9-6-2018

Impact of public release of performance data on the behaviour of healthcare consumers and providers.

David Metcalfe
University of Oxford

Arturo J. Rios Diaz
Thomas Jefferson University

Olubode A. Olufajo
Howard University College of Medicine

M. Sofia Massa
University of Oxford

Nicole A.B.M. Ketelaar
Saxion University of Applied Sciences

Recommended Citation

Authors

David Metcalfe, Arturo J. Rios Diaz, Olubode A. Olufajo, M. Sofia Massa, Nicole A.B.M. Ketelaar, Signe A. Flottorp, and Daniel C Perry

Cochrane Database of Systematic Reviews

Impact of public release of performance data on the behaviour of healthcare consumers and providers (Review)

Metcalfe D, Rios Diaz AJ, Olufajo OA, Massa MS, Ketelaar NABM, Flottorp SA, Perry DC

www.cochranelibrary.com

WILEY

T A B L E O F C O N T E N T S

Impact of public release of performance data on the behaviour of healthcare consumers and providers

David Metcalfe[1], Arturo J Rios Diaz[2], Olubode A Olufajo[3], M. Sofia Massa[4], Nicole ABM Ketelaar[5], Signe A. Flottorp[6,7], Daniel C Perry[1]

[1]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, Oxford, UK. [2]Department of Surgery, Thomas Jefferson University Hospital, Philadelphia, PA, USA. [3]Department of Surgery, Howard-Harvard Health Sciences Outcomes Research Center Howard University College of Medicine, Washington, DC, USA. [4]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [5]Social Work Research Group, Saxion University of Applied Sciences, Enschede, Netherlands. [6]Norwegian Institute of Public Health, Oslo, Norway. [7]Institute of Health and Society, University of Oslo, Oslo, Norway

Contact address: David Metcalfe, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, John Radcliffe Hospital, Headley Way, Oxford, OX3 9DU, UK. d.metcalfe@doctors.org.uk.

A B S T R A C T

Background

It is becoming increasingly common to publish information about the quality and performance of healthcare organisations and individual professionals. However, we do not know how this information is used, or the extent to which such reporting leads to quality improvement by changing the behaviour of healthcare consumers, providers, and purchasers.

Objectives

To estimate the effects of public release of performance data, from any source, on changing the healthcare utilisation behaviour of healthcare consumers, providers (professionals and organisations), and purchasers of care. In addition, we sought to estimate the effects on healthcare provider performance, patient outcomes, and staff morale.

Search methods

We searched CENTRAL, MEDLINE, Embase, and two trials registers on 26 June 2017. We checked reference lists of all included studies to identify additional studies.

Selection criteria

We searched for randomised or non-randomised trials, interrupted time series, and controlled before-after studies of the effects of publicly releasing data regarding any aspect of the performance of healthcare organisations or professionals. Each study had to report at least one main outcome related to selecting or changing care.

Data collection and analysis

Two review authors independently screened studies for eligibility and extracted data. For each study, we extracted data about the target groups (healthcare consumers, healthcare providers, and healthcare purchasers), performance data, main outcomes (choice of healthcare provider, and improvement by means of changes in care), and other outcomes (awareness, attitude, knowledge of performance data, and costs). Given the substantial degree of clinical and methodological heterogeneity between the studies, we presented the findings for each policy in a structured format, but did not undertake a meta-analysis.

Main results

We included 12 studies that analysed data from more than 7570 providers (e.g. professionals and organisations), and a further 3,333,386 clinical encounters (e.g. patient referrals, prescriptions). We included four cluster-randomised trials, one cluster-non-randomised trial, six interrupted time series studies, and one controlled before-after study. Eight studies were undertaken in the USA, and one each in Canada, Korea, China, and The Netherlands. Four studies examined the effect of public release of performance data on consumer healthcare choices, and four on improving quality.

There was low-certainty evidence that public release of performance data may make little or no difference to long-term healthcare utilisation by healthcare consumers (3 studies; 18,294 insurance plan beneficiaries), or providers (4 studies; 3,000,000 births, and 67 healthcare providers), or to provider performance (1 study; 82 providers). However, there was also low-certainty evidence to suggest that public release of performance data may slightly improve some patient outcomes (5 studies, 315,092 hospitalisations, and 7502 providers). There was low-certainty evidence from a single study to suggest that public release of performance data may have differential effects on disadvantaged populations. There was no evidence about effects on healthcare utilisation decisions by purchasers, or adverse effects.

Authors' conclusions

The existing evidence base is inadequate to directly inform policy and practice. Further studies should consider whether public release of performance data can improve patient outcomes, as well as healthcare processes.

PLAIN LANGUAGE SUMMARY

Can the public release of performance data in health care influence the behaviour of consumers, healthcare providers, and organisations?

What is the aim of this review?

The aim was to find out if publicly releasing information about the performance of healthcare providers (e.g. hospitals and individual professionals) has a measurable influence on changing the behaviour of consumers, providers, and purchasers of care. We also sought to determine whether this affected the performance of healthcare providers, patient outcomes, and staff morale.

Key messages

Public release of performance data may lead to little or no difference in healthcare choices (made by either consumers or providers), or provider performance. However, it may slightly improve outcomes for patients.

What was studied in the review?

Healthcare providers are increasingly expected to inform the public on how well they are performing. However, it is not yet known whether public release of performance data has a measurable influence on patients' choice of healthcare services, or whether it can truly drive improvements in the quality of health care.

What are the main results of the review?

The authors searched the literature for studies evaluating the effects of publicly releasing healthcare performance information, and found 12 relevant studies that analysed data from more than 7570 providers, and a further 3,333,386 clinical encounters, e.g. individual patients.

There was low-certainty evidence that public release of performance data may lead to little or no difference in the services that patients choose to access, the decisions taken by healthcare providers, or overall provider performance. There was low-certainty evidence suggesting that some patient outcomes may slightly improve following public release of performance data, but that this might have less of an effect on the behaviour of disadvantaged populations. There was no evidence relating to healthcare utilisation decisions by purchasers, or adverse effects.

Although a number of the studies were individually well conducted, there were limitations: in particular, the evidence base varied substantially in terms of setting (e.g. United States or Korea), health condition (e.g. heart attack or hip replacement), type of performance data (e.g. process or patient outcome), and the mode of data publication (e.g. mail shot or poster). Their findings were also inconsistent, with some reporting changes attributed to public release of information, and others reporting no such changes.

How up-to-date is this review?

The review authors searched for studies that had been published up to June 2017.

People: Insurance plan beneficiaries, birthing mothers, GPs
Settings (countries and clinical settings): United States, Canada, South Korea, Netherlands, China / Community, primary care and hospitals
Intervention: Public release of performance data
Comparison: No public reporting

Outcomes	Impact	No of clinical encounters (studies)	Certainty of the evidence (GRADE)*
Changes in healthcare utilisation by consumers	Public release of performance data may make little or no difference to long-term healthcare utilisation by consumers. However, two studies (one cNRT and one ITS) found that some population subgroups might be influenced by public release of performance data	18,294 insurance plan beneficiaries[a] (3: 1 cRT, 1 cNRT, 1 ITS)	⊕⊕◯◯ **low**
Changes in healthcare decisions taken by healthcare providers (professionals and organisations)	Public release of performance data may make little or no difference to decisions taken by healthcare professionals. Two studies (2 cRTs) found that some decisions might be affected by public release of performance data. One study (ITS) found that decisions might be influenced by the initial release of data, but that subsequent releases might have less impact	3,000,000 births[b] and 67 healthcare providers (4: 2 RTs, 2 ITS)	⊕⊕◯◯ **low**[c]
Changes in the healthcare utilisation decisions of purchasers	No studies reported this outcome.	-	-
Changes in provider performance	Public release of performance data may make little or no difference to objective measures of provider performance	82 healthcare providers (1 cRT)	⊕⊕◯◯ **low**[d]
Changes in patient outcome	Public release of performance data may slightly improve patient outcomes	315,092 hospitalisations and 7503 healthcare providers (5: 1 RT, 3 ITS, 1 CBA)	⊕⊕◯◯ **low**[e]

| Adverse effects | No studies reported this outcome. | - | - |
| Impact on equity | Public release of performance data may have a greater effect on provider choice among advantaged populations | Unknown (1 ITS) | ⊕⊕◯◯ **low** |

EPOC adapted statements for GRADE Working Group grades of evidence
High-certainty. This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different[†] is low.
Moderate-certainty. This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different[†] is moderate.
Low-certainty. This research provides some indication of the likely effect. However, the likelihood that it will be substantially different[†] is high.
Very low-certainty. This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different[†] is very high.
[†] Substantially different = a large enough difference that it might affect a decision

[a] Number was based only on Farley 2002a and Farley 2002b studies, as the total number of cases analysed in Romano 2004 was unclear
[b] Number of participants in Jang 2011 (3,000,000) estimated from data presented in Chung 2014
[c] Downgraded one level for inconsistency as effect shown by Zhang 2016, but not IkkersheJang 2011, Ikkersheim 2013, or Flett 201511
[d] Downgraded two levels for risk of bias, as there was attrition of participating hospitals, evidence of contamination of the intervention across intervention and control hospitals, and blinding was not possible given the nature of the intervention
[e] Downgraded two levels for inconsistency, as there was marked disagreement between studies, with two showing improvements in patient outcome (Liu Tu 2009; Liu 20179), and three showing no such improvements (DeVoRinke 2015; DeVore 2016; Joynt 201615)

cluster-randomised trial (cRT); cluster-non-randomised trial (cNRT); controlled before-after (CBA) study; interrupted time
   series (ITS) study; randomised trial (RT)

BACKGROUND

It is becoming increasingly common to release information about the performance of healthcare systems into the public domain. In the present era of accountability, cost-effectiveness, quality improvement, and demand-driven healthcare systems, decision makers such as governments, regulators, purchaser and provider organisations, health professionals, and consumers of health care are becoming more interested in measuring performance (Smith 2009). Such measurements may be presented in consumer reports, provider profiles, or report cards. It is not always clear who the information users are or what the release of data is expected to achieve. However, it is often assumed that the information will influence the behaviours of various stakeholders, and so ultimately lead to health system improvements (Berwick 2003; Smith 2009; Campanella 2016).

One study has conceptualised public reporting of performance data as (1) supporting patient choice, (2) improving accountability, and (3) allowing providers to benchmark their performance against others (Greenhalgh 2018).

Publication of performance data can support patient choice by helping them to identify the highest performing providers. However, there are many barriers to patient use of performance data (Canaway 2017). These include the complexity of the performance data (Hibbard 2010), lack of skills to comprehend and use performance data (Hibbard 2007; Canaway 2017; Canaway 2018), and the way data are presented (Damman 2010; Canaway 2017; Canaway 2018). Such barriers might negate the impact of choice, and even reduce equity in health care. Consumers from poorer backgrounds and with lower educational levels may be less able to choose, and less able to afford travel to better performing, but more distant, providers (Aggarwal 2017; Moscelli 2017). There is also evidence that patients often do not use published performance data when making healthcare choices (Greenhalgh 2018).

Improved accountability may be achieved by encouraging providers to focus on quality issues, as they know that performance measures will be published (Fung 2008; Hendriks 2009). This in turn, may stimulate quality improvements, particularly as providers can see their own performance against that of other clinicians and hospitals. Similarly, patients who preferentially choose high-quality health care might help drive improvements, by concentrating resources with the best performing providers (Hibbard 2009; Kolstad 2009; Werner 2009).

Other proposed goals for performance measurements have been linked to controlling costs (Berwick 1990; Sirio 1996), regulating the overall healthcare system (Rosenthal 1998; Schut 2005), and influencing the decisions of healthcare purchasers (Brook 1994; Hibbard 1997; Mukamel 1998).

Professional concerns to public release of performance data often relate to the validity of both the performance measures themselves, and comparisons between health providers (Sherman 2013; Kiernan 2015; Burns 2016;). There are concerns that failure to adequately adjust for case mix differences might lead to providers that treat higher-risk patients being labelled as poor performers, or to providers preferentially selecting lower-risk patients (Wasfy 2015; Burns 2016; Shahian 2017; Wadhera 2017). In healthcare systems where providers charge for their services, the 'better' performing providers might feel empowered to increase charges, thereby restricting access to better care (Mukamel 1998). An additional risk is that publication of performance data may lead to improved reporting, without necessarily improving performance. It has been argued that the care processes that are easiest to measure are often those that are least important in a quality improvement context, and can result in the de-prioritisation of other tasks (Loeb 2004).

Description of the intervention

Public release of performance data is the release of information about the quality of care, so that patients and consumers can better decide what health care they wish to select, and healthcare professionals and organisations can better decide what to provide, improve, or purchase. This mechanism excludes the use of auditing and feedback as a tool for improving professional practice and healthcare outcomes, which has been reviewed elsewhere (Ivers 2012).

How the intervention might work

Public release of performance data may change individual or organisational behaviour through a number of mechanisms. The goal of improving quality of health care can be achieved through a selection pathway or a change pathway (Berwick 2003). Consumers, patients, and purchaser organisations that are in a position to do so, can select the best healthcare professionals and organisations. This type of selection will not change the quality of the delivered care by itself, but it can be a stimulus for quality improvement. Importantly, such changes might be attenuated by the limited choice that patients have in many cases, e.g. in the case of emergencies, the need to access specialised care that is only available in few centres, or because of resource limitations (Aggarwal 2017; Moscelli 2017). In a change pathway, healthcare professionals and organisations can improve performance by changing their work procedures or professional culture, and organisations can make structural changes.

Why it is important to do this review

Some systematic reviews have suggested positive effects of publicly releasing performance data, but included a broad range of study designs (Marshall 2000; Shekelle 2008; Fung 2008; Faber 2009). This study (which is the first update of Ketelaar 2011) aimed to

review the evidence for the impact of such interventions using more stringent selection criteria.

OBJECTIVES

To estimate the effects of publicly releasing performance data on changing the healthcare utilisation behaviour of healthcare consumers, providers (professionals and organisations), and purchasers of care. In addition, we sought to estimate the effects on healthcare provider performance, patient outcomes, and staff morale.

METHODS

Criteria for considering studies for this review

Types of studies

- Randomised trials, including cluster-randomised trials
- Non-randomised trials, including cluster-non-randomised trials, which use non-random methods of allocation, such as alternation or allocation by case note number
- Controlled before-after studies, with at least two intervention sites and two control sites that are chosen for similarity of main outcome measures at baseline
- Interrupted time series studies, with at least three data points before and three data points after the intervention

We included non-randomised studies in anticipation of a lack of randomised trials, but also because some interventions might not be appropriate for a trial (e.g. randomising participants to not receive important information that might affect their healthcare choices), and others might have a variable effect over time that is best observed by an alternative study design, such as an interrupted time series.

Types of participants

Patients or other healthcare consumers and healthcare providers, including organisations (e.g. hospitals), without any restriction by type of healthcare professional, provider, setting, or purchaser.

Types of interventions

We included interventions that contained the following elements:
- Performance data about any aspect of the healthcare organisations or individuals, including process measures (e.g. waiting times), healthcare outcomes (e.g. mortality), structure measures (e.g. presence of waiting rooms), consumer or patient

experiences (e.g. Consumer Assessment of Healthcare Providers and System (CAHPS) data), with or without expert or peer-assessed measures, e.g. certification, accreditation, and quality ratings given by colleagues. Performance data were included if prepared and released by any organisation, such as the government, insurers, consumer organisations, or providers. We excluded studies that did not evaluate publication of performance data concerning process measures, healthcare outcomes, structure measure, consumer or patient experiences, or expert or peer-assessed measures.
- The release of performance data into the public domain in written or electronic form without regard to any minimum degree of accessibility. For example, this could include a report available in a publicly accessible library, as well as active dissemination directly to consumers through personal mailings.

