APPENDIX A. SEARCH STRATEGIES

The database was Web of Science; the sub-databases were Science (SCI-EXPANDED), Social Science (SSCI), Arts & Humanities (A&HCI) and the Science & Social Sciences Proceedings (CPCI-S & CPCI-SSH).

40 Cited Author=(epstein a*) AND Cited Year=(2000)
Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

PMID- 10770153
TI - Public release of performance data: a progress report from the front.
AU - Epstein AM
PT - Comment
PT - Editorial
SO - JAMA. 2000 Apr 12;283(14):1884-6.

268 Cited Author=(epstein a*) AND Cited Year=(1998)
Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

PMID- 9624015
TI - Rolling down the runway: the challenges ahead for quality report cards.
AU - Epstein AM
SO - JAMA. 1998 Jun 3;279(21):1691-6.

197 Cited Author=(schneider e*) AND Cited Year=(1996)
Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

PMID- 8657242
TI - Influence of cardiac-surgery performance reports on referral practices and access to care. A survey of cardiovascular specialists.
AU - Schneider EC
AU - Epstein AM
SO - N Engl J Med. 1996 Jul 25;335(4):251-6.

180 Cited Author=(schneider e*) AND Cited Year=(1998)
Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

PMID- 9613914
TI - Use of public performance reports: a survey of patients undergoing cardiac surgery.
AU - Schneider EC
AU - Epstein AM
SO - JAMA. 1998 May 27;279(20):1638-42.

104 Cited Author=(fung c*) AND Cited Year=(2008)
Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

Fung, C. H., Y. W. Lim, S. Mattke, C. Damberg and P. G. Shekelle. "Systematic review: the evidence that publishing patient care performance data improves quality of care." Ann Intern Med 148(2): 111-23.

412 Cited Author=(marshall m*) AND Cited Year=(2000)
Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

2000. Marshall, M. N., P. G. Shekelle, R. H. Brook and S. Leatherman. "Use of performance data to change physician behavior." JAMA 284(9): 1079.

2000. Marshall, M. N., P. G. Shekelle, S. Leatherman and R. H. Brook. "The public release of performance data: what do we expect to gain? A review of the evidence." JAMA 283(14): 1866-74.

2000. Marshall, M. N., P. G. Shekelle, S. Leatherman and R. H. Brook. "Public disclosure of performance data: learning from the US experience." Quality in Health Care 9(1): 53-57.

2000. Dying to Know: Public Release of Information about Quality of Health Care by Martin Marshall, Paul G. Shekelle, Robert H. Brook, Sheila Leatherman. RAND MR-1255

APPENDIX B. STUDY SELECTION FORM

1. Included in Fung or 2006 and earlier

Included in Fung ..... STOP
2006 or earlier ..... ☐

2. What types of health care setting are the quality and safety information about?

Health plan/ HMO ..... ☐
Health system ..... ☐
Hospital ..... ☐
Physician/ Individual providers ..... STOP
Other, specify: ..... _____

3. Which Key Question* does this article address?

KQ1 ..... ☐
KQ2 ..... ☐
KQ3 ..... ☐
KQ4 ..... ☐
None ..... STOP
Background ..... STOP

* KQ1: What is known about the most effective way of displaying quality and safety information, comparative data about health system structure, services, and performance so that it is understandable?
KQ2: How do patients prefer to receive or access this information?
KQ3: What is the evidence that patients or their families use publicly reported quality and safety information to make informed health care decisions?
KQ4: What is the evidence that public reporting of quality and safety information leads to improved quality or safety?

4. What is the study design?

RCT ..... ☐
Observational, concurrent comparison ..... ☐
Observational, time series (no concurrent) ..... ☐
Observational, other ..... ☐
Systematic Review ..... ☐
Non-systematic review, commentary or news, other ..... STOP
Misc include ..... _____

5. Does this article discuss one of the following report cards/ reported data?

New York State Reporting System ..... ☐
CAHPS ..... ☐
HEDIS ..... ☐
Cleveland ..... ☐
Wisconsin ..... ☐
Medicare compare ..... ☐
California ..... ☐
Other, specify: ..... _____

6. What country is the data from?

US ☐
Europe ☐
Canada ☐
Australia/New Zealand ☐
Other, specify: _____
Unclear/not stated ☐

7. What level do the data come from?

National or sufficiently representative ☐
Regional ☐
Single state ☐
City/county ☐
Single medical center ☐
Unknown ☐

What outcomes are reported?

8. Individual-level outcomes

Health/clinic outcomes ☐
Patient selection of plan or provider ☐
Patient satisfaction ☐
Provider satisfaction ☐
Patient-provider communication ☐
Self-management ☐
Adherence (medication, visit) ☐
Provider practice patterns ☐
Harms or benefits ☐
Other, specify: _____

9. System-level outcomes

Quality improvement activity ☐
Change in quality rating/scores ☐
Efficiency ☐
Privacy breaches ☐
Patient safety ☐
Attitudes ☐
Usability ☐
Harms or benefits ☐
Other, specify: _____

APPENDIX C. EVIDENCE TABLES FOR KEY QUESTIONS #1 AND #2

Author, Year (ID)	Objective	Subject of public reporting; Hospital/ Health plan; Location	Sample	Design Type	Design Rating; Global Rating	Key Findings
Mazor, 2009[20]	To evaluate consumers' responses to different approaches to public reporting of comparative hospital data on HAIs.	Healthcare-associated Infections (HAIs); Hospital; Worcester, MA	Random sample of residents; 201 completed surveys; Response rate 34% of those sent to valid addresses; 25% of all selected addresses. Age: Mean 51.7; 37.8% male; 28.4% HS education or less	Experiment (random sample; random assignment of versions of mock report)	4;2	The three report characteristics tested (consistency of hospital rating across indicators included; presenting results in words or charts, and including confidence intervals) had no significant impact on understandability. More respondents with a higher level of education (at least some college) rated 2 sections of the report as easier to understand verse those with high school education or less. For the other 5 sections the differences between education levels were not statistically significant. Age was not found to effect understandability.
Mazor, 2009[21]	To understand consumer response to public reports and how reports might be improved.	HAIs; Hospital; Worcester, MA	Random sample of residents; 59 participants; 22 (37.3%) male; age range 24-82; mean 53.3; 25.4% high school education or less	In-depth interviews; random sample of residents invited to participant.	2;2	Most respondents had no prior knowledge/ understanding of HCI and this required explanation before different reports could be discussed. Inconsistent rankings across hospitals made it difficult for interviewees to pick the 'best' hospital. Format: number preferred over symbols; confidence intervals were confusing; interviewees unable to paraphrase definition of risk adjustment provided in the mock reports; print preferred to internet.
Richard, 2005[22]	To use qualitative interviews to understand patients' views of report cards on cardiac care.	Cardiac report cards; Hospital; Canada	7 cities with major cardiac programs; 91 cardiac patients	63 individual interviews and 6 focus groups	2;2	Participants endorsed the idea of report cards, wanted to see improvement in report card scores over time, and would use them if relevant. Patients wanted report cards to contain additional information to supplement the traditional outcomes (e.g., mortality, morbidity), specifically the experiences of other cardiac patients and non-medical aspects of care. Dissemination ideas were varied and included important roles for family physicians and cardiologist to provide and explain the report cards.

