled Trial" [Publication Type]	45	2
"United States Agency for Healthcare Research and Quality" [MESH] AND "Randomized Controlled Trial" [Publication Type]	11	0
Total Articles	5413	76

The use of the asterisk expands search terms such that all combinations of terms with the phrase preceding the asterisk will be returned in the search (e.g., cost returns searches for cost, costs, etc.).

MESH = Medical Subject Heading

Cochrane search strategies. We searched the Cochrane databases from January 1, 1990 through December 15, 2003 (OVID, Evidence Based Medicine Reviews Multifile) using the search terms described in Table 3.

Table 3: Search terms and citations for Cochrane databases

Search terms	Citations reviewed	Articles retrieved
Pay	6	2
Incentive	4	0
Efficiency	74	0
Safety	264	0
Cost	210	2
Error	12	0
Performance	60	0
Value	95	0
Insurance	0	0
Reimbursement	0	0
Total	725	4

The use of the asterisk expands search terms such that all combinations of terms with the phrase preceding the asterisk will be returned in the search (e.g., cost returns searches for cost, costs, cost effectiveness, etc.).

Abstract Review

To identify potentially relevant articles for focused searching, at least two investigators (to ensure consistent application of the inclusion and exclusion criteria) reviewed each citation and, whenever an abstract was available, the abstract. Discrepancies in inclusion were resolved by discussion and re-review.

Evaluating Published Articles for Completeness of Reporting

We assessed each of the published articles for their completeness in reporting the factors we identified in our conceptual model that could influence a provider's response to incentives. Specifically, we scored them for the inclusion (or not) of descriptions of the elements in Table 4. We also recorded the type of care (preventive care, acute care, or chronic care) to which the quality measured pertained.

Table 4: Evaluating randomized controlled trials for completeness of reporting

Domain of the Conceptual Model	Specific Variable
Financial Characteristics of Incentive	<i>Recipient:</i> individual provider vs. provider group <i>Revenue potential:</i> magnitude of the financial incentive <i>Revenue potential:</i> incentive as a proportion of total income <i>Impact on cost:</i> direct costs and opportunity costs of complying
Nonfinancial Characteristics of Incentive	<i>Perceived attainability:</i> how easy/difficult it is to accomplish the task of the incentive <i>Performance domain measured:</i> structure, process, outcome
Predisposing Factors	<i>Financial characteristics of the environment:</i> proportion of income from: fee for service, salary, capitation <i>Financial characteristics of the environment:</i> number of other financial incentives in place <i>Provider characteristics:</i> demographics, specialty, and other immutable factors <i>Provider characteristics:</i> workload, proportion of patients if service where incentive relevant <i>Market characteristics:</i> community initiatives or performance standards
Enabling Factors	<i>Organizational characteristics:</i> size, type of practice, specialty, etc. <i>Organizational characteristics:</i> capabilities such as information systems, use of guidelines and feedback, etc. <i>Organizational characteristics:</i> leadership, culture, etc. <i>Patient characteristics:</i> demographics and other immutable factors <i>Patient characteristics:</i> type of insurance, benefits structure

Identifying Ongoing Research

Based on input from our expert advisors, our conceptual model, and practical considerations, we developed methods to catalog ongoing research into QBP that involved specifying: inclusion and exclusion criteria to identify potentially relevant research projects, search strategies to retrieve project abstracts, abstract review protocols, and a system of describing the study design of ongoing research projects.

Inclusion and Exclusion Criteria

Since the search for ongoing research focused on projects not yet reported in the literature, the criteria for identifying relevant projects focused on the planned intervention. Two types of research potentially met our inclusion criteria: projects designed as randomized controlled trials, or projects with interventions using QBP methods as described above (i.e., payment or performance reporting strategies) and applied at the community level (or in a broader geographic region, such as a State) that included historical or contemporaneous non-randomized control groups.

Search Strategy

We searched online health services research databases (HSRProj and AHRQ's Grants-On-Line Database or GOLD). We also searched the Web sites of other funders or coordinators of projects (e.g., the Leapfrog Group at www.leapfroggroup.org/RewardingResults/). Finally, we inquired of staff at AHRQ, the Robert Wood Johnson Foundation, the California HealthCare Foundation, and the Commonwealth Fund whether there was ongoing research that met our inclusion criteria being funded by those organizations. Table 5 lists our information sources for this aspect of the report.