Comparators

The following comparisons were planned:
1. Public release of performance data compared to settings in which data were not released to the public
2. Different modes of releasing performance data to the public

Types of outcome measures

Primary outcomes

We planned the primary outcome measures according to two key aims of publicly releasing performance data.

1. Improvement by selection
- Changes in healthcare utilisation by consumers
  - Objective measures of changing consumer behaviour, such as increased use of a specific healthcare provider
- Changes in healthcare decisions taken by healthcare providers (professionals and organisations)
  - Objective measures of changing healthcare provider behaviour, such as changes to drug prescribing
- Changes in the healthcare utilisation decisions of purchasers
  - Objective measures of changing purchaser behaviour, such as increased or decreased funding for services

2. Improvement by changes in care
- Changes in provider performance
  - Objective changes, such as reaching the correct diagnosis or time to treatment
  - Including measures that were made both public and others that were not
- Changes in patient outcome
  - Objective changes, such as mortality or patient-reported outcome measures
- Changes in staff morale
  - Using a previously validated assessment tool

Secondary outcomes

We considered unintended and adverse effects or harms, and any potential impact on equity (e.g. differential effects between advantaged and disadvantaged populations), and awareness, knowledge, attitude, or costs.

We excluded studies that reported awareness, attitude, perspectives, and knowledge of performance data and cost data in the absence of objective measures of decision behaviour, provider performance or patient outcomes.

Search methods for identification of studies

Electronic searches

We searched the Database of Abstracts of Reviews of Effects (DARE) for primary studies included in related systematic reviews. We searched the following databases on 26 June 2017:

- Cochrane Central Register of Controlled Trials (CENTRAL; 2017, Issue 5) in the Cochrane Library;
- MEDLINE Ovid (including Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Versions);
- Embase Ovid.

The Cochrane Effective Practice and Organisation of Care (EPOC) Information Specialist developed the search strategies in consultation with the authors. Search strategies are comprised of keywords and controlled vocabulary terms. We applied no language or time limits. We searched all databases from database start date to 26 June 2017.

Searching other resources

Trial Registries

- International Clinical Trials Registry Platform (ICTRP), Word Health Organization (WHO) www.who.int/ictrp/en/ (searched 26 June 2017)
- ClinicalTrials.gov, US National Institutes of Health (NIH) clinicaltrials.gov/ (searched 26 June 2017)

We manually searched the reference lists of all included studies. We provided all search strategies used in Appendix 1.

Data collection and analysis

Selection of studies

We downloaded all titles and abstracts retrieved in the electronic search to a reference management database. We removed the duplicates, and two review authors then independently examined the remaining references. All review authors recorded their assessments of abstracts with points: '0' for exclusion, '1' for doubtful and '2' for inclusion. Two review authors (DM, ARD) independently rated each abstract; therefore, a minimum score of zero, and a maximum score of four was possible. Abstracts with a combined score of zero or one were excluded. Studies with a combined score of three or four were included. Two review authors resolved the fate of studies with a combined score of two by discussion. A third review author (OO) adjudicated on any disagreements that remained unresolved. Figure 1 shows the PRISMA flow diagram that accounts for exclusion of all items received by the search strategy.

Figure 1. Flowchart for study selection
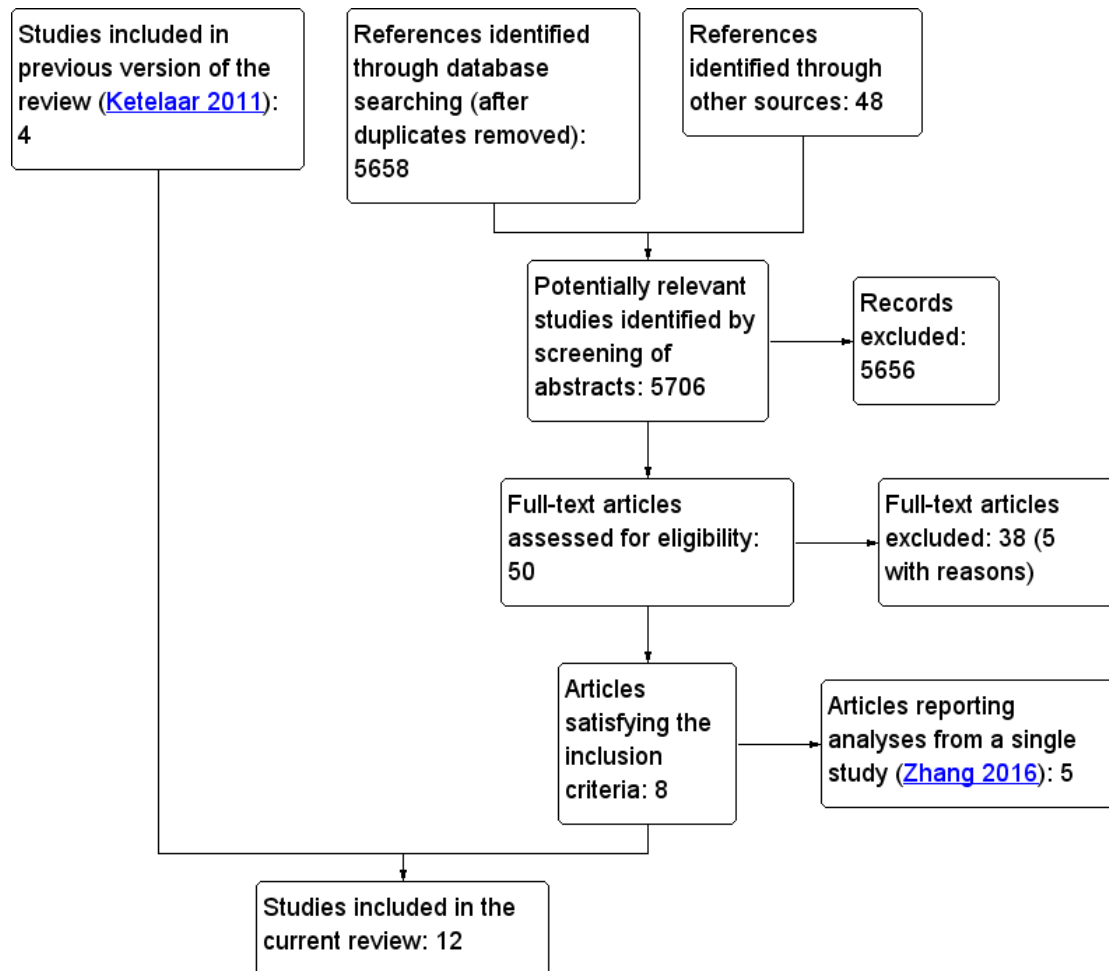
Data extraction and management

Two authors (DM, OO) independently extracted the data about the study design, patient and provider characteristics, interventions, outcome measures, and healthcare choices to a form specially designed for our review. Disagreements were resolved by discussion, and we accepted the judgement of a third author (ARD) in the event of continued disagreement.

Assessment of risk of bias in included studies

We assessed risk of bias by applying the guidance from the *Cochrane Handbook for Systematic Reviews of Interventions*, which recommends using the following items: (i) adequate sequence generation, (ii) concealment of allocation, (iii) blinding, (iv) incomplete outcome data, (v) selective reporting, and (vi) no risk of bias from

other sources (Higgins 2011). However, we deviated from this guidance: we used three additional criteria that are specified by the Cochrane Effective Practice and Organisation of Care Group: (vii) baseline characteristic similarity, (viii) baseline outcome similarity, and (ix) adequate protection against contamination (EPOC 2013). We used these nine standard criteria for randomised trials, non-randomised trials, and controlled before-after studies. We used seven criteria for interrupted time series studies, and applied these as recommended by EPOC 2013: (i) the intervention is independent of other changes, (ii) the shape of the intervention effect is pre-specified, (iii) the intervention is unlikely to affect data collection, (iv) knowledge of the allocated interventions is adequately prevented during the study, (v) the outcome data are incomplete, (vi) reporting is not selective, and (vii) there is no risk of bias from other sources. Two review authors (DM, ARD) inde-

pendently reached judgements about risk of bias using the guidance provided by Higgins 2011 and EPOC 2013, and resolved disagreements by discussion. A third review author (OO) dealt with any disagreements that the two review authors could not resolve.

Measures of treatment effect

In order to standardise reporting of effect sizes, we re-analysed data from individual studies to ensure that randomised trials and controlled before-after studies could be reported as relative effects. Interrupted time series were reported as change in level and change in slope. We described the methods used for re-analysing and presenting these data in Data synthesis.

Unit of analysis issues

We noted whether randomised trials randomised patients or healthcare providers. If analysis did not allow for clustering of patients within healthcare providers, we recorded a unit of analysis error, because such analyses tend to overestimate the precision of the treatment effect. In the event of a unit of analysis error and insufficient data to account for clustering, we did not report P values or confidence intervals.

Dealing with missing data

In the event of important missing data, contacted the authors of individual studies. As described in Data synthesis, we electronically extracted missing interrupted time series data that were presented in graphs.

Assessment of heterogeneity

There were substantial differences between the policies and interventions described. There were also differences between the settings, in terms of culture and health system delivery. Although some studies evaluated similar interventions, there were still important clinical and methodological differences. As statistical tests for heterogeneity lack power when few studies are included, we elected not to calculate average effects across studies, or to estimate statistical heterogeneity (Schroll 2011).

Assessment of reporting biases

We did not present funnel plots as we did not undertake a meta-analysis and there were not more than 10 studies contributing to any individual analysis (Higgins 2011).

Data synthesis

We followed the EPOC recommendations with regard to analysing data from individual studies and meta-analysis (EPOC 2013). We expressed the findings from controlled before-after studies as relative effects. To achieve this, we reported continuous variables as relative change in outcome measures, adjusted for baseline differences. We undertook absolute difference-in-difference analyses that were adjusted for differences in the postintervention control group using: ((postintervention intervention group - postintervention control group) - (preintervention intervention - pre-intervention control))/postintervention control. For ease of comparison with the findings of controlled before-after studies, we reported the findings of randomised and non-randomised trials using the same difference-in-difference analysis.

Interrupted time series are typically reported using regression analysis, such as autoregressive integrated moving average (ARIMA) analysis. Pursuant to the EPOC recommendations, we present outcomes along two dimensions: change in level and change in slope (EPOC 2013). The former represents the immediate effect of the intervention as measured by the difference between the fitted value for the first post-intervention time point and the predicted outcome at the same point, based only on an extrapolation of the pre-intervention slope. Change in slope is an expression of any longer-term effect of the intervention. We decided to use a similar method to the change in level, but a later follow-up period, e.g. six months.

In the event that appropriate interrupted time series analyses were not reported but that data were presented graphically, we read values from graphs using Plot Digitizer v2.6.8 (Huwaldt 2004). We extracted 'actual' data points from all studies and only planned to use lines of best fit in the event that true points were not available. A segmented time series model ($Y(t) = B0 + B1*$preslope $+ B2*$postslope $+ B3*$ intervention $+ e(t)$) was specified, in which $Y(t)$ was the outcome in month t. Preslope is a continuous variable that indicates time from the beginning of the study until the end of the pre-intervention phase, after which it was coded as a constant. Postslope is assigned the value 0 until after the intervention takes place, after which it is coded sequentially from 1 (i.e. 1, 2, 3). Intervention is assigned the value 0 pre-intervention and 1 in the postintervention time period. In this model, B1 estimates the pre-intervention slope, B2 the postintervention slope, and B3 the change in level, i.e. the difference between the first postintervention time point and the extrapolated first postintervention time point had the pre-intervention line continued into the postintervention period. The difference in slope was determined using B2 - B1.

We reported effects at 3, 6, 9, 12, and 24 months postintervention when the data were available. Given the substantial degree of clinical and methodological heterogeneity between the studies, we presented the findings for each policy in a structured format, but did not undertake a meta-analysis.

Summary of findings

We summarised the findings of the main intervention comparisons in a 'Summary of findings' table to illustrate the certainty of the evidence. One review author (DM) categorised the certainty of the evidence as high, moderate, low, or very low, using the five GRADE domains (study limitations, consistency of effect, imprecision, indirectness, and publication bias (Guyatt 2011)). We undertook this pursuant to Chapter 12 of the *Cochrane Handbook for Systematic Reviews of Interventions* and worksheets created by EPOC (Higgins 2011; EPOC 2013). All other co-authors checked these judgments, and resolved disagreements through discussion. When ratings were up- or down-graded, we justified these decisions using footnotes in Appendix 2 and Summary of findings for the main comparison. Standardised statements for reporting effects and certainty of evidence were selected, based on the GRADE assessments for each outcome, and used throughout the review (EPOC 2017). The seven outcomes reported in Summary of findings for the main comparison are:

- Changes in healthcare utilisation by consumers
- Changes in healthcare decisions taken by healthcare providers (professionals and organisations)
- Changes in the healthcare utilisation decisions by healthcare purchasers
- Changes in provider performance
- Changes in patient outcome
- Adverse effects
- Impact on equity

Subgroup analysis and investigation of heterogeneity

As described in Data synthesis, we presented the findings of individual studies in a structured format rather than attempting meta-analysis, given the substantial heterogeneity between the studies. Therefore, it was not possible to undertake subgroup analyses.

Sensitivity analysis

In the absence of a formal meta-analysis, we did not undertake any sensitivity analyses.

R E S U L T S

Description of studies

The included studies are summarised in Table 1 and described fully in Characteristics of included studies. A number of studies that narrowly failed to satisfy our selection criteria are described in Characteristics of excluded studies.

Results of the search

The electronic searches for this update retrieved 5658 individual items; a further 48 were identified from other sources, e.g. manual searching of reference lists. We excluded 5656 items because the titles and abstracts did not meet our inclusion criteria. We retrieved the full-text versions of the remaining 50 articles; 38 of these did not satisfy the inclusion criteria; five with reasons, see (Characteristics of excluded studies). Five of the remaining 16 articles reported separate analyses of a single cluster randomised trial, and so we treated them as a single study for the purposes of this review (Zhang 2016). Therefore, we included 12 studies in the review. As described in Data synthesis, we did not undertake formal meta-analyses due to substantial inter-study heterogeneity. We presented the study flow chart in Figure 1 (Moher 1999).

Included studies

We included 12 studies that comprised more than 7570 providers (e.g. professionals and organisations) and a further 3,333,386 clinical encounters (e.g. patient referrals, prescriptions). There were four cluster randomised trials (Farley 2002a; Tu 2009; Ikkersheim 2013; Zhang 2016), one cluster-non-randomised trial (Farley 2002b), six interrupted time series studies (Romano 2004; Jang 2011; Flett 2015; DeVore 2016; Joynt 2016; Liu 2017), and one controlled before-after study (Rinke 2015). Eight were conducted in the USA (Farley 2002a; Farley 2002b; Romano 2004; Flett 2015; Rinke 2015; DeVore 2016; Joynt 2016; Liu 2017), and one each in Canada (Tu 2009), the Netherlands (Ikkersheim 2013), Korea (Jang 2011), and China (Zhang 2016).

Three studies focused on changes in the healthcare utilisation decisions of consumers (Farley 2002a; Farley 2002b; Romano 2004), four of providers (Jang 2011; Ikkersheim 2013; Flett 2015; Zhang 2016), and none of purchasers. Two studies reported data on changes to provider performance (Tu 2009; Rinke 2015), five on patient outcomes (Tu 2009; Flett 2015; DeVore 2016; Joynt 2016; Liu 2017), and none on staff morale. No study explicitly reported adverse events as a separate outcome, or gave particular consideration to effects on equitable health care.