Design ratings: 4 stars indicate a strong study design rating; while 1 star indicates a weaker study design rating.
Global ratings: 3 indicates great weight in the stratum's body of evidence; and 1 indicates little weight.

APPENDIX D. EVIDENCE TABLES FOR KEY QUESTION #3

Author, Year (ID)	Objective	Subject of public reporting; Hospital/Health plan; Location	Sample	Design Type	Design Rating; Global Rating	Key Findings
Articles from Fung and colleagues						
Farley, 2002[37]	To assess effects of providing CAHPS information on plan choices	CAHPS; Health plan; New Jersey	HMO Medical Plans in New Jersey; Medicaid beneficiaries (1998)	Randomized controlled trial	4;3	No effect on HMO choices overall; Participants who read the report card and did not select the dominant HMO chose the HMO with higher CAHPS scores.
Farley, 2002[38]	To assess effects of providing CAHPS information on plan choices	CAHPS; Health plan; Iowa	HMO Medical Plans in Iowa; Medicaid Beneficiaries (2000)	Randomized controlled trial	4;2	No effect on HMO choices overall
Spranca, 2000[49]	To assess effects of providing CAHPS information about hypothetical health plans on plan choices	Hypothetical plans; Health plan; Los Angeles	Hypothetical plans in laboratory setting; adults with private insurance	Experimental study	4;2	When plans had high CAHPS ratings, participants were willing to enroll in less expensive plans that restrict services
Harris, 2002[39]	To investigate the impact of expert-assessed and consumer-assessed quality ratings on willingness to enroll in hypothetical health plans that restrict provider access	Hypothetical plans; Health plan; Los Angeles	Laboratory setting; Privately insured adults (2000)	Experimental study	4;2	Provision of report cards with information about quality of health plan reduced importance of provider network features
Beaulieu, 2002[35]	To assess effects of providing health plan performance data (HEDIS measures, patient satisfaction) on consumers' enrollment decisions	HEDIS; Health plan; Harvard University	Private health plans available to Harvard employees; Harvard employees (1994 to 1997)	Observational cohort	3;2	Provision of quality information had a small but statistically significant effect on health plan choices.
Wedig, 2002[45]	To assess effects of providing quality ratings from the Federal Employee Health Benefit guide on consumers' plan choices	Federal Employee Health Benefit guide; Health plan; U.S.	Private health plans available to federal employees; Federal employees with single person HMO coverage residing in counties with 5 or fewer unique plans (1995 to 1996)	Observational cohort	3;2	Dissemination of report cards influenced plan selection. Employees were more likely to select plans with better quality ratings.

Design ratings: 4 stars indicate a strong study design rating; while 1 star indicates a weaker study design rating.
Global ratings: 3 indicates great weight in the stratum's body of evidence; and 1 indicates little weight.

Author, Year (ID)	Objective	Subject of public reporting; Hospital/Health plan; Location	Sample	Design Type	Design Rating; Global Rating	Key Findings
Jin, 2005[30]	To assess effects of providing quality ratings from the Federal Employee Health Benefit guide on plan choices	Federal Employee Health Benefit guide; Health plan; U.S.	Private health plans serving federal employees; Federal employees, retirees, and surviving family of deceased federal employees (1998-1999)	Observational cohort	3;3	Overall, inertia in health plan enrollment decisions. For individuals affected by performance ratings, better scores were associated with increased likelihood of selecting the plan.
Scanlon, 2002[43]	To assess effects of providing HEDIS and patient satisfaction ratings on plan choices	HEDIS; Health plan; General Motors;	General Motors employees (1996-1997); Private health plans (HMO only)	Observational cohort	3;3	Employees avoided plans with many below average ratings and would be willing to pay more to avoid plans with lower ratings, but were not strongly attacted to plans with many superior ratings.
Mennemeyer, 1997[52]	To assess the relationship between the release of HCFA hospital-specific mortality rates and utilization (discharges); to compare the impact of releasing HCFA mortality rates to press reports of unexpected deaths, on utilizations.	HCFA; Hospital; U.S.	Community hospitals treating Medicare patients (1984-1992)	Observational cohort	3;2	Hospitals with mortality rates two times that expected by HCFA had less than one fewer discharge per week in the first year; press reports of single, unexpected deaths was associated with 9% reduction in hospital discharges within one year.
Vladeck, 1988[55]	To examine relationship between mortality rate outlier status and hospital CABG volume/quality improvement activity following CSRS implementation	NYS CSRS; Hospital; New York	All New York general acute hospitals serving Medicare patients (~1985 to ~1986)	Analysis of Time Trend	2;1	No significant effect on occupancy rates
Mukamel, 1998[51]	To measure the relationship between provider (hospital, physician) ratings in the CSRS and rates of growth in fee-for-service market share	NYS CSRS; Hospital; New York	All New York hospitals performing CABG (1990 to 1993)	Observational cohort	2;1	Hospitals with better outcomes experienced higher rates of growth in market share
Hannan, 1994[54]	To determine if mortality rate outlier status was associated with overall improvement in risk-adjusted mortality and changes in provider volume of CABG operations performed following the implementation of the CSRS	NYS CSRS; Hospital; New York	All New York hospital performing CABG (1989 to 1992)	Observational cohort	3;2	No association between mortality rate outlier status and hospital volume

Design ratings: 4 stars indicate a strong study design rating; while 1 star indicates a weaker study design rating.
Global ratings: 3 indicates great weight in the stratum's body of evidence; and 1 indicates little weight.