Table 5: Information sources for the catalog of ongoing research

Goal of Search	Databases searched	Relevant Organizations (for Web-based searches and staff interviews)
Identify ongoing research evaluating quality-based purchasing strategies	GOLD (www.gold.ahrq.gov), HSRProj (via the National Library of Medicine at gateway.nlm.nih.gov/gw/Cmd)	AHRQ Leapfrog Group Robert Wood Johnson Foundation California HealthCare Foundation Commonwealth Fund

Database Searches

We searched the two available databases for ongoing health services research, using a similar search strategy for each (Tables 6 and 7). We accessed HSRProj through the National Library of Medicine's Gateway database at gateway.nlm.nih.gov/gw/Cmd and GOLD at www.gold.ahrq.gov.

GOLD search strategies. We searched GOLD through February 15, 2004 for grants funded by AHRQ using the categories described in Table 6. Through our combination of searches, we eventually evaluated all projects in GOLD.

Table 6: Search terms and citations for GOLD

Search by Category	Grants reviewed	Grants retrieved
Quality Outcomes	319	2
Quality Measures	189	2
Quality Improvement	256	2
Managed Care/Market Forces	98	1
Payment Strategies	22	1
Cost	121	0
New Knowledge	374	2
Total Grants	1379	10

HSRProj search strategies. We searched the HSRProj database through February 15, 2004 using the categories described in Table 7.

Table 7: Search terms and citations for HSRProj database

Search terms	Grant abstracts reviewed	Grants retrieved
Pay	49	1
Incentive	165	6
Efficiency	144	2
Safety	374	4
Error	160	1
Performance	546	7
Value	219	6
Reimbursement	136	2
Total Grants	1793	29

The use of the asterisk expands search terms such that all combinations of terms with the phrase preceding the asterisk will be returned in the search (e.g., cost returns searches for cost, costs, cost effectiveness, etc.).

Grant Abstract Review

Two investigators reviewed the abstracts of projects identified from the database searches to assess relevance to the technical review. Discrepancies in inclusion were resolved by discussion and re-review and by discussion with project officers at funding agencies or with the principal investigator of the project under consideration.

Describing the Study Design of Ongoing Research

For each research project, we interviewed either project staff (usually the principal investigator) or the project officer to determine the study design. We obtained information about the intervention—performance measures and incentives used—and the control group. The information sought is described in Table 8.

Table 8: Design information sought about ongoing research

Design Issue	Examples of Possible Responses
Patient Population from an Insurance Perspective	Privately Insured, Medicare, Medicaid, or multiple populations
Health Plan Setting	Health maintenance organization, preferred provider organization, point of service
Control Group	Randomized controlled trial vs. non-randomly selected contemporaneous control vs. historical control
Incentive Structure	Describe financial or reputational gains from superior performance
Performance Measures	Participation vs. clinical performance (for the latter, describe determinants of performance assessment, including weighting given when multiple measures are used)
Evaluation Plan/Goals	Assess determinants of participation in the program, catalog incentives used, test impact of incentives on clinical performance

Table 9. Evaluating Randomized, Controlled Trials for Completeness of Reporting

Domain of the Conceptual Model	Specific Variable	Christensen	Davidson	Fairbrother	Hickson	Hillman '98	Hillman '99	Kouides	Roski
Financial Characteristics of Incentive	<i>Recipient of the incentive:</i> individual provider vs. group	Reported (individual pharmacist)	Reported (individual physician)	Reported (individual physician)	Reported (individual physician)	Reported (individual physician or group)	Reported (medical group)	Reported (individual physician or group)	Reported (individual physician)
	<i>Revenue potential:</i> magnitude of the financial incentive	Reported (schedule of fees-for-service)	Reported (schedule of fees-for-service)	Reported (bonus up to \$7,500 vs. fees-for-service vs. control)	Reported (\$2/visit fee-for-service)	Reported (~33% chance to receive a bonus up to \$5,000)	Reported (~10% chance to receive a bonus; total potential \$ not reported)	Reported (fee-for-service, \$0.80-\$1.60 per vaccination)	Reported (bonus up to \$10,000)
	<i>Revenue potential:</i> incentive as a proportion of total income	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
	<i>Impact on cost:</i> direct costs and opportunity costs of complying	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
Non-financial Characteristics of Incentive	<i>Perceived attainability:</i> How easy/difficult it is to accomplish the task of the incentive	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
	<i>Performance domain measured:</i> structure, process, outcome	Reported (chronic care process: medication instruction)	Reported (preventive care process: well child continuity visits)	Reported (preventive care process: vaccinations)	Reported (preventive care process: well child visits)	Reported (preventive care process: vaccinations)	Reported (preventive care process: cancer screening)	Reported (preventive care process: vaccinations)	Reported (preventive care process: tobacco screening, tobacco cessation)