Three US studies examined the effect of a single suite of interventions (i.e. laws mandating public reporting of healthcare-associated infections in the United States), which were introduced by some state legislatures between 2006 and 2009 (Flett 2015; Rinke 2015; Liu 2017). Liu 2017 examined the effect of mandatory reporting on central line-associated bloodstream infection rates in adult intensive care units. They undertook an interrupted time series study using data from hospitals contributing to the National Healthcare Safety Network between 2006 and 2012. States that did not introduce mandatory reporting were used to control for secular trends through a difference-in-difference analysis. The other two studies focused their analyses on healthcare-associated infections in paediatric inpatients (Flett 2015; Rinke 2015). Rinke 2015 sought to determine whether mandatory central line-asso-

ciated bloodstream infection public reporting was associated with a reduction in a specific paediatric safety indicator (PDI12, i.e. selected infections due to medical care), which is defined using diagnosis codes on hospital discharge. They undertook a controlled before-after study using the Kids' Inpatient Database, which is one of a suite of administrative healthcare databases coordinated by the Healthcare Cost and Utilization Project at the US Agency for Healthcare Research and Quality. Flett 2015 did not examine patient outcomes, but aimed to test the hypothesis that clinicians in hospitals that are required to report central line-associated bloodstream infections would modify their behaviour by sending fewer blood culture tests or prescribing longer courses of antibiotics. They undertook an interrupted time series using data from the Pediatric Health Information System, which is a collaborative venture between children's hospitals that is used for clinical audit and quality improvement. The data were analysed using generalised linear mixed-effects models with auto-correlated residuals to compare central line-associated bloodstream infections adjusted rate ratios before and after implementation of mandatory reporting laws.

Two US studies studied the effect of providing information about plan performance on choice of insurance plan by new Medicaid beneficiaries (Farley 2002a; Farley 2002b). Farley 2002a was a cluster-randomised trial, using data from new Medicaid beneficiaries in Iowa. Under Iowa Medicaid, new enrollees were automatically assigned, by default, to one of four private health maintenance organisations or the Medicaid primary care case management programme. They were sent a packet of information about their specific health plan and benefits under Medicaid. The control group received the standard packet of information and the intervention group received this, plus an additional report that described the performance of each health plan, along domains such as 'overall health care rating', and 'personal doctor rating'. The authors used multinomial logistic regression to model the odds of new beneficiaries electing to continue with or change their allocated plan. In Farley 2002b, the same author team undertook a cluster-non-randomised trial to evaluate the same performance reports on beneficiary choice within the New Jersey Medicaid programme. The study design was very similar to Farley 2002a, in terms of control and intervention groups, although this was technically an non-randomised trial, because participants were allocated according to the last digit of their Medicaid case ID number. The objective outcome measure reported was the effect of performance reports on Medicaid beneficiary plan choices.

The other three US studies each examined the impact of different public reporting initiatives on patient outcomes (Romano 2004; DeVore 2016; Joynt 2016). Two used Medicare claims data, and so confined their analyses to the Medicare population, i.e. those aged 65 years or older (DeVore 2016; Joynt 2016). DeVore 2016 undertook an interrupted time series to study the effect on 30-day re-admissions, of publicly reporting risk-adjusted hospital readmission rates for patients with selected conditions (acute my-

ocardial infarction, heart failure, and pneumonia) on the Hospital Compare website. Joynt 2016 reported an interrupted time series with a similar study design to DeVore 2016, but examined the impact on mortality rates, of public reporting of mortality (for patients with the same three selected conditions) on the Hospital Compare website. They used hierarchical modelling to compare 30-day mortality in the pre- and postreporting periods. The final US study presented an interrupted time series based on the California Hospital Outcomes Project in California and the Cardiac Surgery Reporting System in New York (Romano 2004). This study evaluated the effects of publishing report cards on trends in hospital volumes for specific diagnoses, i.e. coronary artery bypass surgery mortality in New York, and both acute myocardial infarction and postdiscectomy complications in California. The interrupted time series examined hospital case volumes, determined using administrative data sets in each state (the California Patient Discharge Data Set and the New York Statewide Planning and Research Cooperative System) before and after the publication of reports that identified hospitals as performance outliers. These reports were published by the California Hospital Outcomes Project and the New York Cardiac Surgery Reporting System.

There were three cluster-randomised trials outside the US; one each in Canada (Tu 2009), the Netherlands (Ikkersheim 2013), and China (Zhang 2016). In Canada, Tu 2009 evaluated the public release of performance data about 12 care quality indicators for acute myocardial infarction and six for congestive heart failure in 86 hospitals. Participating hospitals were randomised to either early (January 2004) or delayed (September 2005) publication of performance report cards. The performance data were provided to individual hospitals, and then publicised both online and through popular media, with coverage achieved through television, radio, and newspapers. The outcomes reported by this study were any change in hospital performance, measured using the 18 care quality indicators. The cluster-randomised trial in the Netherlands randomised 26 GPs to receive either individualised hospital report cards (65.4%), or to a control group (34.6%) that did not receive this information (Ikkersheim 2013). The study then captured individual patient referrals (for breast cancer, cataract surgery, and hip or knee replacement) to one of four hospitals in the region, using an electronic referral system. Zhang 2016 undertook a cluster-randomised trial in Hubei Province, south central China. They matched 20 primary care providers within a single city, based on similar organisational characteristics. In this matched-pair cluster-randomised trial, half the providers were randomised to public reporting of injection prescribing, by way of league tables that were posted on outpatient bulletin boards. Performance data were also disseminated to both local health authorities and the leaders of hospitals in the intervention group. The outcomes were the percentage of prescriptions requiring antibiotics, percentage requiring intravenous antibiotics, and the average expenditure per prescription.

Finally, a single interrupted time series study was undertaken in

Seoul, South Korea by Jang 2011. In this study, the intervention was public release of data (online and in media releases) about caesarean section rates for 1194 institutions across the country. These rates were publicised as part of a series of public releases, which were not described in detail. The outcome was change in risk-adjusted institutional caesarean section rates over the whole study period, and after each public release of data.

Excluded studies

In total, we excluded 38 studies after assessing full copies of the papers. The main reasons for exclusion were: ineligible study design (24), interventions did not contain process measures, health care outcomes, structure measures, consumer or patient experiences, expert- or peer-assessed measures (8), no objective outcome data were recorded or available for one or both arms (3), the study

was about hypothetical choices (3). We listed selected studies that readers might reasonably have expected to find included in this review in the 'Characteristics of excluded studies' table.

Risk of bias in included studies

The included studies were rated on different risk of bias items as appropriate for each study design (randomised trial, non-randomised trial, controlled before-after, or interrupted time series). We described this in Assessment of risk of bias in included studies, but in summary, we rated randomised trials, non-randomised trials, and controlled before-after studies using the same nine criteria, and used seven criteria for interrupted time series studies. We showed the results of these risk of bias assessments in the 'Characteristics of included studies' tables and summarised them in both Figure 2 and Figure 3.

Figure 2. Risk of bias graph: review authors' judgements about each risk of bias item, presented as percentages across all included studies. The blank spaces represent risk of bias criteria that were not applicable to the study design.
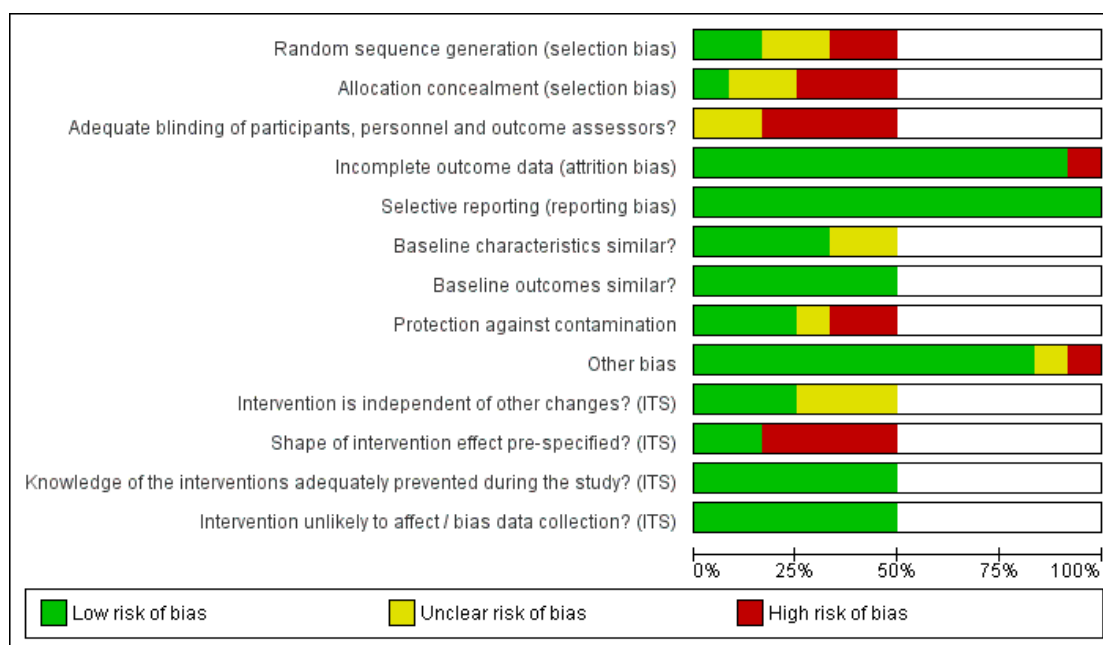
Figure 3.   Risk of bias summary: review authors' judgements about each risk of bias item for each included study. The blank cells represent risk of bias criteria that were not applicable to the study design.

Allocation

The extent of possible selection bias due to the random sequence generation process was unclear in two studies, because the precise method of random sequence generation was not described (Farley 2002a; Ikkersheim 2013). Two studies were at high risk, as Rinke 2015 was a controlled before-after study, and Farley 2002b was a cluster-non-randomised trial, and so used a non-random method of sequence generation. We judged risk of selection bias as low for Zhang 2016 who 'flipped a coin to randomly assign' paired primary care institutions, and Tu 2009 who employed a dedicated study statistician to implement a stratified randomisation process. We made the same judgements for allocation concealment as for random sequence generation, except for Zhang 2016, which was judged to be at high risk for allocation concealment given their use of a coin flip.

Blinding

Although hospitals and healthcare providers could not be blinded to their allocated groups, individual participants were unlikely to have been aware that a study was taking place. No study explicitly contacted individual patients or members of the public to inform them about the research question, intervention, or measured outcomes. For this reason, two studies were considered to be at unclear risk, as it was not stated whether individuals in those trials were informed that a study was taking place (Farley 2002a; Farley 2002b). Four studies were at high risk, because providers were likely to know that a study was taking place, and it was not possible to blind them to their group allocation (Tu 2009; Ikkersheim 2013; Rinke 2015; Zhang 2016).

Incomplete outcome data

We judged 11/12 included studies to be at low risk of attrition bias, because these studies based their outcomes on routinely collected administrative data, e.g. electronic prescriptions or hospital referrals. Only Tu 2009 was judged to be at high risk of bias, because five randomised hospitals withdrew due to resource constraints; one after randomisation and four during follow-up. Although only a small proportion (5.8%) of the hospitals randomised in this cluster-randomised trial withdrew, it is plausible that poorly performing institutions would be more likely to withdraw than those with average or high performance.

Selective reporting

Only Tu 2009 registered a trial protocol with ClinicalTrials.gov (NCT00187460) in advance of undertaking the study. All outcomes described in this protocol were presented in the final report, which also included all-cause mortality as an additional outcome.

Therefore, we judged it to be at low risk of reporting bias. Although Zhang 2016 presented a trial protocol, this was published in March 2015, eighteen months after the intervention began in October 2013. None of the remaining ten studies registered a protocol in advance of randomisation (randomised and non-randomised trials) or data analysis (interrupted time series and controlled before-after series).

Other potential sources of bias

As outlined in the 'Assessment of risk of bias in included studies' section, the four cluster-randomised trials (Farley 2002a; Tu 2009; Ikkersheim 2013; Zhang 2016), cluster-non-randomised trial (Farley 2002b), and controlled before-after study (Rinke 2015), were assessed for bias in terms of baseline characteristics, baseline outcome measures, and protection against contamination. In addition, we assessed these sources of bias for the six interrupted time series studies: intervention is independent of other changes, shape of the intervention is prespecified, intervention is unlikely to affect data collection, and knowledge of the allocated interventions is adequately prevented during the study (Romano 2004; Jang 2011; Flett 2015; DeVore 2016; Joynt 2016; Liu 2017).

Baseline characteristics

We considered four studies to be at low risk of bias for baseline characteristics because the intervention and control groups were shown to be similar (Tu 2009; Ikkersheim 2013; Rinke 2015; Zhang 2016). Two studies did not report baseline characteristics, and we considered them to be at unclear risk of bias (Farley 2002a; Farley 2002b).

Baseline outcome measures

All six interrupted time series studies presented baseline outcome measures that differed between the intervention and control groups. However, all six also used appropriate statistical techniques, including multivariable regression (Farley 2002b; Tu 2009; Ikkersheim 2013; Rinke 2015; Zhang 2016), and difference-in-differences analyses (Tu 2009; Rinke 2015; Zhang 2016) to account for differences in baseline between the groups. They were therefore all considered to be at low risk of bias from this source.

Protection against contamination

We judged three studies to be at low risk of contamination, either because they randomised healthcare professionals (Ikkersheim 2013), or because their intervention was sent by post, and so unlikely to reach individuals in the control group (Farley 2002a; Farley 2002b).

We assessed two studies to be at high risk. The authors of Tu 2009 stated that several hospitals in the delayed feedback group reported that they also initiated quality improvement activities after becoming aware that performance measures were due to be released publicly. As this was not quantified, it was difficult to determine the degree to which hospitals in the control group modified their activities in anticipation of having to publicly release performance data. We also assessed Rinke 2015 at high risk because hospitals in states that did not mandate healthcare-associated infection reporting might still have modified their practice, given that such laws were being introduced elsewhere in the USA.

We judged Zhang 2016 to be at unclear risk, because no specific efforts were taken to protect against contamination. However, it is not certain that their intervention (posters on bulletin boards in outpatient areas of intervention organisations) would necessarily have influenced behaviour in control institutions.

Intervention independent of other changes

In three interrupted time series studies, it was unclear whether the intervention occurred independently of other changes over time, or whether the outcome was influenced by other confounding variables and events during the study period (Romano 2004; Jang 2011; DeVore 2016). We judged the remaining three interrupted time series studies to be at low risk of bias. In the two studies that examined public reporting of healthcare-associated infections, this was because they analysed data from a number of states that introduced legislation at different times (Flett 2015; Liu 2017). We judged Joynt 2016 to be at low risk, because they did not demonstrate a substantial change in the postintervention period, so this was unlikely to be attributable to other factors.

Shape of intervention effect prespecified

Two interrupted time series studies prespecified the shape of the intervention effect, so we assessed both to be at low risk of bias in this domain (Romano 2004; Jang 2011). The remaining four interrupted time series studies did not, and we judged them to be at high risk.

Knowledge of the allocated interventions adequately prevented during the study

All six interrupted time series studies reported objective outcome measures, so we judged them to be at low risk of bias for this domain.

Intervention unlikely to affect data collection

The intervention was unlikely to affect data collection in any of the six interrupted time series studies, as all were undertaken retrospectively, using routinely collected data. In all cases, the methods of data collection were the same before and after the intervention. Therefore, we judged all six studies to be at low risk of bias.

Effects of interventions

See: **Summary of findings for the main comparison** Public reporting of performance data versus no public reporting

The studies included in this review used a wide range of different interventions, which are described in the 'Characteristics of included studies' tables. We presented the effect sizes reported by each outcome and study in Table 2, Table 3, Table 4, and Table 5, together with the relative effects, for ease of comparison between different study designs and outcome measures. We also provided a 'Summary of findings' table, together with our decisions on how we determined levels of certainty (Summary of findings for the main comparison; Appendix 2).

Primary outcomes

Changes in healthcare utilisation by consumers

This review provided an indication of the likely effect of public release of performance data on healthcare utilisation by consumers. There was low-certainty evidence from three studies that public release of performance data may make little or no difference to long-term healthcare utilisation by consumers. Two studies included data from over 18,294 insurance beneficiaries (Farley 2002a; Farley 2002b), and it was unclear how many consumers were analysed by Romano 2004.