Author, Year (ID)	Objective	Subject of public reporting; Hospital/Health plan; Location	Sample	Design Type	Design Rating; Global Rating	Key Findings
Chassin, 2002[36]	To examine relationship between mortality rate outlier status and hospital CABG volume/quality improvement activity following the CSRS implementation	NYS CSRS; Hospital; New York	New York hospitals with the highest and lowest CABG mortality from 1989-1995	Analysis of Time Trend	2;1	Small changes in market share and less than half the time in the expected direction
Jha, 2006[29]	To examine the relationship between providers' CSRS rankings and market share; to examine impact of cardiac surgeons' performance on the likelihood of ceasing practice in New York	NYS CSRS; Hospital; New York	All New York hospitals performing CABG for more than 3 years (1989 to 2002)	Time Series (for market share analysis)	3;2	No significant relationship between ranking and subsequent market share
Baker, 2003[34]	To examine market share following the release of risk-adjusted 30-day mortality rates for six acute conditions as part of the CHQC program	CHQC; Hospital; Northeast Ohio	30 nonfederal hospitals (1991 to 1997)	Time Series	3;2	No statistically significant relationship overall between higher than expected mortality rates and market share
Romano, 2004[33]	To examine the relationship between outlier status in California & New York public reports in three conditions/ procedures (CABG mortality in New York, AMI and postdiskectomy complications in California) and hospital volume	NYS CSRS and CA; New York and California	All licensed hospitals in New York State performing CABG , non-federal hospitals in California except Kaiser hospitals and state developmental and correctional hospitals	Time Series	3;2	No significant AMI-related volume changes among outlier hospitals. Slight increase in lumbar diskectomy-related volume for low-complication outliers. Significant transient increase in CABG volume for low-mortality hospitals and transient decrease in volume for high-mortality outliers.
Hibbard, 2005[4]	To compare the impact of public (QualityCounts), internal (private) and no reporting on quality improvement activity, market share, and risk-adjusted performance (three clinical areas--hip/knee surgery, cardiac care, and obstetric care)	QualityCounts; Hospital; South central Wisconsin	Hospitals participating in Quality Counts; 24 Hospitals	Analysis of Time Trend	2;2	No significant changes in market share for hospital with publicly-reported data. No results given for internal or no reporting groups.

Design ratings: 4 stars indicate a strong study design rating; while 1 star indicates a weaker study design rating.
Global ratings: 3 indicates great weight in the stratum's body of evidence; and 1 indicates little weight.

Author, Year (ID)	Objective	Subject of public reporting; Hospital/Health plan; Location	Sample	Design Type	Design Rating; Global Rating	Key Findings
New articles not in Fung and colleagues						
Cutler , 2004[32]	To examine whether where patients go for bypass surgery (the distribution of patients across providers) affected by report cards	NYS CSRS; Hospital, New York State	All hospitals performing bypass surgery in New York (3,406 patients in the baseline year)	Observational, time series--across hospital rather than statewide trends.	3;3	Hospitals identified as high-mortality by the report experienced an approximated 10% decline in bypass surgery (4.9 fewer patients with hospital averages of 50 surgeries per month, significant at the 0.5 level); while low mortality hospitals do not experience an increase. The reduction is in low-severity, not high severity patients.
Harris, 2002[40]	To determine if consumers perceive the quality of health plans and how quality relates to their choice of health plan.	N/A; Health plan; Minneapolis and St. Paul, MN	Randomly-selected eligible employees interviewed by phone. 721 interviewed. 91% response rate. Limited to unmarried employees with no dependents	Observational: cross sectional	3;2	Incorporating information from consumers about how important to them different attributes of health plans are improves models that explain health plan choice.
Dixon, 2008[27]	To examine the influence of health plan (consumer driven health plan versus preferred provider organization) on the use of health-related information and health services	N/A; Health plan; Large manufacturing company	Health plan/HMO; US; Employees of a large manufacturing company	Observational, time series (no concurrent)	2;1	Enrollees in lower-deductible CDHP were most likely to start using information. Enrollees in high-deductible CDHP were more likely to use cost information than PPO enrollees. Variation in information seeking decreased throughout study.

Design ratings: 4 stars indicate a strong study design rating; while 1 star indicates a weaker study design rating.
Global ratings: 3 indicates great weight in the stratum's body of evidence; and 1 indicates little weight.

APPENDIX E. EVIDENCE TABLES FOR KEY QUESTION #4

Author, Year (ID)	Objective	Subject of public reporting; Hospital/Health plan; Location	Sample	Design Type	Design Rating; Global Rating	Key Findings
Impact on Quality Improvement Activity Articles from Fung						
Chassin, 2002[36]	To examine relationship between mortality rate outlier status and hospital CABG volume/quality improvement activity following the implementation of the CSRS	NYS CSRS; Hospital; New York State	Key informants at four hospitals and state officials directly involved in efforts to quality improvement efforts at the hospitals	Case Series	1;1	Increase in quality improvement activity (e.g., staffing policy changes, multidisciplinary approach to examining care processes, changes in operating room schedule)
Dziuban, 1994[59]	To document a hospital's response to being identified as a high risk-adjusted mortality outlier in the CSRS	NYS CSRS; Hospital; New York State	One outlier hospital	Case Study	1;1	Quality improvement activity increased (change in timing & technique used for patients undergoing emergent CABG, change in hospital policies)
Bentley, 1998[60]	To determine whether Pennsylvania Health Care Cost Containment Council's Consumer Guide to CABG, which compared in-hospital mortality rates, led to more changes in Pennsylvania hospitals' CABG policies/practices than in New Jersey hospitals, which were not required to publicly-report performance results	Pennsylvania consumer guide; Hospital; Pennsylvania and New Jersey	Key informants at the hospitals identified by the chief executive officers of these hospitals; Hospitals providing CABG surgery	Survey (Descriptive)	1;1	Response in Pennsylvania hospitals (e.g., recruited staff, started continuous quality improvement program to improve CABG procedures). More changes in Pennsylvania than New Jersey hospitals (no formal statistical testing because small sample size)
Hibbard, 2003[58]	To compare the effects of public reporting (QualityCounts) to confidential reporting and no reporting, on quality improvement activity, market share (hospital discharges), and risk-adjusted performance (two summary indices of adverse events and indices in three clinical areas--hip/knee surgery, cardiac care, and obstetric care)	QualityCounts; Hospital; South central Wisconsin	Hospitals participating in Quality Counts (n=24)	Controlled Before/After Trial	3;1	Compared to hospitals that received confidential reports or no reports, QualityCounts hospital did not engage in more quality improvements overall, but they did engage in a statistically higher number of quality improvement efforts specific to the areas included in the reports.

Design ratings: 4 stars indicate a strong study design rating; while 1 star indicates a weaker study design rating.
Global ratings: 3 indicates great weight in the stratum's body of evidence; and 1 indicates little weight.