Table 9. Evaluating Randomized, Controlled Trials for Completeness of Reporting (cont'd)

Domain of the Conceptual Model	Specific Variable	Christensen	Davidson	Fairbrother	Hickson	Hillman '98	Hillman '99	Kouides	Roski
Predisposing Factors	<i>Financial characteristics of the environment:</i> proportion of income from: fee for service, salary, capitation	Not reported	Not reported	Report that all apply, but do not give percentages	Reported (salary)	Not reported	Not reported	Not reported	Report that all apply, but do not give percentages
	<i>Financial characteristics of the environment:</i> number of other financial incentives in place	Not reported	Not reported	Not reported	Not reported	Report many other incentives, but do not describe them	Report many other incentives, but do not describe them	Not reported	Not reported
	<i>Provider characteristics:</i> demographics, specialty, and other immutable factors	Not reported	Not reported	Only board certification reported	Only specialty reported	Only specialty reported	Only specialty reported	Not reported	Specialty reported
	<i>Provider characteristics:</i> workload, proportion of patients where incentive is relevant	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
	<i>Market characteristics:</i> community initiatives or performance standards	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported

Table 9. Evaluating Randomized, Controlled Trials for Completeness of Reporting (cont'd)

Domain of the Conceptual Model	Specific Variable	Christensen	Davidson	Fairbrother	Hickson	Hillman '98	Hillman '99	Kouides	Roski
Enabling Factors	<i>Organizational characteristics:</i> size, type of practice, specialty, etc.	Not reported	Not reported	Not reported	Reported (type and specialty)	Reported (varies)	Reported (varies)	Size reported	Reported
	<i>Organizational characteristics:</i> capabilities such as information systems, use of guidelines and feedback, etc.	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
	<i>Organizational characteristics:</i> leadership, culture, etc.	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported	Not reported
	<i>Patient characteristics:</i> demographics and other immutable factors	Not reported	Not reported	Age, high poverty levels reported	Age, high poverty levels reported	Age and race reported	Age reported	Age reported	Not reported
	<i>Patient characteristics:</i> type of insurance, benefits structure	Reported (Medicaid patients, no cost to patients)	Medicaid patients, benefits structure not reported	Not reported	Reported (most uninsured)	Medicaid patients, benefits structure not reported	Medicaid patients, benefits structure not reported	Medicare demonstration project patients, benefits structure not reported	Not reported

positive and one was negative. With bonuses tied to performance, there were two positive results and three negative.

Performance domain measured. Among the articles included, there were seven studies of preventive care with nine dependent variables assessed. Among these nine outcomes, five were positive and four negative. The single study addressing chronic care was positive.⁶¹

Patient factors. Authors did not report the burden adherence would place on patients in any of the articles we found. However, in a general sense, we found that incentives to achieve performance were more effective when the indicator to be followed required less patient cooperation (e.g., receiving vaccinations or answering questions about smoking) than when significant patient cooperation was needed (e.g., to quit smoking, Table 10).

Table 10: Available results by conceptual model domains tested

Conceptual Domain and Specific Variable	Results
Financial Characteristics of the Incentive: Recipient Individual vs. Group	<ul style="list-style-type: none"> • Individual: 5 positive, 2 negative • Group or Individual: 1 positive, 2 negative
Financial Characteristics of the Incentive: Recipient Provider Type	<ul style="list-style-type: none"> • Physicians: 5 positive, 4 negative • Pharmacists: 1 positive
Financial Characteristics of the Incentive: Magnitude	<ul style="list-style-type: none"> • No clear relationship between magnitude and result • Both trials in which the performance required to achieve a bonus was unknown were negative
Nonfinancial Characteristics of the Incentive: Performance Domain Measured	<ul style="list-style-type: none"> • Preventive care: 5 positive (3 immunizations, 1 well-child, 1 tobacco screening); 4 negative (1 cancer screening, 1 well-child, 1 immunizations, 1 tobacco cessation) • Chronic care: 1 positive
Patient Factors	<ul style="list-style-type: none"> • Goals likely to encounter fewer patient barriers (immunizations, tobacco screening): mostly positive • Goals that required modest patient cooperation (e.g., well child visits and cancer screening): mixed • Goals that require significant patient cooperation (e.g., tobacco cessation): negative