There was low-certainty evidence from one study that public release of performance data can lead to small and transient effects on healthcare utilisation behaviour by consumers (Romano 2004). This study analysed hospital patient volumes following implementation of the California Hospital Outcomes Project, which classified acute hospitals as better, worse or neither better nor worse than expected, based on the adjusted-mortality of patients with acute myocardial infarction, or undergoing diskectomy. They found that hospitals, which were high performing for adjusted mortality from acute myocardial infarction, received higher volumes of acute myocardial infarction than expected in the third and fourth quarters after publication of the California Hospital Outcomes Project, although there was no measurable effect in the early period following publication. Similarly, inconsistent trends were observed amongst diskectomy patients; the only reported association was between high performing (low complication) hospitals and volume of patients with lumbar diskectomy. However, this effect size was very small (less than one additional patient per month per hospital), and so may not have been an important effect. Performance data from New York was released as part of the Cardiac Surgery Reporting System. Romano 2004 analysed Cardiac Surgery Reporting System data from New York, and found that high performing (low mortality) hospitals received a higher number of cases in the

month following publication of a report (74.5 actual cases versus 61.1 expected). In the six months following designation as a high performance outlier, hospitals admitted 24 (22%) additional patients for coronary artery bypass surgery, and within two months after designation as a low performance outlier, hospitals treated 11 (16%) fewer patients. However, all volume effects had disappeared within three months of data publication.

There was low-certainty evidence that suggested that public release of performance data might effect the behaviour of specific subgroups. For example, Farley 2002b reported that the subgroup of enrollees who actually read the Consumer Assessment of Healthcare Providers and Systems report chose plans with higher standardised Consumer Assessment of Healthcare Providers and Systems ratings than those in the control group (2.58 versus 1.81, P < 0.01). Similarly, Romano 2004 found that the only detectable changes in hospital volume were among patients undergoing coronary artery bypass grafting in New York, and this change was entirely driven by patients who identified as 'white and other race'. They did not find evidence that black or Hispanic patient volumes were affected by designating a hospital as a high coronary artery bypass graft mortality outlier.

It is possible that restrictions on patient choice might act as an effect modifier (Aggarwal 2017; Moscelli 2017). However, the interventions in Farley 2002a and Farley 2002b were presented as 'true' choices, since new insurance beneficiaries should not have been limited by concerns around cost and distance. Similarly, Romano 2004 studied hospital choice amongst elective surgical populations seeking treatment at hospitals within a single city.

Changes in healthcare decisions taken by healthcare providers (professionals and organisations)

This review provides some indication of the likely effect of public release of performance data on decision making by healthcare professionals. There was low-certainty evidence from four studies that public release of performance data may make little or no difference to decisions taken by healthcare professionals. These studies included three million births (Jang 2011), and 67 healthcare providers (Ikkersheim 2013; Flett 2015; Zhang 2016).

Two studies reported modest effects on some outcomes. Ikkersheim 2013 did not find any clear affect on referral patterns following public release of data about cataract surgery, or hip and knee replacement. However, there was a small effect on referrals for breast cancer, with general practitioners in the intervention group referring 1.0% more cases (P = 0.01) to hospitals per incremental percentage point on the report card scale of medical effectiveness. Similarly, Zhang 2016 found that the effect of displaying prescription performance data in outpatient areas varied across outcomes and disease groups. Public release of performance data did not change the number of prescriptions containing antibiotics in the bronchitis group, two or more antibiotics in the gastritis group, injections in the hypertension group, or antibiotic injec-

tions in the bronchitis and hypertension groups. Similarly, the average prescription cost did not change for patients with hypertension. However, public release of performance data did appear to reduce prescriptions containing antibiotics for gastritis (intervention effect -12.7%, P < 0.001), two or more antibiotics for gastritis (-3.8%, P = 0.005), injections for gastritis (-10.6%, P < 0.001), and antibiotic injections for gastritis (-10.7%, P < 0.001). Average antibiotic prescription cost fell for patients with bronchitis (-7.9%, P < 0.001) and gastritis (-5.7%, P = 0.005). These mixed findings were also complicated by evidence that public release of prescribing data increased prescriptions containing antibiotics for patients with hypertension (intervention effect 2.0%, P = 0.08), and injections for bronchitis (2.0%, P = 0.012).

One study found that the first public release of hospital caesarean section rate data may have slightly reduced the number of patients undergoing this procedure (-0.8%, P < 0.01), and that this persisted until the end of the study, 20 months later. However, further public releases of data did not exhibit any further effect on caesarean section rates (Jang 2011).

Finally, Flett 2015 did not find any evidence that mandatory public reporting of central line-associated bloodstream infections had any effect on blood culture testing or antibiotic utilisation in paediatric and neonatal intensive care units in the United States.

Changes in the healthcare utilisation decisions of purchasers

We found no evidence on the effect of public release of performance data on this outcome.

Changes in provider performance

This review provides some indication of the likely effect of public release of performance data on healthcare provider performance. There was low-certainty evidence from one study that public release of performance data may make little or no difference to objective measures of provider performance. Tu 2009 included data from 82 healthcare providers.

Tu 2009 found that a media campaign and release of hospital performance data online had no effect on 11 of 12 acute myocardial infarction process-of-care quality indicators. The twelfth acute myocardial infarction quality indicator (fibrinolytics given prior to transfer to the Coronary Care Unit or Intensive Care Unit) increased by 5.8% (P = 0.02), although no statistical correction was made for multiple hypothesis testing. Similarly, public release of performance data did not clearly effect five of six congestive heart failure quality indicators, although the sixth (Angiotensin-Converting Enzyme (ACE) inhibitor or Angiotensin Receptor Blocker (ARB) for left ventricular dysfunction) increased by 5.9% (P = 0.02). Neither the acute myocardial infarction nor congestive heart failure composite process-of-care quality indicators improved following the public release of performance data.

The main outcomes in two studies described above, are sometimes considered evidence of provider performance (Jang 2011; Zhang

2016). However, as these outcomes (caesarean section and antibiotic prescribing) may be appropriate clinical decisions, they are not direct evidence of poor performance, so we have considered them under 'Changes in healthcare decisions taken by healthcare providers (professionals and organisations)' instead of 'Provider performance'.

Changes in patient outcome

Low-certainty evidence from five studies suggested that public release of performance data may slightly improve patient outcomes. We graded the certainty as low, because the evidence was mixed, with two studies reporting improvements (Tu 2009; Liu 2017), and three finding no evidence of improved patient outcomes (Rinke 2015; DeVore 2016; Joynt 2016). These five studies included 7503 healthcare providers and 315,092 hospitalisations. Two studies reported that patient outcomes were not changed by publication of hospital-level quality metrics on Hospital Compare, which is a website run by the Centers for Medicare & Medicaid Services. DeVore 2016 did not find any evidence that publication of hospital re-admission rates had an effect on 30-day re-admissions for patients with myocardial infarction, heart failure, or pneumonia. Similarly, Joynt 2016 reported a very small slowing in a pre-existing trend (change 0.13% per quarter; 95% CI 0.12% to 0.14%) towards reduced 30-day mortality following publication of mortality rates on Hospital Compare.

Rinke 2015 did not find any evidence that mandatory hospital reporting of central line-associated blood stream infections had any effect on the rate of paediatric central line-associated bloodstream infections. However, Liu 2017 reported a 34% reduction (incidence rate ratio 0.66, P < 0.001) in adult central line-associated bloodstream infections after mandatory reporting, when compared with the 25-month period before each state introduced legislation. This discrepancy between the findings of Rinke 2015 and Liu 2017 might reflect a genuine difference in terms of impact on children and adult central line-associated bloodstream infection rates. Importantly, both studies found that central line-associated bloodstream infection rates declined across the USA during their study period, including in states that did not introduce mandatory reporting. It is unclear whether public release of performance data in some states contributed to this national decline, even within states that did not introduce mandatory reporting.

Tu 2009 found that public release of hospital performance data online and through the media was associated with a 2.5% reduction in 30-day mortality (P = 0.045) for patients with acute myocardial infarction, although no such effect was observed in patients with congestive heart failure.

Changes in staff morale

We found no evidence on the effect of public release of performance data on this outcome.

Secondary outcomes

Unintended and adverse effects or harms

We found no evidence on the effect of public release of performance data on this outcome.

Impact on equity

Low-certainty evidence from one study suggested that public release of performance data may have different effects on advantaged and disadvantaged populations (Romano 2004). As described in 'Changes in healthcare utilisation by consumers', this study reported that patients who identified as white and other race in New York might have been influenced by publicly released hospital mortality rates when choosing a hospital in which to undergo coronary artery bypass grafting. However, this same effect was not observed in black or Hispanic patients undergoing the same procedure at hospitals in New York.

Other outcome measures

Two studies reported on awareness, knowledge of performance data, attitude, and cost data (Farley 2002b; Ikkersheim 2013). Farley 2002b reported secondary outcomes as a result of a survey, although this was disseminated using a 3:1 ratio, and the results were further complicated by low response rates. Ikkersheim 2013 undertook semi-structured interviews with 17 GPs but these were largely focused on the specific intervention (report cards) and the findings were poorly reported. Therefore, we decided to exclude these results, and did not report these outcomes.

D I S C U S S I O N

Summary of main results

Changes in healthcare utilisation by consumers

Changes in healthcare utilisation are one of the two key ways in which public release of performance data might improve healthcare quality (Berwick 2003). However, only three studies addressed the impact on healthcare utilisation decisions by consumers (Farley 2002a; Farley 2002b; Romano 2004). We judged that they provided low-certainty evidence of little or no effect. There were consistent results from two studies that showed some consumers may engage with published performance data, and change their healthcare choices accordingly; this group was too small to register an effect in the population as a whole (Farley 2002b; Romano 2004).

Changes in healthcare decisions taken by healthcare providers (professionals and organisations)

There was low-certainty evidence with mixed findings from four studies, which reported either modest effects (Jang 2011; Ikkersheim 2013; Zhang 2016), or no effect (Flett 2015), on healthcare decisions taken by healthcare providers. Two studies found evidence that public release of performance data had modest effects on some of the healthcare decisions taken by healthcare providers, but not all of the decisions measured (Ikkersheim 2013; Zhang 2016). One study found that the first public release of data had a small but sustained effect on caesarean rates, but that subsequent releases did not affect the rate any further (Jang 2011).

Changes in provider performance

There was low-certainty evidence from one study that informed conclusions about the effect of public release of performance data on provider performance. A single randomised trial addressed this question, and found that 2/18 (11.1%) of measured processes appeared to improve in the intervention hospitals (Tu 2009). However, as no correction was made for multiple hypothesis testing (Bender 2001), this did not provide convincing evidence that provider performance was affected by public release of performance data.

Changes in patient outcome

Low-certainty evidence showed that five studies that included patient outcomes had inconsistent findings, with two reporting improvements (Tu 2009; Liu 2017), and three reporting no difference (Rinke 2015; DeVore 2016; Joynt 2016).

Impact on equity

Only one study undertook a subgroup analysis to identify differential effects of public release of performance data (Romano 2004). Low-certainty evidence from one study reported that white and other race patients, undergoing coronary artery bypass grafting in New York, may have been influenced by publicly released mortality rates. However, this finding was not reproduced among black and Hispanic patients. Although Farley 2002b did not study equity directly, their finding that only consumers who read the Consumer Assessment of Healthcare Providers and Systems report were influenced, raises the possibility that some groups (e.g. those with greater rates of literacy) might be preferentially influenced by public release of performance data.

Other outcomes

There were no studies that considered the effect of public release of performance data on changes in the healthcare utilisation decisions of purchasers, changes in staff morale, or adverse effects.

Two studies reported on awareness, knowledge of performance data, attitude, and cost data but we did not include the data due to concerns about reporting and high attrition bias.

Overall completeness and applicability of evidence

There are many systems around the world that include public release of performance data. However, only a small proportion were represented in this review, so it is likely that most have either not been evaluated, or were subject only to low-quality studies. It is notable that some interventions have been evaluated more robustly than others, with two studies in this review considering the Centers for Medicare & Medicaid Services website Hospital Care (DeVore 2016; Joynt 2016), and three, the introduction of state-based mandatory reporting of central line-associated blood stream infections (Flett 2015; Rinke 2015; Liu 2017). Similarly, the majority of the studies included in this review (9/12, 75%) were based in North America, with no representation from South America, Africa, or Australasia. Therefore, it is likely that a small number of initiatives have attracted a disproportionate number of studies, and there is clearly work that needs to be done to robustly evaluate similar interventions in other settings. There was also insufficient evidence to draw any conclusions about the healthcare utilisation decisions of purchasers, staff morale, or adverse effects. The applicability of the evidence was also limited by considerable heterogeneity in interventions. For example, it was possible that the freedom of patients to choose healthcare providers was curtailed in some cases, which might have acted as an effect modifier that explains some of the differences in findings between included studies. However, only three studies included interventions that might lead to improved consumer selection, and consumer choice would not obviously have been restricted by considerations around distance and cost (Farley 2002a; Farley 2002b; Romano 2004). These studies suggested that those engaging with publicly reported performance data (Farley 2002b), and those from privileged backgrounds (Romano 2004), might be more likely to modify their choice of healthcare provider. This raises the possibility that lack of education and health literacy might restrict patient choice, and act as an effect modifier in some cases.

The three studies that took place in the USA involved only a small proportion of the numerous major reporting systems available. We included one new study from Canada, which was published after the latest systematic reviews by Fung 2008, Shekelle 2008, and Faber 2009 (Tu 2009). We excluded many of the more recent studies, because they did not have a rigorous study design, or did not report the defined primary outcome measures. The studies we included evaluated interventions that used data that might have been originally collected for a purpose other than influencing behaviour or improving outcomes. It is possible that custom-made interventions, using data collected for the specific purpose of influencing behaviour or improving outcomes, would have a greater

impact. However, the lack of such interventions in the literature highlighted the fact that their delivery may be excessively resource intensive, and that future initiatives aimed at public release of performance data will continue to draw on data initially collected for a different purpose.

Despite evidence that secondary outcome measures (e.g. awareness, attitude, knowledge of performance data) are crucial, since public reporting can only change behaviour if the target population (healthcare consumers, providers, or purchasers of care) understands the information, these measures were lacking in the included studies (Hibbard 2010). Therefore, it was difficult to explain the lack of effect. For example, Faber 2009 found that the effect of performance data was higher for those who understand the information, which might be consistent with the evidence from Farley 2002b. Damman 2011 showed that comparative performance information was complex, and consumers had difficulties in interpreting and using performance data. However, it is notable that this review did not find that healthcare providers (who might be in a better position to interpret such data) were necessarily influenced more than consumers.

Certainty of the evidence

We deemed the certainty of the evidence that examined the effect of public release of performance data on a number of outcomes to be low. These outcomes were:

- Changes in healthcare utilisation by consumers;
- Changes in healthcare utilisation by providers (organisations and professionals); and
- Changes in patient outcome.

Only 4/12 included studies (33.3%) were randomised trials, so the evidence for these outcomes was partly informed by non-randomised study designs. However, the use of EPOC study design criteria ensured that all included observational studies took considerable steps to minimise the risk of bias (EPOC 2013). There was also considerable heterogeneity in the settings, outcomes, and modes of public release, and inconsistent effects reported between studies.

We also judged the certainty of the evidence that examined the effect on changes in provider performance to be low. Although this outcome was informed by a single randomised trial, we had concerns about risk of bias in the following items: (1) allocation concealment, (2) adequate blinding of participants, personnel and outcome assessors, and (3) protection against contamination (Tu 2009). It is also uncertain whether the findings of a single randomised trial, in a narrowly defined patient group, within one region of Canada, can be generalised to other settings.

Due to lack of evidence, we were unable to draw any conclusions about the following primary outcomes:

- Changes in the healthcare utilisation decisions of purchasers;
- Changes in staff morale.