Author, Year (ID)	Objective	Subject of public reporting; Hospital/Health plan; Location	Sample	Design Type	Design Rating; Global Rating	Key Findings
Hibbard, 2005[4]	To compare the impact of public (QualityCounts), internal (private) and no reporting, on quality improvement activity, market share (hospital discharges), and risk-adjusted performance (two summary indices of adverse events and indices in three clinical areas--hip/knee surgery, cardiac care, and obstetric care)	QualityCounts; Hospital; South central Wisconsin	Hospitals participating in Quality Counts (n=24)	Descriptive (survey) (for quality improvement analysis)	1;1	Out of seven possible activities, mean number of quality improvement activities was 4.1 overall; 5.7 for hospitals with improved ratings; 2.6 with no change in ratings; 4 with decrease in ratings (no formal statistical testing)
Rosenthal, 1998[62]	To study quality improvement activities following release of CHQC reports of mortality rates, length of stay, and cesarean section rates (all measures severity-adjusted)	CHQC; Hospital; Cleveland	One academic and three community hospitals of varying size in the Cleveland area	Case Series	1;1	Quality improvement activities increased (e.g., interdisciplinary process improvement teams, detailed review of processes of care, development of practice guidelines)
Tu, 2003[61]	To study the impact of the "Cardiovascular Health and Services in Ontario: AN ICES Atlas," which reports hospital-specific acute myocardial infarction performance measures, on quality improvement activity	ICES; Hospital; Ontario, Canada	All Ontario hospitals providing acute myocardial infarction care; Physicians working in Ontario hospitals representing 62 of 121 eligible hospitals (52% overall hospital response rate)	Descriptive (survey)	1;1	54% of respondents indicated that one or more changes were made at their hospital
Longo, 1997[63]	To examine the impact of Missouri Department of Health's obstetrics consumer report, which provides structure, process, and outcomes measures, on quality improvement activity and clinical outcomes	MO Dept. Health obstetrics consumer report; Hospital; Missouri	All hospitals providing obstetric care; Key informant designated by hospital administrators at 82 hospitals (93% response rate)	Descriptive (survey)	1;1	Hospitals instituted services (e.g., hospital policy for that infants ride in car seats upon discharge, formal neonatal transfer agreements) after the reports were published
Luce, 1996[65]	To describe quality improvement activity following the California OSHPD's CHOP report featuring risk-adjusted outcomes	OSHPD CHOP; Hospital; California	All California non-federal hospitals; 17 out of 22 public hospitals that are members of the California Association of Public Hospitals and Health Systems	Descriptive (survey)	1;1	Minimal impact on quality improvement activity

Design ratings: 4 stars indicate a strong study design rating; while 1 star indicates a weaker study design rating.
Global ratings: 3 indicates great weight in the stratum's body of evidence; and 1 indicates little weight.

Author, Year (ID)	Objective	Subject of public reporting; Hospital/Health plan; Location	Sample	Design Type	Design Rating; Global Rating	Key Findings
Rainwater, 1998[66]	To describe the impact of publicly reporting California's CHOP risk-adjusted 30-day inpatient mortality rates for patients with acute myocardial infarction, on quality improvement activity	OSHPD CHOP; Hospital; California	California non-federal acute care hospitals; 39 key informants at a sample of acute care hospitals in California	Interviews	1;1	Minimal impact on quality improvement activity (2/3 respondents indicated no specific QI activity)
Mannion, 2005[64]	To describe impact of the National Health Service (NHS) star performance ratings on quality improvement efforts	NHS; Hospital; United Kingdom	All hospital trusts; Staff at four low performing hospital trusts and two high performing hospital trusts	Case series	1;1	Ratings transmitted important priorities from central government and helped direct and concentrate front-line resources. Public reporting led to tunnel vision and distortion of clinical priorities and disincentive to improve performance among high-rated organizations.
Impact on Quality Improvement Articles, not in Fung						
Wang, 2010[57]	To examine the impact of report cards on provider volume (hospital and surgeons) and on patient matching with surgeons.	Hospital CABG Volume	Hospitals in PA who perform 30 or more CABG per year between 3rd Q 1998 and 1st Q 2006	Observational Cohort	3; 2	Report cards have no significant impact on hospital surgical volume and do not change the population of patients who have CABG. Report cards have a larger impact on the distribution of healthier patients as opposed to sicker across hospitals. Bad rating takes a year to have an effect on volume which was estimated as a decrease in quarterly CABG cases of about 15%. These were almost all among low severity CABG cases. This effect did not persist past one year.

Design ratings: 4 stars indicate a strong study design rating; while 1 star indicates a weaker study design rating.
Global ratings: 3 indicates great weight in the stratum's body of evidence; and 1 indicates little weight.

Author, Year (ID)	Objective	Subject of public reporting; Hospital/Health plan; Location	Sample	Design Type	Design Rating; Global Rating	Key Findings
Rainwater, 2005[56]	To evaluate the use and impact of California's Quality of Care Report Card (QRC), based on three questions: 1. Do consumers use the QRC? 2. How useful to consumers are the quality measures included in the QRC? 3. What is the impact of the QRC on quality improvement and other activities in the participating HMOs and medical groups?	California's Quality of Care Report Card (QRC); Health Plan; California	6 consumer focus groups, 2,341 respondents to mail and internet surveys, 56 key informants	Mixed methods: focus groups, surveys, interviews	3;3	Use is reported at over 28,000 visitors to the QRC website annually, and over 100,000 booklets distributed. Users are most interested in comparing HMOs in the plan service domain, and find features like the specialty care information, specific measures such as mental health care, and comparative performance information by health topic or disease most helpful.
Impact on Clinical Outcomes Articles From Fung						
Hannan, 1994[71]	To assess changes in in-hospital mortality rates of CABG patients following the publication of mortality data in the CSRS	NYS CSRS; Hospital; New York	All New York hospitals performing CABG; 57187 patients undergoing CABG (1989-1992)	Analysis of Time Trend	2;2	RAMR decreased from 4.17% to 2.45%.
Dziuban, 1994[59]	To document a hospital's response to being identified as a high risk-adjusted mortality outlier in the CSRS	NYS CSRS; Hospital; New York	One poor performing hospital	Case Study	1;1	Excess mortality was localized to high-acuity patients undergoing emergent CABG. Mortality decreased to zero following focused effort to optimize management of these patients.
Hannan, 1994[54]	To determine if mortality rate outlier status was associated with changes in CABG-related in-hospital risk-adjusted mortality rates following the implementation of the CSRS	NYS CSRS; Hospital; New York	All New York hospitals performing CABG; All New York patients discharged after CABG (1989 to 1992)	Analysis of Time Trend	2;2	Reductions in RAMR, especially among hospitals that had highest initial mortality rates. Convergence in risk-adjusted mortality rates among hospitals initially identified as high, medium, and low performers.
Peterson, 1998[72]	To examine the impact of the CSRS on in-hospital mortality rates by comparing unadjusted mortality rates in New York to other states. To examine the impact of the CSRS on in-state access to CABG and referral out-of-state of patients in need of CABG	NYS CSRS; Hospital; New York	All hospitals performing CABG; Medicare patients 65 or older who underwent CABG in a U.S. hospital (1987 to 1992)	Observational cohort	3;3	Both unadjusted and risk-adjusted mortality rates in New York declined more than in other states. NY MI patients were less likely to receive CABG, but the overall percentage of NY MI patients receiving CABG rose, paralleling national trends, even among higher risk elderly subsets; out-of-state CABG rates declined