Synopses of the available studies. As there were so few available studies, we are able to include synopses of each in this report. Rather than use the original abstracts, which varied in structure and content, we have put each into a uniform format. The eight randomized controlled trials of performance-based payment, presented in alphabetical order by first author, were:

- **Christensen DB, Holmes G, Fassett WE, et al. Influence of a financial incentive on cognitive services: CARE project design/implementation. *J Am Pharm Assoc.* Sep-Oct 1999;39(5):629-639.**

and

- **Christensen DB, Hansen RW. Characteristics of pharmacies and pharmacists associated with the provision of cognitive services in the community setting. *J Am Pharm Assoc.* Sep-Oct 1999;39(5):640-649.**

Setting and Design: This study took place in Washington State from February 1994 – September 1995. Incentives were offered by the Washington State Cognitive Activities and Reimbursement Effectiveness Project to community pharmacies that served primarily

Table 11: Ongoing quality-based purchasing research: Projects in the Rewarding Results initiative

Project Name (Project Director) (Funder)	Patient Population	Health Plan Setting	Control Group	Incentive Structure/ Intervention	Performance Measures	Evaluation Plan	Organization	Project End Date
Bridges to Excellence— Physician Component (F De Brantes) (RWJF)	Privately Insured	HMO	None	Physicians can receive up to \$55/patient for meeting all standards	Structure, process, clinical outcomes and efficiency for primary care providers and three specialties	Assess determinants of participation in the program	General Electric	12/05
Bridges to Excellence— Hospital Component (F De Brantes) (RWJF)	Privately Insured	HMO	None	To be determined: may include additional 2% above usual payment for hospitals in top decile, 1% for hospitals in 2 nd decile	Leapfrog standards and CMS/Premier demonstration	Assess determinants of participation in the program	General Electric	12/05
Blue Cross Blue Shield of Michigan— Rewarding Results (M Ortwine) (RWJF)	All patients	Indemnity and PPO	None	Hospitals can earn up to 4% augmentation of Diagnosis Related Group payments	Condition-specific quality measures (50%), general patient safety practices (40%), community health activities (10%)	Assess determinants of participation in the program	Blue Cross Blue Shield of Michigan	9/05
Massachusetts Health Quality Partners— Rewarding Results (B Rabson) (RWJF)	Privately Insured	HMO	None	Varies among five health plans	HEDIS [®] measures	Catalog the incentives health plans offer as they change; follow provider performance responses	Massachusetts Health Quality Partners	8/05
Excellus— Rewarding Results (K Curtin) (RWJF)	Privately Insured	HMO	Historical controls	Withholds paid out to physicians based on performance plus share of savings from efficiency measures	Clinical quality (esp. HEDIS [®] measures), member satisfaction, efficiency	Assess the impact of incentive on performance	Excellus Health Plan	9/05

Blue Cross of California-Rewarding Results (C Volpe) (RWJF)	Privately Insured	PPO	Contemporaneous control group in different geographic region	Up to \$5,000/physician bonus plus benchmarking information (not public)	HEDIS [®] and Medicare Health Care Quality Improvement measures, several efficiency and administrative measures	Assess impact of incentive on performance and determinants of changing practice management in response to the program	WellPoint Health Networks	12/05
Integrated Healthcare Association—Pay for Performance (A Bowers) (CHCF)	Privately Insured	HMO and POS	Historical controls	Medical groups can receive enhanced fee-for-service or bonus (varies by plan) plus public reporting	Condition-specific quality measures, patient satisfaction, and information technology (weights vary year to year)	Assess impact of incentive on performance and determinants of changing practice management in response to the program	Integrated Healthcare Association	8/05
MediCal Local Initiative (E Payne) (CHCF)	Medicaid	HMO	None	Enhanced fee-for-service or capitation bonus plus provider recognition awards	Access, service, HEDIS [®] scores	To be determined	Center for Healthcare Strategies	8/05
Evaluation of Rewarding Results (G Young) (AHRQ)	Varies across Rewarding Results projects	Varies across Rewarding Results projects	Varies across Rewarding Results projects	Varies across Rewarding Results projects	Varies across Rewarding Results projects	Explain observed variations among sites in response to Rewarding Results incentives by evaluating provider attitudes toward the incentive arrangements (awareness, perception of salience) and the clinical value of the quality targets used.	Boston University	9/06