In terms of secondary outcomes, there were no studies that set out to consider adverse effects or harms. We deemed the evidence for any potential impact on equity to be low, as it was based on a subgroup analysis from a single interrupted time series study (Romano 2004).

Potential biases in the review process

Although our search was comprehensive, we could not exclude the possibility of having missed relevant studies. However, we minimised this risk by asking an Information Specialist to help design and implement the search strategy, and ensured that two review authors independently examined all items retrieved from our search. We also ensured that data extraction and 'Risk of bias' assessments were independently undertaken by two review authors. Although the GRADE assessments were determined by a single author (DM), these were checked by all review authors, and disagreements resolved through discussion. These steps ensured that potential biases in the review processes were mitigated as much as possible. However, this stringent approach to study collection also meant excluding most of the studies that have evaluated public release of performance data in other settings, and using a range of study designs. It was possible that this approach biased our review against settings that were less likely to deliver studies that satisfied the EPOC inclusion criteria, and this might have accounted for the over-representation of studies from North America, Europe, and Asia. It might also have led to the exclusion of studies (e.g. those utilising qualitative designs) that contained important information about the impact of public release of performance data. However, it was necessary to limit our review to studies that were at the lowest possible risk of bias, to maximise the certainty of its findings. There may nevertheless be scope for future reviews to synthesise evidence from studies using a broader range of designs.

Agreements and disagreements with other studies or reviews

Our systematic literature search and a further PubMed search of studies citing an earlier version of this review (Ketelaar 2011), identified three relevant systematic reviews (Fung 2008; Faber 2009; Campanella 2016). Our review agreed with these earlier publications that previous studies were limited by risk of bias, inconsistent findings, and heterogeneity of interventions, healthcare settings, and outcomes.

Faber 2009 considered public release of performance data on consumer choice, and concluded that there was only evidence to support an effect on the small subgroup of participants that actively engaged with the published performance data. This was consistent with our findings, and those of Fung 2008.

Campanella 2016 attempted a meta-analysis of data from ten studies, and reported improved mortality (risk ratio 0.85, 95% confi-

dence interval 0.79 to 0.92). However, this finding was reported in the context of very high heterogeneity (P < 0.0001; I² = 100%). The authors limited their meta-analysis to studies that reported sufficient data, and excluded those with inappropriate study designs, or those that were at high risk of bias. Our review only considered studies that proffered the highest certainty of evidence, and did not consider a meta-analysis appropriate in view of the considerable degree of heterogeneity between studies (see Assessment of heterogeneity). Instead, our findings were consistent with those of Fung 2008, which concluded that "studies of the effect of public reporting on outcomes provide mixed signals, and the usefulness of public reporting in improving patient safety and patient-centeredness remains unknown, because few studies assessed these end points".

AUTHORS' CONCLUSIONS

Implications for practice

The existing evidence base on the effects of public release of performance data on changing behaviour of healthcare decision makers was inadequate to directly inform practice.

Implications for research

In order to understand the effectiveness of the public release of performance data, we need more longitudinal studies with robust evaluation designs. In particular, the evidence base would benefit from more studies that consider whether public release of performance data can improve patient outcomes, rather than simply healthcare processes. In this review, only one of the included studies reported data on patient outcomes (Tu 2009). Further work should also specifically consider whether public release of performance data might result in adverse effects or harms.

Unfortunately, most studies were unable to guarantee that disseminated performance data actually reached its intended audience, i.e. that lack of effect was not simply a result of failed exposure to the intervention. Importantly, Farley 2002b reported evidence to suggest that the subgroup of patients that read the reports sent by post were influenced when choosing a health insurance programme. Therefore, future studies should consider carefully how they might maximise the number of people exposed to their intervention, and whether this can be quantified. However, the effect of public release of performance information in the 'real world' is likely to be limited by difficulties in reaching its intended audience (Hibbard 2007; Damman 2010; Aggarwal 2017; Canaway 2017; Moscelli 2017; Canaway 2018; Greenhalgh 2018). Therefore, the need to ensure that performance data reach those who are intended to be influenced, needs to be balanced against the risk of reducing study validity by creating artificial conditions that cannot be replicated when the intervention is used in practice.

Berwick's model suggests that public release of performance data may improve quality of care by means of a pathway of change or selection (Berwick 2003). The studies we included focused exclusively on either one or the other of these pathways. In addition, one assumption underlying public release of performance data is that provider choice is a rational decision, i.e. consumers prefer the healthcare provider or health plan that is rated as the best. However, there is little evidence to confirm this assumption (Faber 2009; Kolstad 2009), although a number of other factors are known to influence consumer choice, e.g. established relationships with local physicians, health plans (Schwartz 2005; Hibbard 2009), hospitals, distance, and opinions of friends, and family (Harris 2008; The King's Fund 2010). Similarly, Ikkersheim 2013 found that decisions taken by healthcare professionals were often informed by their personal preferences, experience of, and communication with other providers, and personal relationships with other professionals. These factors influenced hospital referral decisions even when professionals were provided with objective performance data. Future studies may wish to consider the mechanism(s) by which public release of performance data can effect change, as well as whether such changes can be demonstrated empirically.

ACKNOWLEDGEMENTS

reflect those of the Systematic Reviews Programme, NIHR, NHS or the Department of Health.

REFERENCES

References to studies included in this review

DeVore 2016 *{published data only}*
DeVore AD, Hammill BG, Hardy NC, Eapen ZJ, Peterson ED, Hernandez AF. Has public reporting of hospital readmission rates affected patient outcomes? Analysis of Medicare claims data. *Journal of the American College of Cardiology* 2016;**67**(8):963–72. PUBMED: 26916487]

Farley 2002a *{published data only}*
* Farley DO, Elliott MN, Short PF, Damiano P, Kanouse DE, Hays RD. Effect of CAPHS Performance Information on health plan choices by Iowa Medicaid. *Medical Care Research and Review* 2002;**59**(3):319–36. PUBMED: 12205831]

Farley 2002b *{published data only}*
* Farley DO, Short PF, Elliott MN, Kanouse DE, Brown JA, Hays RD. Effect of CAPHS health plan performance information on plan choices by New Jersey. *Health Services Research* 2002;**37**(4):985–1007. PUBMED: 12205831]

Flett 2015 *{published data only}*
Flett KB, Ozonoff A, Graham DA, Sandora TJ, Priebe GP. Impact of mandatory public reporting of central line-associated bloodstream infections on blood culture and antibiotic utilization in pediatric and neonatal intensive care units. *Infection Control and Hospital Epidemiology* 2015;**36**(8):878–85. PUBMED: 25913602]

Ikkersheim 2013 *{published data only}*
Ikkersheim D, Koolman X. The use of quality information by general practitioners: does it alter choices? A randomized clustered study. *BMC Family Practice* 2013;**14**:95. PUBMED: 23824745]

Jang 2011 *{published data only}*
Jang WM, Eun SJ, Lee C, Kim Y. Effect of repeated public releases on cesarean section rates. *Journal of Preventative Medicine and Public Health* 2011;**44**(1):2–8. PUBMED: 21483217]

Joynt 2016 *{published data only}*
Joynt KE, Orav EJ, Zheng J, Jha AK. Public reporting of mortality rates for hospitalized Medicare patients and trends in mortality for reported conditions. *Annals of Internal Medicine* 2016;**165**(3):153–60. PUBMED: 27239794]

Liu 2017 *{published data only}*
Liu H, Herzig CTA, Dick AW, Furuya EY, Larson E, Reagan J, et al. Impact of state reporting laws on central line-associated bloodstream infection rates in U.S. adult intensive care units. *Health Services Research* 2017;**52**(3):1079–98. PUBMED: 27451968]

Rinke 2015 *{published data only}*
Rinke ML, Bundy DG, Abdullah F, Colantuoni E, Zhang Y, Miller MR. State-mandated hospital infection reporting is not associated with decreased pediatric health care-associated infections. *Journal of Patient Safety* 2015;**11**:123–34. PUBMED: 24681422]

Romano 2004 *{published data only}*
* Romano PS, Zhou H. Do well-publicized risk-adjusted outcomes reports affect hospital volume?. *Medical Care* 2004;**42**(4):367–77. PUBMED: 15076814]

Tu 2009 *{published data only}*
Tu JV, Donovan LR, Douglas SL, Wang JT, Austin PC, Alter DA, et al. Effectiveness of public report cards for improving the quality of cardiac care. The EFFECT study: a randomized trial. *JAMA* 2009;**302**(21):2330–7. PUBMED: 19923205]

Zhang 2016 *{published data only}*
Du X, Wang D, Wang X, Yang S, Zhang X. Exploring the transparency mechanism and evaluating the effect of public reporting on prescription: a protocol for a cluster randomized controlled trial. *BMC Public Health* 2015;**21**(15):277. PUBMED: 25881035]
Liu C, Zhang X, Wan J. Public reporting influences antibiotic and injection prescription in primary care: a segmented regression analysis. *Journal of Evaluation in Clinical Practice* 2015;**21**(4):597–603. PUBMED: 25902726]
* Liu C, Zhang X, Wang X, Wan J, Zhong F. Does public reporting influence antibiotic and injection prescribing to all patients? A cluster-randomized matched-pair trial in China. *Medicine (Baltimore)* 2016;**95**(26):e3965. PUBMED: 27367995]
Tang Y, Liu C, Zhang X. Public reporting as a prescriptions quality improvement measure in primary care settings in China: variations in effects associated with diagnoses. *Scientific Reports* 2016;**6**:39361. PUBMED: 27996026]
Wang X, Tang Y, Zhang X, Yin X, Du X, Zhang X. Effect of publicly reporting performance data of medicine use on injection use: a quasi-experimental study. *PLOS One* 2014;**9**(10):e109594. PUBMED: 25313853]
Yang L, Liu C, Wang L, Yin X, Zhang X. Public reporting improves antibiotic prescribing for upper respiratory tract infections in primary care: a matched-pair cluster-randomized trial in China. *Health Research Policy and Systems* 2014;**12**:61. PUBMED: 25304996]

References to studies excluded from this review

Cavender 2015 *{published data only}*
Cavender MA, Joynt KE, Parzynski CS, Resnic FS, Rumsfeld JS, Moscucci M, et al. State mandated public reporting and outcomes of percutaneous coronary intervention in the United States. *American Journal of Cardiology* 2015;**115**(11):1494–501. PUBMED: 25891991]

Moscucci 2005 *{published data only}*

Moscucci M, Eagle KA, Share D, Smith D, De Franco AC, O'Donnell M, et al. Public reporting and case selection for percutaneous coronary interventions: an analysis from two large multicenter percutaneous coronary intervention databases. *Journal of the American College of Cardiology* 2005;**45**(11):1759–65. PUBMED: 15936602]

Paris 2013 *{published data only}*

Paris B, Arahood T, Asche C, Amundson G. Lessons from voluntary reporting of Illinois hospital employee seasonal influenza vaccination rates (2009-2013). *Value in Health* 2013;**16**(3):A96.

Park 2011 *{published data only}*

Park J, Konetzka RT, Werner RM. Performing well on nursing home report cards: does it pay off?. *Health Services Research* 2011;**46**(2):531–54. PUBMED: 21029093]

Saratzis 2017 *{published data only}*

Saratzis A, Thatcher A, Bath MF, Sidloff DA, Bown MJ, Shakespeare J, et al. Reporting individual surgeon outcomes does not lead to risk aversion in abdominal aortic aneurysm surgery. *Annals of the Royal College of Surgeons of England* 2017;**99**(2):161–5. PUBMED: 28071950]

Additional references

Aggarwal 2017

Aggarwal A, Lewis D, Mason M, Sullivan R, van der Meulen J. Patient mobility for elective secondary health care services in response to patient choice policies: a systematic review. *Medical Care Research and Review: MCRR* 2017;**74**(4):379–403. [PUBMED: 27357394]

Bender 2001

Bender R, Lange S. Adjusting for multiple testing - when and how?. *Journal of Clinical Epidemiology* 2001;**54**(4):343–9. [PUBMED: 11297884]

Berwick 1990

Berwick DM, Wald DL. Hospital leaders' opinions of the HCFA mortality data. *JAMA* 1990;**263**(2):247–9.

Berwick 2003

Berwick D, Jamer B, Coye M. Connections between quality measurements and improvement. *Medical Care* 2003;**41**(1):130–8.

Brook 1994

Brook RH. Health care reform is on the way: do we want to compete on quality?. *Annals of Internal Medicine* 1994;**120**(1):84–6.

Burns 2016

Burns EM, Pettengell C, Athanasiou T, Darzi A. Understanding the strengths and weaknesses of public reporting of surgeon-specific outcome data. *Health Affairs* 2016;**35**(3):415–21.

Campanella 2016

Campanella P, Vukovic V, Parente P, Sulejmani A, Ricciardi W, Specchia ML. The impact of public reporting on clinical outcomes: a systematic review and meta-analysis. *BMC Health Services Research* 2016;**16**:296.

Canaway 2017

Canaway R, Bismark M, Dunt D, Kelaher M. Perceived barriers to effective implementation of public reporting of hospital performance data in Australia: a qualitative study. *BMC Health Services Research* 2017;**17**(1):391. [PUBMED: 28592277]

Canaway 2018

Canaway R, Mismark M, Dunt D, Prang KH, Kelaher M. "What is meant by public?": stakeholder views on strengthening impacts of public reporting of hospital performance data. *Social Science & Medicine* 2018;**202**:143–50. [PUBMED: 29524870]

Chung 2014

Chung SH, Seol HJ, Choi YS, Oh SY, Kim A, Bae CW. Changes in the cesarean section rate in Korea (1982-2012) and a review of the associated factors. *Journal of Korean Medical Science* 2014;**29**(10):1341–52. [PUBMED: 25368486]

Damman 2010

Damman OC, Van den Hengel YK, Van Loon AJ, Rademakers J. An international comparison of Web-based reporting about healthcare quality: content analysis. *Journal of Medical Internet Research* 2010;**13**(12):e8.

Damman 2011

Damman OC, Hendrik M, Rademakers J, Spreeuwenberg P, Delnoij DM, Groenewegen PP. Consumers interpretation and use of comparative information on the quality of health care: the effect of presentation approaches. Health Expectations 25 May 2011 Epub ahead of print]. DOI: 10.1111/j.1369-7625.2011.00671.x

EPOC 2013

Effective Practice, Organisation of Care (EPOC). Analysis in EPOC reviews. EPOC resources for review authors. Available at: epoc.cochrane.org/epoc-resources-review-authors (accessed 22 August 2018).

EPOC 2017

Effective Practice, Organisation of Care (EPOC). Reporting the effects of an intervention in EPOC reviews. Available at: epoc.cochrane.org/epoc-resources-review-authors (accessed 22 August 2018).

Faber 2009

Faber M, Bosch M, Wollersheim H, Leatherman S, Grol R. Public reporting in health care: how do consumers use quality of care information? A systematic review. *Medical Care* 2009;**47**(1):1–8.

Fung 2008

Fung C, Yee-Wei L, Soeren M, Damberg C, Shekelle P. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Annals of Internal Medicine* 2008;**148**(2):111–23.

Greenhalgh 2018

Greenhalgh J, Dalkin S, Gibbons E, Wright J, Valderas JM, Meads D, et al. How do aggregated patient-reported outcome measures data stimulate health care improvement? A realist synthesis. *Journal of Health Services Research & Policy* 20183;**23**(1):57–65. [PUBMED: 29260592]

Harris 2008

Harris KM, Beeuwkes Buntin M, The RAND Cooperation. *Research Synthesis Report. Choosing a healthcare provider: the role of quality information.* Princeton: Robert Wood Johnson Foundation, 2008.