Design ratings: 4 stars indicate a strong study design rating; while 1 star indicates a weaker study design rating.
Global ratings: 3 indicates great weight in the stratum's body of evidence; and 1 indicates little weight.

Author, Year (ID)	Objective	Subject of public reporting; Hospital/Health plan; Location	Sample	Design Type	Design Rating; Global Rating	Key Findings
Ghali, 1997[73]	To compare trends in CABG-related mortality in Massachusetts (a state without statewide public reporting of CABG outcomes) to New York (a state with public reporting) and northern New England	NYS CSRS; Hospitals; New York and Massachusetts	All NY hospitals performing CABG; 12 Massachusetts hospitals performing cardiac surgery (except Veterans Affairs hospitals) and hospitals contained in the HCFA hospital 30-day unadjusted mortality dataset (1990, 1992, and 1994)	Observational cohort	3;2	RAMR reductions in Massachusetts were comparable to mortality reduction in New York and northern New England; unadjusted mortality trends were similar in Massachusetts, New York, northern New England, and the United States
Rosenthal, 1997[77]	To measure changes in hospital mortality that occurred following the implementation of the CHQC reporting initiative, which publicly-released in-hospital mortality rates	CHQC; Hospital; Cleveland	Hospitals in the Cleveland area; 101,060 consecutive eligible discharges with eight diagnoses (acute myocardial infarction, heart failure, obstructive airway disease, gastrointestinal hemorrhage, pneumonia, stroke, CABG, and lower bowel resection) from 30 northeastern Ohio hospitals (1992 to 1993)	Time Series	3;1	Risk-adjusted mortality for most conditions declined from 7.5% to 6.8%, 6.8%, and 6.5% for 3 periods following publication. Declines in mortality rates were significant in weighted linear regression analyses for heart failure (0.50% per period) and pneumonia (0.38% per period)
Baker, 2003[34]	To examine hospitals' market share and 30-day risk-adjusted mortality at hospitals participating in CHQC	CHQC; Hospital; Cleveland	Medicare patients receiving care at these Cleveland-area hospitals (1991 to 1997)	Time Series	3;2	Hospital outlier status was not significantly related to changes in risk-adjusted 30-day mortality between 1991 and 1997.
Clough, 2002[78]	To measure changes in in-hospital mortality rates associated with the implementation of the CHQC reporting initiative	CHQC; Hospital; Cleveland	Hospitals included in the Ohio Hospital Association's inpatient discharge data (1992 to 1995)	Observational cohort	3;2	No statistical difference in rate of decline in combined mortality in Cleveland compared to the rest of the Ohio
Longo, 1997[63]	To examine the impact of Missouri Department of Health's obstetrics consumer report, which provides structure, process, and outcomes measures	MO Dept. Health obstetrics consumer report; Hospital; Missouri	All Missouri hospitals providing obstetrics care (1989 to 1993)	Observational cohort	3;2	Improvements in ultrasound rates, vaginal birth after cesarean rates, and cesarean rates were noted among outlier hospitals

Design ratings: 4 stars indicate a strong study design rating; while 1 star indicates a weaker study design rating.
Global ratings: 3 indicates great weight in the stratum's body of evidence; and 1 indicates little weight.

Author, Year (ID)	Objective	Subject of public reporting; Hospital/Health plan; Location	Sample	Design Type	Design Rating; Global Rating	Key Findings
Hibbard, 2005[4]	To compare the impact of public (QualityCounts), internal (private) and no reporting, on quality improvement activity, market share (hospital discharges), and risk-adjusted performance (two summary indices of adverse events and indices in three clinical areas--hip/knee surgery, cardiac care, and obstetric care)	QualityCounts; Hospital; South central Wisconsin	Hospitals participating in Quality Counts (2001 to 2003, n=24)	Controlled Before/After Trial (for outcomes analysis)	3;2	Performance feedback, whether public or private, was associated with improved performance
Moscucci, 2005[74]	To measure the effect of the New York State PCI report on case selection for percutaneous coronary intervention (PCI) by comparing Michigan's and New York's adjusted and unadjusted in-hospital mortality rates	NYS PCI (CSRS); Hospital; New York and Michigan	All New York hospitals performing CABG; 11,374 patients in a multicenter (eight hospital) PCI database in Michigan and 69,048 patients in a statewide (34 hospital) PCI database in New York (1998 to 1999)	Observational cohort	3;2	Unadjusted mortality rates were significantly lower in New York than Michigan, but adjusted mortality rates were not statistically different.
Omoigui, 1996[75]	To determine if dissemination of CSRS mortality data was associated with outmigration of high-risk patients to undergo treatment at the Cleveland Clinic	NYS CSRS; Hospital; New York and Cleveland	All hospital performing CABG in New York State; 9,442 patients receiving CABG at the Cleveland Clinic (1989 to 1993)	Observational cohort	3;2	Patients from New York State receiving CABG at the Cleveland Clinic had higher RAMR than patients from Ohio, other states, and other countries
Dranove, 2003[76]	To study the effects of public reporting in New York and Pennsylvania	NYS CSRS and Pennsylvania public reporting system; Hospital; New York and Pennsylvania	All New York and Pennsylvania hospitals performing CABG; Medicare beneficiaries and hospitals found in a Medicare claims data set (not specified) and hospitals participating in the American Hospital Association annual survey (1987 to 1994)	Observational cohort	3;2	Report cards shifted CABG use to healthier patients, leading to worse outcomes, especially among sicker patients (defined as higher hospital expenditures and days in hospital)

Design ratings: 4 stars indicate a strong study design rating; while 1 star indicates a weaker study design rating.
Global ratings: 3 indicates great weight in the stratum's body of evidence; and 1 indicates little weight.