Key: AHRQ =Agency for Healthcare Research and Quality; CHCF = California HealthCare Foundation; CMS = Centers for Medicare & Medicaid Services; RWJF = The Robert Wood Johnson Foundation; HEDIS = Health Plan Employer Data and Information Set; HMO = health maintenance organization; POS = point of service; PPO = preferred provider organization

Table 12: Ongoing quality-based purchasing research: Other QBP projects

Project Name (Project Director) (Funder)	Patient Population	Health Plan Setting	Control Group	Incentive Structure/ Intervention	Performance Measures	Evaluation Plan	Organization	Project End Date
Purchaser/ Provider Evaluation: Hospital Quality Data in Tennessee (B Braun) (AHRQ)	Privately Insured	All commercial	None	Public report	Leapfrog standards	Obtain purchaser and hospital reactions through semi-structured interviews	Park Nicollet Institute	1/04
Determining Whether Pay-for-performance Incentives Improve Health Care Quality in Medical Groups (M Rosenthal) (CMWF)	Privately Insured + Medicare Plus Choice	HMO - commercial and Medicare	Contempora- neous control group in different geographic region	Medical groups can receive bonuses over capitation plus public reporting	Condition-specific quality measures, general patient safety practices, patient experience/ satisfaction, and information technology (weights vary year to year)	Assess the impact of incentive on performance (measuring performance using indices included among the pay-for- performance indicators and indices that are not included, to look for spillover effects)	Harvard School of Public Health	1/05
The Impact of Provider Performance Reporting on Consumer and Physician Organization Behavior (M Rosenthal) (RWJF)	Privately Insured + Medicare Plus Choice	HMO - commercial and Medicare	National contempora- neous control group	Reporting medical group performance to health plan enrollees	Condition-specific quality measures, general patient safety practices, patient experience/ satisfaction, and information technology (weights vary year to year)	Assess the impact of incentive on medical group performance and on patients' choices of medical group	Harvard School of Public Health	10/04
The Changing Patterns and Impact of Value Based Purchasing: State Medicaid Agencies (A Epstein) (RWJF)	Medicaid	Varies	Contempora- neous control groups in commercial plans from same region	Varies by State	Varies by State	Describing QBP activities in Medicaid programs and assessing the impact of the various incentive strategies	Harvard School of Public Health	11/04

The Patterns and Impact of Value-based Purchasing (A Epstein) (AHRQ)	Privately Insured	HMO	None	Varies across plans and markets	Varies across plans and markets (study focuses on five condition-specific measures plus patient satisfaction; looks at others too)	Describing QBP activities and health plan responses in 40 markets	Harvard School of Public Health	9/06
Assessment of the Effectiveness of Consumer-driven Health Plans for Consumers (S Sofaer) (RWJF)	All enrollees in consumer-driven health plans	Potentially relevant to all	N/A	All users of information materials and decision support tools	Varies by the vendor or health plan providing the information	Evaluating the tools themselves, not the response of providers to the tools	City University of New York	7/04 (will extend)
A National Center for Value Purchasing Methods (M Callahan) (AHRQ)	All	All	None	Multiple	Varies by health plan or purchasers	Case studies of provider group management response to incentive approaches	Healthfront	9/06
Using Incentives to Drive Leaps in Patient Safety (Evaluation of 4 Leapfrog pilot programs) (S Delbanco) (AHRQ)	Privately Insured	PPO	Contemporaneous controls (intervention--unionized beneficiaries, control--union beneficiaries at same company); also pre/post design for both groups	Reduced co-pays for consumers, public reporting	Leapfrog standards	Assess the impact of incentive on choice of hospital and determinants of consumer response to the program, impact of the benefit design change, and salience of the program and quality information to consumers	Leapfrog	9/06

Key: AHRQ =Agency for Healthcare Research and Quality; CMWF = The Commonwealth Fund; RWJF = The Robert Wood Johnson Foundation; HMO = health maintenance organization; PPO = preferred provider organization; QBP = quality-based purchasing

Table 13 summarizes the six scenarios to be simulated. (See Appendix B, available at www.ahrq.gov/clinic/epcindex.htm, for the simulation algorithm.)