Hendriks 2009

Hendriks M, Spreeuwenberg P, Rademakers J, Delnoij DM. Dutch healthcare reform: did it result in performance improvement of health plans? A comparison of consumer experiences over time. *BMC Health Services Research* 2009; **9**:167. DOI: 10.1186/1472-6963-9-167

Hibbard 1997

Hibbard JH, Jewett JJ, Legnini MW, Tusler M. Choosing a health plan: do large employers use the data?. *Health Affairs* 1997;**16**(6):172–80.

Hibbard 2007

Hibbard JH, Peters E, Dixon A, Tusler M. Consumers competencies and the use of comparative quality information: it isn't just about literacy. *Medical Care Research and Review* 2007;**64**(4):379–94.

Hibbard 2009

Hibbard JH. Using systematic measurement to target consumer activation strategies. *Medical Care Research and Review* 2009;**66**(1):9S–27S.

Hibbard 2010

Hibbard JH, Greene J, Daniel D. What is quality anyway? Performance reports that clearly communicate to consumers the meaning of quality of care. *Medical Care Research and Review* 2010;**67**(3):275–93.

Higgins 2011

Higgins JPT, Green S, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Huwaldt 2004 [Computer program]

Huwaldt JA, Steinhorst S, Paul G. Plot Digitizer. Version accessed 27 October 2015. USA: Huwaldt JA; SourceForge, 2004.

Ivers 2012

Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews* 2012, Issue 6. DOI: 10.1002/14651858.CD000259.pub3; PUBMED: 22696318

Kiernan 2015

Kiernan F, Rahman F. Measuring surgical performance: a risky game?. *Surgeon* 2015;**13**(4):213–7.

Kolstad 2009

Kolstad JT, Chernew ME. Quality and consumer decision making in the market for health insurance and health care services. *Medical Care Research and Review* 2009;**66**(Suppl 1):28S–52S.

Liu 2015

Liu C, Zhang X, Wan J. Public reporting influences antibiotic and injection prescription in primary care:

a segmented regression analysis. *Journal of Evaluation in Clinical Practice* 2015;**21**(4):597–603. [PUBMED: 25902726]

Liu 2016

Liu C, Zhang X, Wang X, Zhang X, Wan J, Zhong F. Does public reporting influence antibiotic and injection prescribing to all patients? A cluster-randomized matched-pair trial in China. *Medicine (Baltimore)* 2016;**95**(26): e3965. [PUBMED: 27367995]

Loeb 2004

Loeb JM. The current state of performance measurement in health care. *International Journal for Quality in Health Care* 2004;**16**(Suppl 1):i5–9.

Marshall 2000

Marshall MN, Shekelle PG, Leatherman S, Brook RH. The public release of performance data: what do we expect to gain? A review of the evidence. *JAMA* 2000;**283**(14): 1866–74. [PUBMED: 10770149]

Moher 1999

Moher D, Cook DJ, Eastwood S, Olkin I, Drummon R, Stroup DF, et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet* 1999;**354**(9193):1896–900.

Moscelli 2017

Moscelli G, Siciliani L, Gutacker N, Cookson R. Socioeconomic inequality of access to healthcare: Does choice explain the gradient?. Journal of Health Economics 23 June 2017 Epub ahead of print]. DOI: 10.1016/j.jhealeco.2017.06.005; PUBMED: 28935158

Mukamel 1998

Mukamel DB, Mushlin AI. Quality of care information makes a difference: an analysis of market share and price changes after publication of the New York State Cardiac Surgery Mortality Reports. *Medical Care* 1998;**36**(7): 945–54.

Rosenthal 1998

Rosenthal GE, Hammar PJ, Way LE, Shipley SA, Doner D, Wojtala B, et al. Using hospital performance data in quality improvement: the Cleveland Health Quality Choice experience. *Joint Commission Journal of Quality Improvement* 1998;**24**(7):347–60.

Schroll 2011

Schroll JP, Moustgaard R, Gøtzsche PC. Dealing with substantial heterogeneity in Cochrane reviews. Cross-sectional study. *BMC Medical Research Methodology* 2011; **11**:22. [PUBMED: 21349195]

Schut 2005

Schut FT, van de Ven WP. Rationing and competition in the Dutch health-care system. *Health Economics* 2005;**14** (Suppl 1):S59–74.

Schwartz 2005

Schwartz LM, Woloshin S, Birkmeyer JD. How do elderly patients decide where to go for major surgery? Telephone

interview survey. *BMJ* 2005;**331**(7520):821. DOI: 10.1136/bmj.38614.449016.DE

Shahian 2017

Shahian DM, Jacobs JP, Badhwar V, D'Agostino RS, Bavaria JE, Prager RL. Risk aversion and public reporting. Part 1: observations from cardiac surgery and interventional cardiology.. *Annals of Thoracic Surgery* 2017;**104**(6): 2093–101. [PUBMED: 29100643]

Shekelle 2008

Shekelle PG, Lim Y-W, Mattke S, Damberg C, Southern California Evidence-based Practice Centre, RAND Corporation. *Does public release of performance results improve quality of care? A systematic review.* London: The Health Foundation, 2008.

Sherman 2013

Sherman KL, Gordon EJ, Mahvi DM, Chung J, Bentrem DJ, Holl JL, et al. Surgeons' perceptions of public reporting of hospital and individual surgeon quality. *Medical Care* 2013;**51**(12):1069–75.

Sirio 1996

Sirio CA, McGee JL. Public reporting of clinical outcomes - the data needs of health care stakeholders. *American Journal of Medical Quality* 1996;**11**(1):S78–81.

Smith 2009

Smith PC, Mossialos E, Papanicolas I, Leatherman S. *Performance measurement for health system improvement. Experiences, Challenges and Prospects.* Cambridge: Cambridge University Press, 2009.

Tang 2016

Tang Y, Liu C, Zhang X. Public reporting as a prescriptions quality improvement measure in primary care settings in China: variations in effects associated with diagnoses. *Scientific Reports* 2016;**6**:39361. [PUBMED: 27996026]

The King's Fund 2010

Dixon A, Roberson R, Appleby J, Burge P, Devlin N, Magee H. *Patient Choice. How patients choose and how providers respond.* London: The King's Fund, 2010.

Wadhera 2017

Wahera RK, Anderson JD, Yeh RW. High-risk percutaneous coronary intervention in public reporting states: the

evidence, exclusion of critically ill patients, and implications. *Current Heart Failure Reports* 2017;**14**(6):514–8. [PUBMED: 29101664]

Wang 2014

Wang X, Tang Y, Zhang X, Yin X, Du X, Zhang X. Effect of publicly reporting performance data of medicine use on injection use: a quasi-experimental study. *PLoS One* 2014;**9** (10):e109594. [PUBMED: 25313853]

Wasfy 2015

Wasfy JH, Borden WB, Secemsky EA, McCabe JM, Yeh RW. Public reporting in cardiovascular medicine: accountability, unintended consequences, and promise for improvement.. *Circulation* 2015;**131**(17):1518–27. [PUBMED: 25918041]

Werner 2009

Werner RM, Konetzka RM, Stuart EA, Norton EC, Polsky D, Park J. Impact of public reporting on quality of postacute care. *Health Services Research* 2009;**44**(4):1169–87.

Yang 2014

Yang L, Liu C, Wang L, Yin X, Zhang X. Public reporting improves antibiotic prescribing for upper respiratory tract infections in primary care: a matched-pair cluster-randomized trial in China. *Health Research Policy and Systems* 2014;**12**:61. [PUBMED: 25304996]

Zhang 2018 [pers comm]

Zhang X (School of Medicine and Health Management, Tongji Medical College, HuaZhong University of Science and Technology, Wuhan, Hubei Province, China). [personal communication]. David Metcalfe (Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences [NDORMS], University of Oxford, UK) 15 March 2018.

References to other published versions of this review

Ketelaar 2011

Ketelaar NA, Faber MJ, Flottorp S, Rygh LH, Deane KH, Eccles MP. Public release of performance data in changing the behaviour of healthcare consumers, professionals or organisations. *Cochrane Database of Systematic Reviews* 2011, Issue 11. DOI: 10.1002/14651858.CD004538

* *Indicates the major publication for the study*

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

DeVore 2016

Methods	**Design:** ITS **Country:** USA **Care setting:** acute hospitals **Duration:** 1 July 2006 to 30 June 2012 **Dataset:** 5% nationally representative sample of Medicare beneficiaries **Total participants:** 315,092 hospitalisations **Unit of analysis:** individual hospitalisations; accounted for clustering within hospitals **Data analysis:** regression models
Participants	**Inclusion criteria:** all patients enrolled with Medicare, i.e. predominantly those aged 65 years or older **Hospitals:** more than 4,100 hospitals in the USA **Participants:** 315,092: 37,829 acute myocardial infarction (16.0%), 100,189 heart failure (42.5%), 17,907 diabetes (7.6%), 80,091 chronic obstructive pulmonary disease (33.9%)
Interventions	**Intervention:** public reporting of risk-standardised hospital readmission rates on a public website, Hospital Compare **Duration:** June 2009 until the study end date in 2012 **Deliverer:** Centers for Medicare & Medicaid Services (CMS), US Department of Health and Human Services **Funding:** CMS (federal government funding)
Outcomes	**Main outcome** • 30-day post-discharge re-admission to hospital **Secondary outcomes** • 30-day post-discharge outpatients visits • 30-day post-discharge emergency department visits • 30-day post discharge observation stays without readmission
Notes	Abbreviations: Interrupted Time Series (ITS) study

Risk of bias

Bias	Authors' judgement	Support for judgement
Incomplete outcome data (attrition bias) All outcomes	Low risk	Outcome data available for all patients
Selective reporting (reporting bias)	Low risk	All outcomes and results outlined in the Method section were reported in tables, text, or both
Other bias	Low risk	No additional biases identified

Intervention is independent of other changes? (ITS)	Unclear risk	did not state whether there were other confounding events that might have changed performance over time
Shape of intervention effect pre-specified? (ITS)	High risk	Shape of intervention effect not pre-specified
Knowledge of the interventions adequately prevented during the study? (ITS)	Low risk	Individuals would not have been aware of the study, as this was performed retrospectively, using the Medicare data set
Intervention unlikely to affect / bias data collection? (ITS)	Low risk	Routinely collected administrative data, and so data collection was unlikely to be biased by the intervention

Farley 2002a

Methods	**Design:** cRT **Country:** USA (Iowa) **Care setting:** insurance plan beneficiaries in the community **Duration:** February to May 2000 **Dataset:** data provided by the Iowa Medicaid office **Total participants:** 13,077 **Unit of allocation:** household units **Unit of analysis:** individual Medicaid beneficiaries; accounted for clustering of beneficiaries within household units **Sample size calculation:** not reported; statistical significance was assessed at the 0.05 level **Data analysis:** multinomial logistic regression to model the outcomes (1) stayed in assigned HMO, (2) switched to another HMO, or (3) switched to MediPass
Participants	**Inclusion criteria:** all new cases (i.e. household units) newly eligible to participate in Iowa Medicaid **Participants:** 13,077 new beneficiaries in 7016 cases with 6515 beneficiaries in the control group and 6562 in the intervention group **Health plans:** two HMOs under contract with the Medicaid programme and 1 primary care case management plan (MediPass). One HMO scored more highly on the publicly reported performance measures than the other
Interventions	**Intervention:** standard enrolment materials and Consumer Assessment of Healthcare Providers and Systems (CAHPS) report delivered by personal mail **Control:** standard enrolment materials delivered by personal mail **Duration:** February to May 2000 **Deliverer:** the Iowa Medicaid office posted beneficiaries a packet health plan enrolment materials that included items, such as a plan enrolment form and the CAHPS report for the intervention group **Funding:** co-operative agreement 5U18HS09204-05; the Agency for Healthcare Research and Quality and the Center for Medicare and Medicare Services

Outcomes	**Main outcome**
	• decision to remain with allocated HMO, switch HMO, or switch to MediPass

Notes	The star charts in the CAHPS report were based on each HMO's performance. The bar charts included 3 charts with ratings of the health plan, health care, and personal doctor. Five charts were included by the providers or health plan
	Abbreviations: cluster randomised trial (cRT); health maintenance organization (HMO) ; Consumer Assessment of Healthcare Providers and Systems (CAHPs)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	The new cases enrolled during the study period were randomly assigned to an experimental or control group. This random assignment was independent of case size, county of residence, and initial plan assignment
Allocation concealment (selection bias)	Unclear risk	The new cases enrolled during the study period were randomly assigned to an experimental or control group
Adequate blinding of participants, personnel and outcome assessors?	Unclear risk	did not state whether or not participants knew that they were part of a study and so had been allocated to an intervention or control group
Incomplete outcome data (attrition bias) All outcomes	Low risk	Quote: 'Medicaid office supplied us with data files for the full sample of new beneficiaries'
Selective reporting (reporting bias)	Low risk	All outcomes and results outlined in the Method section were reported in tables, text, or both
Baseline characteristics similar?	Unclear risk	Did not explicitly describe baseline characteristics, although attempted to take these into account when determining risk-adjusted outcomes
Baseline outcomes similar?	Low risk	Did not explicitly describe baseline outcomes, but accounted for differences appropriately using multinomial logistic regression

Protection against contamination	Low risk	No specific safeguards against contamination, but reports were sent by post, so it was unlikely that the control group received the intervention
Other bias	Low risk	No additional biases identified

Farley 2002b

Methods	**Design:** cNRT (non-randomised as participants allocated based on their Medicaid case ID number) **Country:** USA (New Jersey) **Care setting:** insurance plan beneficiaries in the community **Duration:** March to October 1998 **Dataset:** data provided by the New Jersey Medicaid office **Total participants:** 5217 **Unit of allocation:** household units **Unit of analysis:** individual Medicaid beneficiaries; did not account for clustering of beneficiaries within household unit**s** **Sample size calculation:** not reported; statistical significance was assessed at the 0.05 level **Data analysis:** multivariable logistic regression, using enrolment with the dominant Healthcare Maintenance Organisation (HMO), despite this being shown to perform poorly by the publicly released performance data
Participants	**Inclusion criteria:** all new cases (i.e. household units) newly eligible to participate in Iowa Medicaid **Participants:** 5217 new beneficiaries with 2568 in the control group and 2649 in the intervention group **Health plans:** the Medicaid program has a form of mandatory (auto-assignment) voluntary managed care programme, which includes one or more HMOs or (sometimes) a primary care case management plan. New enrollees have an option to switch programme around the time of enrolment
Interventions	**Intervention:** standard enrolment materials and Consumer Assessment of Healthcare Providers and Systems (CAHPS) report delivered by personal mail **Control:** standard enrolment materials delivered by personal mail **Duration:** 25 March to 15 April 1998 **Deliverer:** the New Jersey Medicaid office published a 7-page brochure ("Choosing an HMO") that compared the Medicaid HMO consumer ratings and experiences reported in the CAHPS survey **Funding:** co-operative agreement 5U18HS09204-05; the Agency for Healthcare Research and Quality and the Center for Medicare and Medicare Services
Outcomes	**Main outcome** • decision to remain with the dominant HM

Notes	The star charts in CAHPS report were based on a HMO's performance compared to the average in every county of residence. The counts ranged from 20 to 29 stars. The resulting standardised CAHPS ratings ranged from -8.40 (well below the average) to 6. 26 (well above the county average) Abbreviations: cluster non-randomised trial (cNRT); Consumer Assessment of Healthcare Providers and Systems (CAHPs)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	Quote: 'Based on whether the last digit of the case ID was odd or even, half the cases were randomly assigned to an experimental group and half were assigned to a control group'
Allocation concealment (selection bias)	High risk	Allocation concealment was based on case ID number, therefore research investigators enrolling participants could possibly foresee assignment
Adequate blinding of participants, personnel and outcome assessors?	Unclear risk	Not stated whether or not participants knew that they were part of a study and so had been allocated to an intervention or control group
Incomplete outcome data (attrition bias) All outcomes	Low risk	Quote: 'The analysis of the overall effects of CAHPS included the entire April 1998 sample of enrollees, and is therefore not subject to non-response bias'
Selective reporting (reporting bias)	Low risk	All outcomes and results outlined in the Method section were reported in tables, text, or both
Baseline characteristics similar?	Unclear risk	Did not explicitly describe baseline characteristics, although attempted to take these into account when determining risk-adjusted outcomes
Baseline outcomes similar?	Low risk	Did not explicitly describe baseline outcomes, but accounted for differences appropriately using multinomial logistic regression