Author, Year (ID)	Objective	Subject of public reporting; Hospital/Health plan; Location	Sample	Design Type	Design Rating; Global Rating	Key Findings
Baker, 2002[79]	To examine mortality trends associated with the CHQC program	CHQC; Hospital; Cleveland	Hospitals in the Cleveland area; Medicare patients hospitalized with acute myocardial infarction, heart failure, gastrointestinal hemorrhage, obstructive pulmonary disease, pneumonia, or stroke (1991 to 1999)	Time Series	3;2	Risk-adjusted in-hospital mortality declined significantly for most conditions, but the mortality rate in the early post discharge period rose significantly for most conditions and the 30-day mortality rate declined significantly for only heart failure and obstructive pulmonary disease
Bost, 2001[80]	To compare HEDIS and CAHPS results for plans that publicly report data with those who do not, over a three-year period	HEDIS and CAHPS; Health plan U.S.	Commercial health plans (1997-1999)	Observational cohort	2;1	Technical performance measures and patient experience measures (except communication) were higher for health plans that publicly report data.
McCormick, 2002[81]	To assess the relationship between health plan performance and participation in public reporting programs	HMO commercial health plans; Health plan; U.S.	HMO health plans (1997 to 1999)	Observational cohort	2;2	Lower-scoring plans are significantly more likely than plans with higher-scoring plans to stop disclosing publicly their quality data
Impact on Clinical Outcomes Articles, not in Fung						
Bevan, 2009[70]	To assess the impact of public reporting on the performance of ambulance services	Ambulance service response times; UK	Yearly data from 2000 to 2005	"natural experiment' Comparison of UK countries with the same target but one had reporting and the others did not.	3;2	Response times improved in the countries with public reporting and did not in the others. Examination of potential harms found evidence that some types of gaming occurred (data was changed) but that others types that were suspected (changes in the classification of the event) did not.
Cutler, 2004[32]	To examine whether medical quality among hospitals are affected by report cards	NYS CSRS; Hospital; New York State	All hospitals performing bypass surgery in New York (3,406 patients in the baseline year)	Observational, time series--across hospital rather than statewide trends.	3;2	Hospitals identified as high mortality improve performance in terms of decreased risk-adjusted mortality rates: mortality declined 1.2 percentage points (significant at the 0.01 level) in these low quality hospitals during the 12 months after the reporting.

Design ratings: 4 stars indicate a strong study design rating; while 1 star indicates a weaker study design rating.
Global ratings: 3 indicates great weight in the stratum's body of evidence; and 1 indicates little weight.

Author, Year (ID)	Objective	Subject of public reporting; Hospital/Health plan; Location	Sample	Design Type	Design Rating; Global Rating	Key Findings
Elliott, 2010[68]	To determine if hospitals improved in terms of patient experience over the initial 2 years of public reporting of HCAHPS results	HCAHPS; Hospital; US	Hospital, National CAHPS US 61% of hospitals in 3/08 3,864; 84 % of hospitals in 3/09 3,863 Patient response rate averaged 34%--patients are a random sample of discharges	Observational, Time series, no comparison group	3;2	Hospitals improved in 8 of 9 domains as measured by percent of positive responses (MD communication did not improve). Magnitude of changes was small, but would result in change in ranking. Hospital size and original (both years) vs. later (2nd year only) participation were examined and smaller hospitals who participated later performed better.
Hendriks, 2009[69]	To determine if managed competition and public reporting of quality information is associated with quality improvement in health plans.	National health plans; Health plan; Netherlands	Dutch Health Plans, and Health Plans on a National Level; Random sample of health Plan Members; CQI--based on CAHPS;	Observational, time series, no comparison group	3;1	Plans improved in some domains (health plan information and transparency of copayment, conduct of employees, and general rating and requirements, but not others(access to call center, getting needed help from call center and reimbursement of claims) from 2005 to 2008. Identification of selected domains as areas in need of improvement did not seem to affect whether there was improvement or not.
Kim, 2005[67]	To assess the impact of public release of hospital caesarean rates.	Caesarean Section Rates; Hospital; South Korea	263 hospitals	Observational, time series, no comparison group	2;1	Caesarean rates were 43.0% in 1999. Hospital data for 1999 were published in 2000 and rates declined to 38.6% in 2000 and 39.6% in 2001, which are lower than predicted based on rates for 1985 to 1999 and the first years with any decline. Multiple regression results found that hospitals with higher with higher baseline caesarean rates and higher volume were more likely to decline, while market share and financial incentives were not significantly associated with decline in rates.

Design ratings: 4 stars indicate a strong study design rating; while 1 star indicates a weaker study design rating.
Global ratings: 3 indicates great weight in the stratum's body of evidence; and 1 indicates little weight.

APPENDIX F. CRITERIA USED IN QUALITY ASSESSMENT

Fung and Colleagues' Grading Criteria for Included Studies

Study design ratings:
4 stars indicate a randomized trial or experimental trial;
3 stars indicate a controlled trial, pre-post trial with control (controlled before-after trial);
2 stars indicate a pre-post without control, observational cohort study without multivariable adjustment, cross-sectional study without multivariable adjustment, analysis of time trends without control, or well-designed qualitative study; and
1 star indicates a case series, other qualitative study, or survey (descriptive) study.

Global ratings:
3 indicates great weight in the stratum's body of evidence;
2 indicates moderate weight; and
1 indicates little weight.