Table 13: The six scenarios simulated

Scenario #	Hypothetical (Defined) World of Hospitals							Grading Function		
	Superior Quality		Good Quality		Poor Quality		Average Number of Patients per Hospital	Mean probability mortality of whole population	Low Trim Point < Labeled superior	High Trim Point > Labeled poor
	True Probability of Mortality	% Total Hospitals	True Probability of Mortality	% Total Hospitals	True Probability of Mortality	% Total Hospitals				
1	Only 2 Groups		15.3%	90%	17.3%	10%	200	1 tail distribution: grade is either “good” or “poor”, i.e. if outcome is > high trim point, which includes 2.5% of population		
	Recreation of Thomas and Hofer model, as starting point.							15.5%	N/A	20.5%
2	13.3%	10%	15.3%	80%	17.3%	10%	200	2 tails: with ~2.5% of population above/below each;		
	Thomas and Hofer model; now with three groups; mortality rate for “superior” calculated using assumption that superior hospitals are as much better than good quality hospitals as poor quality hospitals are worse than good quality hospitals (i.e. rate at superior hospitals = rate at good quality hospitals – (rate at poor quality hospitals – rate at good quality hospitals); also assume 10% of hospitals are superior quality.							15.3%	10.3%	20.3%
3	8.6%	10%	12.2%	80%	17.1%	10%	200	2 tails: with ~2.5% of population above/below each; mortality outcomes above high trim point labeled “poor,” below low trim point labeled “superior.”		
								12.1%	7.6%	16.6%
	Mortality values from California AMI study (see text), using Thomas and Hofer hospital group proportions.									
4	8.6%	10%	12.2%	80%	17.1%	10%	100	2 tails: with ~2.5% of population above/below each		
								12.1%	5.7%	18.5%
	As above except number of patients per hospital = 100									
	8.6%	10%	12.2%	80%	17.1%	10%	100	2 tails: with ~10% of population above/below each		
5								12.1%	7.9	16.3
	As above; number of patients per hospital = 100									
6	8.6%	10%	12.2	80%	17.1	10%	400	2 tails: with ~10% of population above/below each trim point.		
								12.1%	10.0%	14.2%
	As above; number of patients per hospital = 400									

5. Results of Simulations To Assess the Usefulness of Outcomes Reports

Scenario 1: Reproducing Thomas and Hofer

In this chapter, we will describe the key findings from our simulations. (See Appendix C, available at www.ahrq.gov/clinic/epcindex.htm, for a fuller description of all the results from all of the simulations.)

For this scenario, we reproduced in our model the assumptions of Thomas and Hofer. The probability of death at *poor* and *good* hospitals was calculated as in their model as described in an unpublished appendix to their paper. The scenario is summarized by Figure 5 and Figure 6 above, and Table 14 and Table 15, below.

Notice that in this scenario, a fairly large part of the *poor* quality hospital distribution is intersected by the trim point (Figure 6). Examining the areas under the *good* quality and *poor* quality hospital curves, to the right of the trim point, it appears that some hospitals that are labeled *poor*, may in fact be of *good* quality. This error is called predictive error, and is reported in Table 14. Other predictive values—positive predictive value (the chance that a hospital which received a *poor* grade is actually a *poor* quality hospital) and negative predictive value (the chance that a hospital receiving a *good* grade is actually a *good* quality hospital)—are shown as well. In the calculation of predictive values, the proportion of the two populations is important. The more rare the condition or state of being “positive” is (in this case, being a *poor* quality hospital), the higher the positive predictive value will tend to be. Since the *poor* quality hospitals only comprise 10% of the population, and their distribution is nearly subsumed by the *good* quality hospitals, it is not surprising that the positive predictive value is so low, and the inversely-related predictive error is so high.

Table 14: Scenario 1: Predictive values, year 1

Score assigned	Hospital really is--	Probability in whole distribution	Probability within this group of scores	2 category test clinical test labels
Poor	Poor	1.1%	38.7%	Positive predictive value
	Good	1.8%	61.3%	Predictive error
	<i>Subtotal</i>	2.9%		
Good	Poor	8.9%	9.1%	Negative predictive value
	Good	88.2%	90.9%	
	<i>Subtotal</i>	97.1%		

Other metrics of test performance are sensitivity (the probability that a hospital that is actually *poor* will be labeled *poor*) and specificity (the probability that a hospital that is actually *good* will be labeled *good*). The measures are independent of the population (or, in this case,

hypothetical world of hospitals) in which they are used. They are measures of the tests themselves, and can be used to compare one test with another. Table 15 shows sensitivity and specificity for scenario 1.