Farley 2002b *(Continued)*

Protection against contamination	Low risk	No specific safeguards against contamination, but reports were sent by post and so it was unlikely that the control group received the intervention
Other bias	Low risk	No additional biases identified

Flett 2015

Methods	**Design:** ITS (with non-intervention control hospitals) **Country:** USA **Care setting:** paediatric and neonatal intensive care units **Duration:** 2004 to 2012 **Dataset:** PHIS **Total participants:** 21 acute hospitals **Unit of analysis:** individual hospitals; accounted for clustering within hospitals **Data analysis:** generalised linear mixed-effects models with auto-correlated residuals
Participants	**Inclusion criteria:** children's hospitals in US states that submitted data to the PHIS **Hospitals:** 17 hospitals in 9 states that introduced public reporting of CLABSI rates, and 4 hospitals in 4 states without public reporting. Minimal data provided about the number or characteristics of individual patients treated within these hospitals
Interventions	**Intervention:** state-based mandatory public reporting of healthcare-associated infections **Duration:** public reporting introduced between July 2005 and April 2010 (depending on state) and lagged behind legislation by 6 to 27 months **Deliverer:** individual state legislatures **Funding:** unclear
Outcomes	**Main outcomes** • blood cultures per 1000 patient days • number of antibiotic days per 1000 patient days
Notes	Abbreviations: Interrupted Time Series (ITS) study; Paediatric Health Information System (PHIS); central line-associated blood stream infection (CLABSI)

Risk of bias

Bias	Authors' judgement	Support for judgement
Incomplete outcome data (attrition bias) All outcomes	Low risk	Outcome data for all included hospitals, except for one that was excluded because of excessive missing data
Selective reporting (reporting bias)	Low risk	All outcomes and results outlined in the Method section are reported in tables, text, or both

Other bias	Unclear risk	No additional biases identified
Intervention is independent of other changes? (ITS)	Low risk	Not stated whether there were other confounding events that might have changed performance over time. However, this was unlikely overall, as each state implemented mandatory reporting at different stages and using different regulatory mechanisms
Shape of intervention effect pre-specified? (ITS)	High risk	Shape of intervention effect not prespecified
Knowledge of the interventions adequately prevented during the study? (ITS)	Low risk	Individuals would not have been aware of the study as this was performed retrospectively, using a clinical registry
Intervention unlikely to affect / bias data collection? (ITS)	Low risk	Routinely collected clinical data, so data collection was unlikely to be biased by the intervention

Ikkersheim 2013

Methods	**Design:** cRT **Country:** the Netherlands (Eindhoven) **Care setting:** primary care **Duration:** 2009 to 2010 **Dataset:** prospective data collection from GPs **Total participants:** 26 GPs (2:1 randomisation to intervention) **Unit of allocation:** individual GPs **Unit of analysis:** individual GPs; accounted for clustering of GPs within practices **Sample size calculation:** not reported; statistical significance was assessed at the 0.05 level **Data analysis:** multivariable logistic regression using a difference-in-difference approach
Participants	**Inclusion criteria:** all GPs within the Eindhoven region **Participants:** 26 GPs, with 17 in the intervention group and 9 in the control group **Participant characteristics:** male 41% (intervention) versus 44% (control), urban 35% (intervention) versus 33% (control)
Interventions	**Intervention:** report cards sent by post to GPs that included a variety of quality indicators that depended on the specific condition (breast cancer, cataract surgery, hip or knee replacement) **Control:** no report cards distributed to control GPs **Duration:** no details provided **Deliverer:** research team **Funding:** the Dutch organisation for health research and development, ZonMw

Outcomes	**Main outcome** • choice of hospital when making patient referrals
Notes	Abbreviations: cluster-randomised trial (cRT); general practitioner (GP)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Method of randomisation unclear
Allocation concealment (selection bias)	Unclear risk	No statement about allocation concealment
Adequate blinding of participants, personnel and outcome assessors?	High risk	No blinding of participants or personnel; the outcomes measured GP behaviour (i.e. referral patterns); individual GPs were not blinded to the group allocation
Incomplete outcome data (attrition bias) All outcomes	Low risk	Data from all participating GPs included
Selective reporting (reporting bias)	Low risk	All outcomes and results outlined in the Method section were reported in tables, text, or both
Baseline characteristics similar?	Low risk	Some baseline characteristics described (health professional sex and urban location), which suggested that the groups were balanced
Baseline outcomes similar?	Low risk	Baseline outcomes varied between hospitals, although multivariable logistic regression was used to adjust for baseline differences
Protection against contamination	Low risk	No specific safeguards against contamination, although it was unlikely that GPs shared hospital report cards amongst themselves when they knew these were the subject of a trial
Other bias	Low risk	No additional biases identified

Jang 2011

Methods	**Design:** ITS **Country:** South Korea **Care setting:** paediatric and neonatal intensive care units **Duration:** 2003 to 2007 **Dataset:** HIRA National Quality Improvement database **Total participants:** not stated; approximately 3,000,000 live births would have been included between January 2003 and May 2007 according to data provided by Chung 2014 **Unit of analysis:** individual hospitals **Data analysis:** time series ARIMA analysis
Participants	**Inclusion criteria:** all hospitals performing 100 or more deliveries per year **Hospital types:** tertiary care hospitals (3.6%), general hospitals (13.1%), hospital (13.1%), clinic (35.4%) **Hospital regions:** capital city (4.9%), metropolis (31.7%), satellite city (22.5%), city (24.5%), rural (16.3%) **Hospital ownership:** public (3.2%), non-public (96.8%) **Hospital deliveries (per year):** > 700 (4.3%), 201 to 700 (26.4%), < 200 (69.4%)
Interventions	**Intervention:** repeated public release of information (online, press releases) on hospital caesarean rates **Duration:** four distinct interventions (September 2005, January 2006, September 2006, January 2007) **Deliverer:** HIRA, South Korea **Funding:** HIRA, South Korea
Outcomes	**Main outcome** • risk-adjusted institutional caesarean section rates
Notes	Abbreviations: Health Insurance Review & Assessment Service (HIRA); Autoregressive Integrated Moving Average (ARIMA)

Risk of bias

Bias	Authors' judgement	Support for judgement
Incomplete outcome data (attrition bias) All outcomes	Low risk	Outcome data available for all patients
Selective reporting (reporting bias)	Low risk	All outcomes and results outlined in the Method section were reported in tables, text, or both
Other bias	Low risk	No additional biases identified
Intervention is independent of other changes? (ITS)	Unclear risk	Not stated whether there were other confounding events that might have changed performance over time

Shape of intervention effect pre-specified? (ITS)	Low risk	The authors pre-specified that RPR would decrease and that cesarean section rates of institutions with higher cesarean section rates in the period before RPR would decrease further after RPR than those with lower starting rates
Knowledge of the interventions adequately prevented during the study? (ITS)	Low risk	Did not state explicitly that those responsible for data collection were informed that the publication of performance data was part of a study
Intervention unlikely to affect / bias data collection? (ITS)	Low risk	Routinely collected administrative data that were collected independently of the individuals at whom the public release of performance data were directed

Joynt 2016

Methods	**Design:** ITS **Country:** USA **Care setting:** acute hospitals **Duration:** January 2005 to November 2012 **Dataset:** Medicare inpatient files **Total participants:** 20,707,266 **Unit of analysis:** individual patients; accounted for clustering within hospitals **Data analysis:** multivariable logistic regression
Participants	**Inclusion criteria:** all Medicare fee-for-service enrollees hospitalised with any of the 15 most common non-surgical discharge diagnoses. Medicare is predominantly composed of patients aged 65 years or older. **Hospitals:** 3970 hospitals **Hospital types:** 6.8% major teaching hospital, 18.3% minor teaching hospital, 74.9% non-teaching **Hospital size:** 42.7% small, 46.3% medium, 11.0% large **Hospital ownership:** 15% for-profit, 62.8% non-profit, 22.1% public **Patients:** 20,707,266 **Patient characteristics:** mean age 79 years, 41% male
Interventions	**Intervention:** Public release of hospital performance data (using 30-day mortality), published on a publicly accessible website. The intervention was the addition of 30-day mortality to publicly accessible hospital performance data in 2008. In the pre-intervention period, hospital performance data were available in the same format, but was limited to process metrics. **Duration:** 4 years **Deliverer:** Hospital Compare, which is maintained by the CMS **Funding:** CMS

Outcomes	**Main outcome**
	• risk-adjusted 30-day mortality

Notes	Abbreviations: interrupted time series (ITS) study; Centers for Medicare & Medicaid Services (CMS)

Risk of bias

Bias	Authors' judgement	Support for judgement
Incomplete outcome data (attrition bias) All outcomes	Low risk	Data from all participating hospitals included
Selective reporting (reporting bias)	Low risk	All outcomes and results outlined in the Method section were reported in tables, text, or both
Other bias	Low risk	No additional biases identified
Intervention is independent of other changes? (ITS)	Low risk	Not stated whether there were other confounding events that might have changed performance over time. However, this was unlikely, given that this study identified few changes in outcome after the intervention
Shape of intervention effect pre-specified? (ITS)	High risk	Shape of intervention effect not prespecified
Knowledge of the interventions adequately prevented during the study? (ITS)	Low risk	Individuals would not have been aware of the study, as this was performed using routinely collected administrative data
Intervention unlikely to affect / bias data collection? (ITS)	Low risk	Routinely collected administrative data, so data collection was unlikely to be biased by the intervention

Liu 2017

Methods	**Design:** ITS (with non-intervention control hospitals) **Country:** USA **Care setting:** adult ICUs **Duration:** 2006 to 2012 **Dataset:** CDC NHSN dataset **Total participants:** 244 acute hospitals **Unit of analysis:** individual CLABSIs; accounted for clustering within hospitals **Data analysis:** multi-variable regression, using a difference-in-difference approach from hospitals in states that did not introduce mandatory reporting
Participants	**Inclusion criteria:** all non-VA acute hospitals enrolled in the NHSN were eligible to participate **Hospitals:** 244 hospitals with 475 ICUs **Hospital teaching hospital status:** control (469 ICU days, 59.1%), intervention (844, 76.2%) **Intensive care unit bed size > 30:** control (45 ICU days, 5.7%), intervention (68, 6.1%) **Number of patient days per year:** control (mean 1384.1, standard deviation (SD) 2152.0), intervention (1855.4, SD 1447.6) **Patient characteristics:** no substantial case mix data provided
Interventions	**Intervention:** mandatory public reporting of healthcare-associated infections **Duration:** variable, depending on the state being studied **Deliverer:** individual state legislatures **Funding:** unclear
Outcomes	**Main outcome** • CLABSIs per 1000 patient days
Notes	Abbreviations: interrupted time series (ITS study); intensive care unit (ICU); Centers for Disease Control and Prevention (CDC); National Healthcare Safety Network (NHSN), central line-associated blood stream infection (CLABSI); Veterans Affairs (VA)

Risk of bias

Bias	Authors' judgement	Support for judgement
Incomplete outcome data (attrition bias) All outcomes	Low risk	Outcome data for all eligible hospitals included
Selective reporting (reporting bias)	Low risk	All outcomes and results outlined in the Method section were reported in tables, text, or both
Other bias	Low risk	No additional biases identified
Intervention is independent of other changes? (ITS)	Low risk	Not stated whether there were other confounding events that might have changed performance over time. However, this was

		unlikely overall, as each state implemented mandatory reporting at different stages, and using different regulatory mechanisms
Shape of intervention effect pre-specified? (ITS)	High risk	Shape of intervention not prespecified
Knowledge of the interventions adequately prevented during the study? (ITS)	Low risk	Individuals would not have been aware of the study as this was performed retrospectively, using administrative data
Intervention unlikely to affect / bias data collection? (ITS)	Low risk	Routinely collected clinical data, so data collection was unlikely to be biased by the intervention

Rinke 2015

Methods	**Design:** CBA **Country:** USA **Care setting:** acute hospitals **Duration:** 2000 to 2009 **Dataset:** HCUP Kids' Inpatient Database **Total participants:** 4,705,857 paediatric hospital discharges **Unit of allocation:** paediatric hospital discharges **Unit of analysis:** paediatric hospital discharges; accounted for clustering of discharges within hospitals and states **Sample size calculation:** not reported; statistical significance was assessed at the 0.05 level **Data analysis:** multivariable logistic regression
Participants	**Inclusion criteria:** all paediatric hospital discharges eligible for PDI2 (i.e. length of stay 2 or more days) in a state that was categorised as 'never reporters' (18 states), '2006 reporters' (2 states), or '2009 reporters' (7 states) **Hospitals:** 3207; 2066 of which were 'never reporters', 135 were '2006 reporters', and 1006 were '2009 reporters' Hospital teaching status: never reporters (52%), 2006 reporters (55%), 2009 reporters (58%) **Participants:** 4,705,857 discharges, 2,580,621 of which were from 'never reporters', 179,322 from '2006 reporters', and 1,945,914 from '2009 reporters' **Participant age:** never reporters (mean 3.5, standard deviation (SD) 5.5), 2006 reporters (4.4, SD 6.0), 2009 reporters (3.6, SD 5.6) **Participant sex:** never reporters (male 54%, female 46%), 2006 reporters (54% male, 46% female), 2009 reporters (55% male, 45% female)
Interventions	**Intervention:** mandatory public reporting of healthcare-associated infections **Control:** no mandatory reporting of healthcare-associated infections **Duration:** mandatory CLABSI reporting introduced in 2006 or 2009 **Deliverer:** individual hospitals, as mandated by state legislatures **Funding:** unclear

Outcomes	**Main outcome**
	• paediatric safety indicator (PDI12), which was defined by the AHRQ as 'selected infections due to medical care', and determined using discharge ICD-9-CM codes 99662 (infection due to other vascular device, implant, and graft), 9993 (other infection), and 99931 (infection due to central venous catheter).
Notes	Abbreviations: controlled before-after (CBA) study; Healthcare Cost and Utilization Project (HCUP); paediatric safety indicator (PDI12); Agency for Healthcare Research and Quality (AHRQ); International Statistical Classification of Diseases 9th Revision Clinical Modification (ICD-9-CM)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	High risk	CBA study, so did not use random sequence allocation
Allocation concealment (selection bias)	High risk	No allocation concealment as hospitals would have known whether or not their state mandated public reporting
Adequate blinding of participants, personnel and outcome assessors?	High risk	No blinding of participants, personnel, or outcome assessors, as all parties would have known whether or not their state mandated public reporting
Incomplete outcome data (attrition bias) All outcomes	Low risk	Outcome data for all included hospitals
Selective reporting (reporting bias)	Low risk	All outcomes and results outlined in the Method section were reported in tables, text, or both
Baseline characteristics similar?	Low risk	Baseline characteristics differed, but were sufficiently similar to undertake the study using appropriate analyses
Baseline outcomes similar?	Low risk	Baseline outcomes differed, but were sufficiently similar to undertake the study using appropriate analyses (2.4 PDI12 per 1000 discharges in the never-reporting states, 2.6 in the 2006 reporter states, and 3.0 in the 3009 reporter states)
Protection against contamination	High risk	Unable to protect against contamination, as hospitals in states without mandatory

		reporting might have been influenced by states in which these laws were introduced
Other bias	Low risk	No additional biases identified