AMSTAR Grading Criteria for Systematic Reviews

1. Was an 'a priori' design provided? The research question and inclusion criteria should be established before the conduct of the review.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
2. Was there duplicate study selection and data extraction? There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
3. Was a comprehensive literature search performed? At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion? The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable

5. Was a list of studies (included and excluded) provided? A list of included and excluded studies should be provided.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
6. Were the characteristics of the included studies provided? In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
7. Was the scientific quality of the included studies assessed and documented? 'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
8. Was the scientific quality of the included studies used appropriately in formulating conclusions? The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
9. Were the methods used to combine the findings of studies appropriate? For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?).	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
10. Was the likelihood of publication bias assessed? An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable
11. Was the conflict of interest stated? Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.	☐ Yes ☐ No ☐ Can't answer ☐ Not applicable

APPENDIX G. PEER REVIEW COMMENTS/AUTHOR RESPONSES

Peer Review Comments	Comment	Response
Scope	Thought the scope was too small. The literature in this area is pretty scant and did not need an ESP to tell us that. Would have expanded review to no just patients but organizations, VSO, other health care systems and other federal agencies	The scope was provided to us by the co-sponsor and is not something we can choose now.
	On pages 17 through 19 two long numbered lists are provided and referenced, but it's not clear if the entire lists are quoted verbatim from the original source. (This should be made clear if it is verbatim.)	These are not quoted verbatim, but rather summarized from the original report, which we now indicate.
	This is a superb and comprehensive review, but may short-change both public reporting and the VA, at least according to one authoritative published opinion. Lucian Leape recently concluded that public reporting was "So far, the most powerful method for reducing preventable injuries", and he went on to cite the VA's own NISQUIP program as the most shining example. (Transparency and public reporting are essential for a safe health care system. LL Leape. The Commonwealth Fund Publication 1381, Vol 4: "Perspectives on Health Reform". March 2010. Accepting that data showing that public reporting improves safety may not yet be strong, Leape's comments point out that this approach seems at least to have more potential than many of the alternatives (regulation, alignment of incentives, accreditation). I'd like to see this perspective mentioned in the discussion. I'd also like to see a brief data summary regarding studies that have looked at VA programs specifically.	This comment quotes Lucian Leape as stating public reporting is a "powerful method for reducing preventative injuries." We would agree that public reporting consistently influences providers to meet the criteria being reported. However NSQUIP cannot be used as an example, since NSQUIP is not publicly reported, at least not at the time of the studies documenting improvements due to NSQUIP.
	2nd paragraph, sentence beginning "Public reporting also…may only be known by providers" – awkward sentence	We have rewritten for clarity.
	2nd full paragraph re: hospitals in South Korea – unclear why data from non-English speaking country was included for this key question but not others.	We restricted KQ 1&2 to English speaking countries only since we judged that the context of the country mattered for questions about "how to most effectively display information" and "how do patients prefer to receive this information?" In other words, we thought data about how patients in non-English speaking countries such as Korea and the Netherlands would have limited relevance to the US. However, for KQ 3&4, about what effects public reporting has, we did not judge country context to be as important and therefore included studies from other countries.
	P. 26, 3rd full paragraph, last sentence – should be "difference was no longer significant" rather than "difference was no long significant". P. 26, 4th full paragraph, 1st sentence – should end with "selection of patients for procedures" instead of "patient". P. 29, Key Questions 1 and 2, last sentence – "assess" is misspelled.	Typo's are corrected

Peer Review Comments	Comment	Response
	P. 30, Acceptability of Findings to the VA Population, second sentence – Would suggest adding: "It is not clear whether public reporting would stimulate further changes in a system such as VA with a robust quality assessment and feedback system already in place." General comment – "Veteran" is sometimes capitalized in document and sometimes lowercase.	Text has been added.
	Methods: the discussion about excluding Fung articles, then adding them in, is confusing. Exclusion criteria should only focus on what is fully excluded from the synthesis.	This section has been updated for clarity.
	Of the 11 articles "rejected" because they focused onn individual providers, were there any global insights that impact questions 2 or 3? Are the trends/insights any different from those for hospitals/facilities?	We did not look into detail at the studies on individual providers. The Fung review did include such studies and concluded that in the few studies found results were mixed in the effect on selection of provider and un-clinical outcomes and unintended consequences; no studies were identified as quality improvement activities (11 studies were identified for hospitals).
	Add a section called Recommendations for VA. The section on applicability to the VA population does not get at issues specific to operations/implementation of public reporting of VA data. The Transparency initiative, for example, would benefit from learning more about Aligning Forces for Quality and their experience with providing community-level data to the public.	"Recommendations for VA" is not a heading in our report template. Rather our report provides evidence for a VA policymaker to make recommendations.
	Seems like a tepid conclusion in light of the actual studies. It seems like there is very little evidence that patients and families use reports. Any impact on market share or volume may, in fact, involve decisions made by payers, or the influence of other factors (like loss of accreditation, program closures, etc).	We agree that there is very little evidence that patients and families use reports. We think our existing statement that use is "moderate at best" accurately conveys this.

Peer Review Comments	Comment	Response
Scope	Consumer vs. Patient. There is some variation in use of the term patient and consumer. Both terms are important, yet will have different conceptual views by readers. To address this issue, it might be valuable to have statements early in the document that patient is meant to convey the Veteran and primary target of VA-related public data, and that consumer is a commonly-used term in discussion about public reporting of data. In general, for the purposes of the report, consider them interchangeable (keep in mind that consumer can include family members and informal caregivers, so is a broader and more inclusive term). Then pick one and use that for the rest of the report.	We have added these terms to the report.
	Definition of Provider. There are several uses of the term "provider" – from an individual doctor or clinician, to a hospital or clinic. This will be confusing for readers. Suggest using the term provider to refer to clinicians, and facility/health care provider spelled out for the latter.	Definition of provider - We disagree with this distinction and use of "provider" in the broad sense, which can be an individual provider (authors of those are excluded from the report) and also hospitals and health plans.
	Topic development: is the Office of Quality and Performance a VA or VHA entity?	It is a VHA entity.
	What were the Fung criteria (could state them, rather than relegate to Appendix).	Language has been modified.
	Literature flow: match the numbers in the narrative with the numbers in the diagram (3 or 6 studies from content experts??)	Numbers are now in agreement.
	Discussion about prior reviews is under Literature Flow. Consider a separate heading, such as Prior Review.	We have added this subheading.
	Figure 2: what do circles numbered 1 and 2 represent? The "one mention" and "two mentions" must be the explanation…but it's not immediately clear to the reader.	We have revised the legend for clarity.
	Key Question #1: why non-U.S. studies taken out? The comment about "particularly sensitive to context" makes the assumption that a person in Germany who gets health data is very different than a person in New York.	The cultural context here is around consumerism. The USA is considered a consumerist society, whereas all European countries have not been, although are becoming more so in the past 10 years.
	Key Question #2: maybe I missed this, but I didn't see much discussion about how patients want to receive or access this information…	This section included all the data and recommendations that were in the Hibbard & Sofaer and the RWJ reports, there is nothing more about this topic that we can include.
	Key Question #3, evidence from systematic review by Fung: Paragraph about the two pathways, and "change pathway" is confusing.	We have included a figure to better illustrate this.
	Impact on Clinical Outcomes: prior discussion excluded non-U.S. studies, and this section discusses S. Korean and Dutch studies; this is confusing (see #12 – consider all non-U.S. studies	We have now excluded non-US studies. Still need to check with AT.