Table 15: Scenario 1, year 1: Sensitivity and specificity calculations

Hospital really is--	Score assigned	Probability in whole distribution	Probability within this group of hospitals	2 category test clinical test labels
Poor	Poor	1.1%	11.2%	Sensitivity
	Good	8.9%	88.8%	
	<i>Subtotal</i>	10.0%		
Good	Poor	1.8%	2.0%	Specificity
	Good	88.2%	98.0%	
	<i>Subtotal</i>	90.0%		

We can see that while the evaluation function will correctly label 98% of *good* hospitals as *good*, it will detect only 11.2% of *poor* quality hospitals in any given year, using Thomas and Hofer's assumptions.

Following is a discussion of assessing the evaluation system over multiple years of use.

The results for calculating *star* scores for 2 years are shown in Table 16 and Table 17. While predictive values, sensitivity, and specificity are generally defined for tests/functions with dichotomous results, the approach of each can be used with more than one possible outcome. We will examine the predictive value and sensitivity and specificity of the most extreme grades: 2 *stars* and 4 *stars* over 2 years.

Table 16: Scenario 1: Probability, given that a hospital has received two, three, or four stars over 2 years, that it is good vs. poor

Number of stars (over 2 years)	Probability of actually being poor is--	Probability of actually being good is--	Overall probability of receiving score
2	78.2%	21.8%	0.2%
3	36.4%	63.6%	5.4%
4	8.4%	91.6%	94.4%

For example, the positive predictive value of 2 *stars* is 78.2%—a large improvement over the 1-year figure of 38.7%, although only a small set of hospitals will be assigned this grade (0.2%); 4 *stars* has a negative predictive value of 91.6%; 3 *stars* has poor discrimination between subgroups, although a hospital in this group is more than three times more likely to truly be poor than if one selected a hospital without any performance information (this would be essentially random and would have a 10% chance of yielding a poor hospital, since they are 10% of the general population, but 36.4% of the population receiving 3 stars).

Sensitivity and specificity calculations show that specificity of *4 stars* is 96.1% and sensitivity of *2 stars* is only 1.2%, as 2 stars is very unlikely in this scenario, whether the hospital is poor or good.

Table 17: Scenario 1: Expected score distribution over 2 years

What hospital <i>really</i> is	Probability (%) hospital will receive score of--			Overall probability of being in this group
	2 stars	3 stars	4 stars	
Poor	1.2%	19.8%	78.9%	10.0%
Good	0.0%	3.8%	96.1%	90.0%

The results for 3 years of testing in this scenario are shown graphically in Figure 7 and by hospital group in Table 18. Hospitals with 3 or 4 stars are almost certainly of *poor* quality—but these scores are rare. Indeed, it is a rare thing to be graded *poor* in this scenario, and to have it occur even once in 3 years happens for only 8.2% of hospitals.

Figure 7: Scenario 1: Percentage of good vs. bad hospitals by 3-year star score

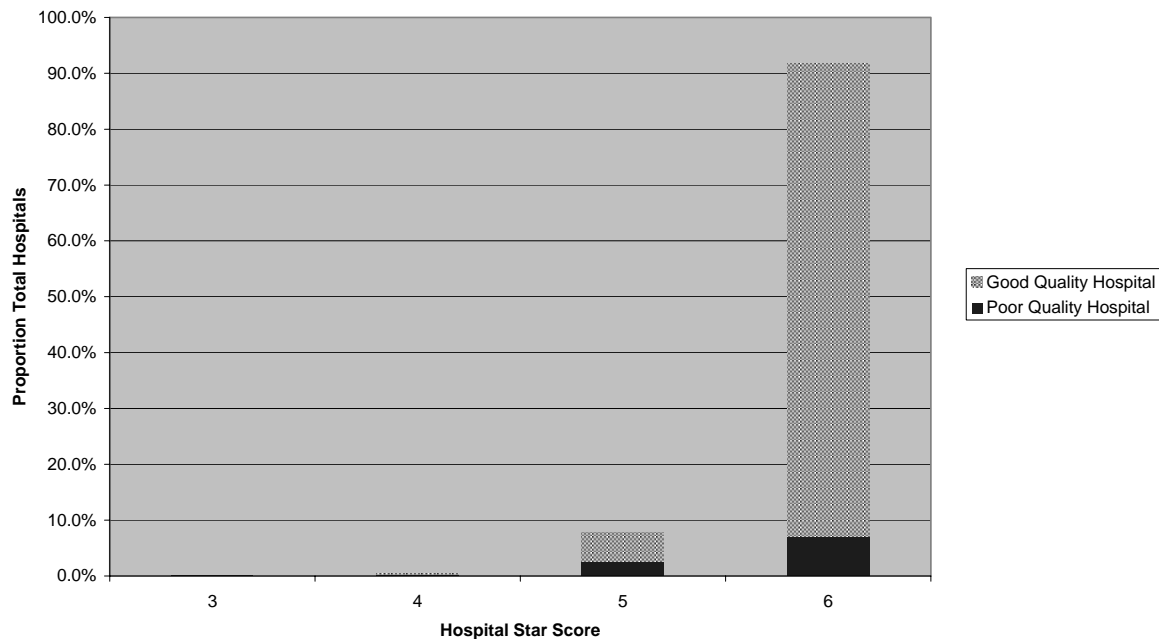


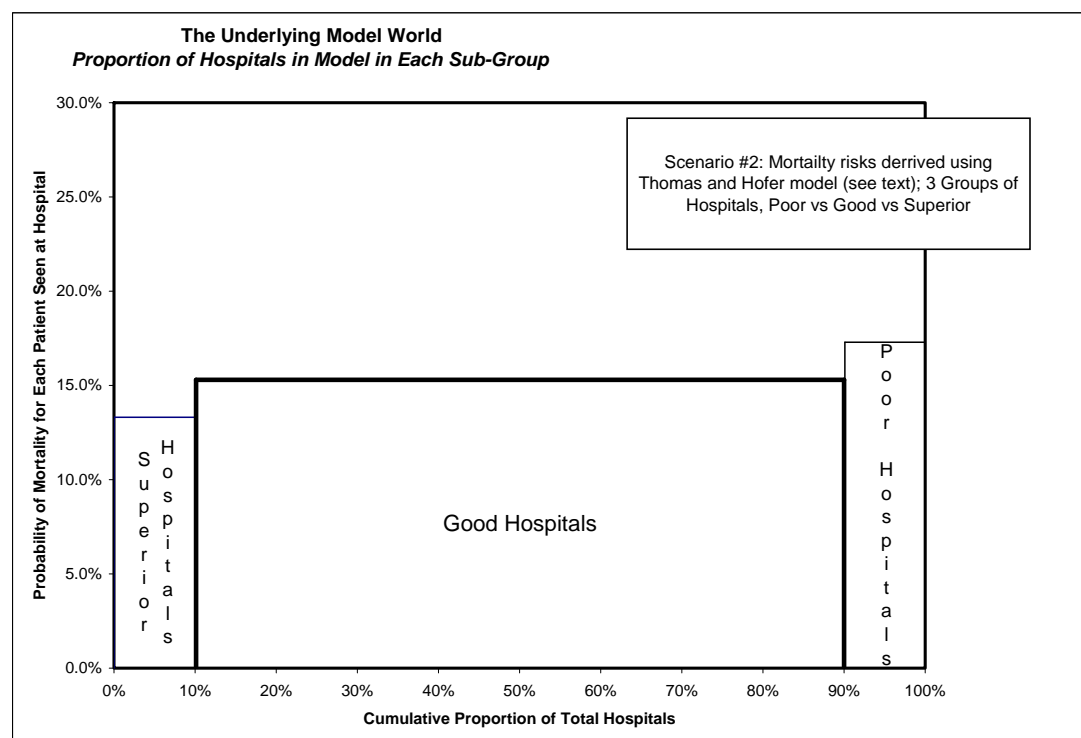
Table 18: Scenario 1: Expected score distribution for good vs. poor hospitals over 3 years

What hospital really is	Probability (%) hospital will receive score of--			
	3 stars	4 stars	5 stars	6 stars
Poor	0.1%	3.3%	26.4%	70.1%
Good	0.0%	0.1%	5.7%	94.2%

Scenario 2: Adding Another Hospital Category

For this scenario, we added the *superior* quality hospital group as 10% of the hypothetical hospital population. The average mortality rate for *superior* hospitals was assumed to be the same percentage difference below the mean performance as Thomas and Hofer's *poor* quality hospitals were above the mean (that is, mortality rates were assumed to be 13.3%, 15.3%, and 17.3% for *superior*, *good*, and *poor* hospitals, respectively, Figure 8). This assumption about *superior* hospitals is arbitrary and meant simply to be approximately as conservative Thomas and Hofer's original assumptions.

Figure 8: Scenario 2: Hypothetical world of hospitals



The trim points were calculated using the normal distribution based on the average mortality rate with trim points defined so that 2.5% of hospitals would lie under the curve beyond each trim point (in a normal distribution with standard deviation defined by the number of patients per