Peer Review Comments	Comment	Response
Scope	Interesting items in Limitations discussion – why was the CHQC program abandoned? Why was the CHOP report not part of this review (only because it wasn't in a peer review journal?)	We did not know why the CHOP was abandoned, that was not in the scope. We have added the newer CHOP report as part of our revision to add relevant evidence identified via internet searches.
	Future research. Given the results of the report, there seem to be more research questions than those proposed. Was there any data about how consumers want to receive/access data? This is an important question that could be study variations in how the data is displayed is important, as well as credibility of the data, trust in the "deliverer" of the data (e.g. government). There are studies on numeracy and literacy and how to present data, although not specifically on publically reported data.	We have added this to the search question.
Search Related	Since there was very little in national work might have been interesting to also look at web sites and high quality blogs? Understand peer review is the best but if the data is not there need to look other places	This is a good suggestion and we have now incorporated a web search into the report. We added the Google search.
	Need large scope of review, maybe look more at social media and web info and not just published standard journals	
	If I were responsible for it the main thing I would want checked is the Google search mentioned above to see if the first few dozen "hits" identify any studies that should be added.	
	On page 30 the authors write that "public reporting of quality information" produces over 19,000,000 hits, but when the text is in quotes, it actually results in only 18,100 hits, and when the word "healthcare" is added separately to the search, then the number drops to 17,500, which is still a high number – but the first page of links look highly relevant to the study. Since I don't know which studies were excluded in Figure 1 there's no way to tell that the 22 new studies included in the report are the complete set of useful studies.	We have now incorporated a web search, but limited to the top 30 hits.
Database	The search methodology is described only briefly and incompletely. The methods say that the literature search was done "using standard search terms" and Appendix A, which is cited as the place to look for clarification, lists only a few author searches. This report would benefit from a comprehensive description of the search strategy, so that it can be checked and repeated in the future. If the exact terms were used as in the Fung report, that should be stated, or how the terms here differed.	Appendix A lists all the search terms and databases searched.
	Search strategy does not include search terms (would be helpful to see).	
	No terms for search presented (maybe this will be an appendix?)	The research terms and databases are in the appendix.

Peer Review Comments	Comment	Response
Nursing Home	Given the VA's provision of nursing home care, it would be preferable to have include nursing home care.	
	The objectives, scope, and methods for this review are clearly described, but I am not certain why the review excluded published information about public reporting of nursing homes, physicians or individual providers. Certainly, the VA statement of transparency does not preclude this, and while present plans have implemented publication of facility level quality information there is great interest and expectancies that the transparency will spread. The review would have far more useful if these areas were INCLUDED (since they represent future needs rather than retrospective) E.G. Having put up the website, now we pay a reviewer to identify how we should have done it (better late than never) , failure to include provider and nursing homes means we will always be chasing our tails.	Nursing homes and individual providers were not included in our scope as provided to us by central office.
	Study Selection #2, "nursing homes" – Since VA provides nursing home care, I would have liked to see this included	Nursing homes were not included in the scope provided to us.
	Scope is reasonably presented; however, the rationale for not including nursing homes and individual providers makes no sense. The key questions are not VA-specific, and if there is important information about the display of information or consumer use of information in decision-making, it would be of value here.	Nursing homes and individual providers were not included in our scope as provided to us by central office.
	Do any of the studies mention how hospital staff used the reports? Perhaps "internal transparency" can motivate quality improvement as much as external reporting. (Shame being, perhaps, an even greater motivator than money).	We did not look at "internal transparency" or reporting back to providers but not the public. The one study that directly assessed this question reported more quality improvement activity with public reporting.
Nursing Home	It is worth asking the question "what is missing". For instance, pure public reporting, without any other organized effort to address poor performance or ensure accountability, may simply be "information noise". On the other hand, if poor performers faced loss of accreditation, loss of business, or other penalties, they may be more likely to take action. It would be useful to know if any of the studies combined such managerial interventions with public reporting.	We did not assess the existence of managerial interventions, but think those can be assumed to exist, since without them plan and facility performance would be unlikely to change.
	Study selection: Need better rationale than "VA public reporting for facilities" for not including individual provider and nursing home data studies. The key questions are not VA specific, so it would be ideal to add these in.	The scope was given to us by central office and this is the rational they gave us.

Peer Review Comments	Comment	Response
Recommendations	Not clear how cites 10-12, 14-15 fit into an evidence review with the stated inclusion criteria. These seem more like suggestions/recommendations from a non-systematic review and seem much different than studies such as cites 37-39, for example.	The article and RWJ reports on public reporting are included because they are recommendations from high profile organizations made by experts and based in the evidence that is available. It seemed to us that our report would seem incomplete if it were missing these two key reports.
	Can you be more precise how Hibbard and Sofaer arrived at their conclusions? It will be essential to know if these recommendations are based on "expert consensus" versus empirical evidence, and what exactly the nature of the evidence is (e.g., user acceptability testing – what types of users, how was testing done, etc)	The Hibbard and Sofaer reports are their recommendations, based on the available evidence, which was somewhat thin. Some of their evidence was usability testing. We judged the Hibbard & Sofaer and the RWJ recommendations to represent the best available blend of evidence and opinion.
	Key Question #1 and #2: this section is well written and quite interesting to read. However, it seems like a summary of global recommendations from a few specific papers (Hibbard; Mazor) and less like a synthesis of data. How were these recommendations developed? What type of research was conducted to support the variety of comments and suggestions? While the literature is not large, it might be valuable to provide example of studies behind the suggestions.	
Update Searches	Is it possible to do a quick check, using the same search terms, for any new articles published since January 2011?	We have preformed and update search though August 2011 and incorporated the one new study meeting the inclusion criteria (N=1).
	Some of the newer reporting systems, such as CMS' Hospital Compare and the RWJ AF4Q pilots, are relatively recent. Hence, my earlier comment that we conduct a "quick peak" at literature published since Jan 2011.	We have done an update search, however, the one study meeting inclusion criteria was not about Hospital Compare or RWJ AF4Q.
Evidence & Summary	Quality of Evidence: the title suggests quality of the studies ("quality of evidence"); narrative is about impact on quality improvement efforts.	The narrative is meant to explain why the quality of evidence for these studies is generally low, namely why one study did a direct comparison.
	It would be good to restate or list what the Key Questions are before addressing them.	We have now restated the key questions in the executive summary.
	Summary: Consider repeating the Questions above each summary.	We have added this.