Romano 2004

Methods	**Design:** ITS **Country:** USA (California and New York) **Care setting:** acute hospitals **Duration:** California (1991 to 1996), New York (1989 to 1996) **Dataset:** California CPDDS, which included discharges from all non-federal hospitals in the state; New York SPARCS, which was similar in scope to the CPDDS **Total participants:** unclear **Unit of analysis:** individual patient admissions; accounted for clustering within hospitals **Data analysis:** time series ARIMA analysis
Participants	**Inclusion criteria:** adults admitted to acute non-federal hospitals in California and New York for a target condition, i.e.: California - target conditions - AMI - CABG (AMI-related) - Percutaneous coronary angioplasty (AMI-related) - Congestive heart failure (AMI-related) - Cervical discectomy - Lumbar discectomy - Back or neck procedures (discectomy-related) - Medical back problems (discectomy-related) - Knee arthroplasty (discectomy-related) - Hip arthroplasty (discectomy-related) New York - target conditions - AMI - CABG - Percutaneous coronary angioplasty (AMI-related) - Congestive heart failure (AMI-related) **Hospital characteristics:** no substantial case mix data provided **Participant characteristics:** no substantial case mix data provided
Interventions	**Intervention:** California (CHOP following legislation mandating the Office of Statewide Health Planning and Development to produce annual reports); New York (New York Cardiac Surgery Reporting System) **Duration:** California (first report published in 1993, and second in 1996); New York (hospital ratings published every 12 to 24 months, from December 1990 until the time of the study) **Deliverer:** report cards were published by agencies in California and New York **Funding:** unclear

Outcomes	**Main outcome** • Change in the utilisation decisions of consumer, healthcare professional or purchasers
Notes	Abbreviations: interrupted time series (ITS) study; California Patient Discharge Data Set (CPDDS); Statewide Planning and Research Cooperative System (SPARCS); Autoregressive Integrated Moving Average (ARIMA); acute myocardial infarction (AMI); coronary artery bypass grafting (CABG); California Hospital Outcomes Projects (CHOP)

Risk of bias

Bias	Authors' judgement	Support for judgement
Incomplete outcome data (attrition bias) All outcomes	Low risk	Outcome data for all included hospitals
Selective reporting (reporting bias)	Low risk	All outcomes and results outlined in the Method section were reported in tables, text, or both
Other bias	High risk	Main analysis based on the assumption of same trend before and after intervention; difference from predicted values was reported, rather than change in trend and level
Intervention is independent of other changes? (ITS)	Unclear risk	Not stated whether there were other confounding events that might have changed performance over time
Shape of intervention effect pre-specified? (ITS)	Low risk	Quote: 'We therefore hypothesized that hospitals with lower than expected mortality or complication rates experience significant volume increases, and hospitals with higher than expected mortality or complication rates experience significant volume decreases in the year after publication of a report card'
Knowledge of the interventions adequately prevented during the study? (ITS)	Low risk	Individuals would not have been aware of the study, as this was performed retrospectively, using administrative data
Intervention unlikely to affect / bias data collection? (ITS)	Low risk	Routinely collected administrative data that were collected independently of the individuals at whom the public release of performance data were directed

Tu 2009

Methods	**Design:** cRT **Country:** Canada (Ontario) **Care setting:** acute hospitals **Duration:** 1 April 2004 to 31 March 2005 **Dataset:** prospective chart review by research nurses, and study linkage to the Ontario Registered Persons Vital Statistics Database for mortality outcomes **Total participants:** 82 hospital organisations **Unit of allocation:** hospital organisations **Unit of analysis:** individual patients; accounted for clustering of patients within hospitals **Sample size calculation:** Quote: 'The study had 84% power to detect 5% absolute difference on the composite quality indicators. The power calculation assumed a baseline performance rate on each composite indicator of 70% (standard deviation 10%) in each study group, and that there would be a secular improvement of 75% (SD 7.5%) in the composite indicator, independent of the study intervention' **Data analysis:** multivariable logistic regression
Participants	**Inclusion criteria:** acute hospitals participating in Ontario, Canada that were identified from the Canadian Institute for Health Information hospital discharge administrative database 1999 to 2001 and treated more than 15 patients with acute myocardial infarction (AMI) annually **Participants:** 86 hospital corporations **Institution characteristics:** 12% teaching hospitals in the intervention group versus 10% in the control group; 74% community hospitals in the intervention versus 79% in the control group; 14% small hospitals in the intervention versus 10% in the control group **AMI patient characteristics:** median age 69 (interquartile range 57 to 78) both groups; female 35.4% versus 36.7% **CHF patient characteristics:** median age 77 (interquartile range 70 to 84) versus 77 (69-84); female 51.3% versus 49.2%
Interventions	**Intervention:** report cards with baseline performance data publicly released online and at a press conference **Control:** report cards publicly released after data had been collected, i.e. a delayed release of data for the control group **Duration:** January to 1 April 2004 **Deliverer:** The Canadian Cardiovascular Outcomes Research Team, which is a national team of cardiovascular outcomes researchers from across Canada **Funding:** Canadian Institutes of Health Research team grant in cardiovascular outcomes research

Outcomes	**Main outcomes**
	• Composite AMI indicators
	• Composite CHF indicators
	Secondary outcomes
	• 12 AMI process-of-care indicators
	• 6 CHF process-of-care indicators
	• 30-day and 1-year mortality for patients in the following subgroups:
	    ○ AMI
	    ○ ST-elevation myocardial infarction (STEMI)
	    ○ Non-STEMI

	○ CHF ○ CHF with left ventricular dysfunction
Notes	Abbreviations: cluster-randomised trial (cRT); acute myocardial infarction (AMI); congestive heart failure (CHF); ST-elevation MI (STEMI)

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Method of randomisation not explicitly stated, but this was undertaken by a dedicated study statistician who used a stratified randomisation process
Allocation concealment (selection bias)	Low risk	Quote 'This random assignment was stratified by type of hospital and performed by a study statistician'
Adequate blinding of participants, personnel and outcome assessors?	High risk	Quote: 'It was not possible to blind the hospitals to their status' Quote: 'We could not blind the delayed feedback group to the media coverage and associated publicity surrounding the study results' Quote: 'Patient charts were abstracted by an experienced research nurse', but it is unclear whether or not the nurse was blinded to allocation
Incomplete outcome data (attrition bias) All outcomes	High risk	One hospital withdrew from the baseline phase after randomisation, and 4 withdrew from the follow-up phase, all due to resource constraints. No intention-to-treat analysis was performed. Additional exclusions of patients were not reported
Selective reporting (reporting bias)	Low risk	A protocol was registered in advance of randomisation and all outcomes were reported in the final report, which also included a new outcome (all-cause mortality)
Baseline characteristics similar?	Low risk	Baseline characteristics of patients and hospitals between the intervention and control groups were similar
Baseline outcomes similar?	Low risk	Baseline outcomes presented and varied between hospitals, although results were presented as absolute change, and so ac-

		counted for baseline differences
Protection against contamination	High risk	Quote: 'There was extensive media coverage following the release of the baseline performance for the early feedback hospitals' Quote: 'One unanticipated observation was that several hospitals in the delayed feedback group reported that they also initiated some quality improvement activities after becoming aware of the publicly released early feedback report cards, before receiving their own hospital- specific results'
Other bias	Low risk	No additional biases identified

Zhang 2016

Methods	**Design:** cRT **Country:** China (Hubei Province) **Care setting:** primary healthcare institutions **Duration:** 2013 to 2014 **Dataset:** Data collected from patient electronic health records **Total participants:** 748,632 outpatient prescriptions from 20 primary healthcare institutions **Unit of allocation:** primary healthcare institutions (paired and matched for similar characteristics) **Unit of analysis:** individual prescriptions; accounted for clustering of prescriptions by individual prescribers **Sample size calculation:** not reported; statistical significance was assessed at the 0.05 level **Data analysis:** multivariable regression models, using a difference-in-difference approach
Participants	**Inclusion criteria:** primary care institutions selected from within Qian Jiang City **Primary healthcare institutions:** 20 providers, 10 of which were in the intervention group, and 10 in the control group **Institution characteristics:** 60 beds in the intervention group versus 66 in the control group; 28 versus 26 doctors, 50,199 versus 49,108 annual outpatient visits **Patient characteristics:** mean age 37.5 years, 49.5% male
Interventions	**Intervention:** public display of prescription information (percentage of prescriptions requiring antibiotics, percentage requiring injections, and average patient expenditure) on outpatient department bulletin boards in participating institutions **Control:** no public display of prescription information **Duration:** 1 October 2013 to 31 August 2014 **Deliverer:** outpatient departments of participating institutions **Funding:** National Natural Science Foundation of China

Outcomes	**Main outcomes**
	• Percentage of prescriptions requiring antibiotics
	• Percentage of prescriptions requiring combined antibiotics
	• Percentage of prescriptions requiring injections
	• Average expenditure per prescription
Notes	Zhang 2016 represents a single study that was reported in five articles (Wang 2014; Yang 2014; Liu 2015; Liu 2016; Tang 2016) that individually satisfied our inclusion criteria. However, the senior author confirmed that these represented multiple analyses of a single cluster-RT (Zhang 2018 [pers comm]). Therefore, we made the decision to present the cluster-RT (as the original study design and higher level of evidence), rather than the designs (e.g. CBA and ITS) that were presented in the other articles
	Abbreviations: cluster-randomised trial (cRT); controlled before-after (CBA) study; interrupted time series (ITS) study

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: 'We flipped a coin to randomly assign one (primary care institution) into the intervention group and another into the control group'
Allocation concealment (selection bias)	High risk	Healthcare providers could not be blinded to the allocation
Adequate blinding of participants, personnel and outcome assessors?	High risk	It was not possible to blind personnel, who must have been aware of the group to which their primary care institution had been allocated
Incomplete outcome data (attrition bias) All outcomes	Low risk	Injection prescribing data retrieved from a comprehensive administrative database
Selective reporting (reporting bias)	Low risk	All outcomes and results outlined in the Method section were reported in tables, text, or both. Although a protocol for the cRT was published, this appeared eighteen months after the trial reports stated that the intervention began
Baseline characteristics similar?	Low risk	Some baseline characteristics described (e. g. age and sex), which suggested that the groups were balanced
Baseline outcomes similar?	Low risk	Baseline outcomes presented and varied between hospitals. However, the hospitals

		were paired according to characteristics, and the results analysed using a difference-in-difference approach and regression models to account for residual baseline differences
Protection against contamination	Unclear risk	No statement as to whether or not other events might have influenced performance over time
Other bias	Low risk	No additional biases identified

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Cavender 2015	Cross-sectional study comparing outcomes between states with and without mandatory public reporting
Moscucci 2005	Study design, controlled before-after design; no information reported from the 2 included registries. Not enough information was reported regarding the baseline data
Paris 2013	Data reported on a private website, so they were not made available to the public
Park 2011	Interrupted time series with insufficient data points
Saratzis 2017	Interrupted time series with insufficient data points

A D D I T I O N A L   T A B L E S

Table 1. Summary of included studies

Study details[a]			Improvement by selection			Improvement by changes in care			Data available
Study	Design, setting, and participants	Intervention	Consumers	Providers	Purchasers	Provider performance	Patient outcome	Staff morale	
Farley 2002a	cRT; USA; 13,077 insurance plan bene-	Consumer Assessment of Health-	X	-	-	-	-	-	X

Table 1. **Summary of included studies** (*Continued*)

	ficiaries	care Providers and Systems (CAHPS) report							
Farley 2002b	cNRT; USA; 5217 insurance plan beneficiaries	Consumer Assessment of Healthcare Providers and Systems (CAHPS) report	X	-	-	-	-	-	X
Romano 2004	ITS; USA; -	Report cards with risk-adjusted patient outcomes produced by state agencies	X	-	-	-	-	-	b
Flett 2015	ITS; USA; 21 hospitals	State-based mandatory public reporting of healthcare-associated infections	-	X	-	-	-	-	X
Rinke 2015	CBA; USA; 3207 hospitals	Mandatory public reporting of healthcare-associated infections	-	-	-	-	X	-	X
DeVore 2016	ITS; USA; 315,092 hospitalisations	Online reporting of risk-adjusted 30-day re-ad-	-	-	-	-	X	-	b

Table 1. **Summary of included studies** (*Continued*)

Study			C1	C2	C3	C4	C5	C6	C7
		mission rates (Hospital Compare)							
Joynt 2016	ITS; USA; 3970 hospitals	Online reporting of risk-adjusted 30-day mortality rates (Hospital Compare)	-	-	-	-	X	-	X
Liu 2017	ITS; USA; 244 hospitals	Mandatory public reporting of healthcare-associated infections	-	-	-	-	X	-	-[c]
Tu 2009	cRT; Canada; 82 hospital organisations	Report cards with risk-adjusted patient outcomes and a press conference	-	-	-	X	X	-	X
Jang 2011	ITS; South Korea; 3,000,000 live births	Repeated public release of information (online, press releases) on hospital caesarean rates	-	X	-	-	-	-	X
Ikkersheim 2013	cRT; The Netherlands; 26 general practitioners	Report cards with risk-adjusted patient outcomes sent	-	X	-	-	-	-	[b]

Table 1. Summary of included studies *(Continued)*

		to GPs for discussion with patients							
Zhang 2016	cRT; China; 20 primary care providers	Public display of prescription information on outpatient department bulletin boards	-	X	-	-	-	-	X

controlled before-after (CBA) study; cluster-randomised trial (cRT); cluster-non-randomised trial (cNRT); Consumer Assessment of Healthcare Providers and Systems (CAHPs); general practitioners (GPs); interrupted time series (ITS) study

Column headers: changes in healthcare utilisation by consumers (Consumers); changes in healthcare decisions taken by healthcare providers (professionals and organisations; (Providers)); changes in healthcare decisions of purchasers (Purchasers); changes in provider performance (Provider performance); changes in patient outcome (Patient outcome); changes in staff morale (Staff morale); impact on equity (Equity)

Order of studies: listed in chronological order USA, then chronological order for other countries of study

[a] Studies grouped by intervention, i.e. mode of public release of performance data

[b] No change in slope and so re-analysis of the ITS data was uninformative

[c] Presented derived data (e.g. outputs of regression models) that were insufficient for re-analysis

Table 2. Changes in the healthcare utilisation decisions of consumers

Intervention	Outcome	Study	Type of study	Absolute post-intervention difference	Absolute pre-intervention difference	Post-intervention level in control group	Relative effect
Dissemination of consumer reports directly to consumers	Assigned to high-rated HMO (2 choices)	Farley 2002a	cRT	1.5	0	15.9	0.0943
	Assigned to low-rated HMO (2 options)			0.4	0	25	0.0160
	Assigned to high-rated HMO (1 option)			1.3	0	29.5	0.0441

As-signed to low-rated HMO (1 option)			0.1	0	23.7	0.0042
Proportion choosing plan	Farley 2002b	cNRT	0.01	0	0.69	0.0145

cluster-randomised trial (cRT); cluster-non-randomised trial (cNRT); health maintenance organization (HMO)

Table 3. Changes in the healthcare utilisation decisions of healthcare providers (professionals and organisations)

Interven-tion	Outcome	Study	Type of study	Absolute post-in-tervention difference	Absolute pre-inter-vention difference	Post-interven-tion level in control	Relative effect
Public re-porting of injection prescrib-ing rates in outpatient areas	Average ex-penditure per pre-scription	Zhang 2016	cRT	3.4	2.2	41.2	0.0291
	Percent-age of pre-scriptions requiring antibiotics			4.6	6.1	62.8	-0.0249
	Percent-age of pre-scriptions requiring combined antibiotics			2.1	4.1	18.6	-0.1083
	Percent-age of pre-scriptions requiring injections			9.0	13.2	64.9	-0.0643
	Average ex-penditure per pre-scription			7.2	6.9	44.3	0.0070

Table 3. Changes in the healthcare utilisation decisions of healthcare providers (professionals and organisations) *(Continued)*

Intervention	Outcome	Study	Type of study	Absolute level effect (95% CI)				
Mandatory public reporting of healthcare-associated infections	Pediatric quality indicator per 1000 eligible discharges	Rinke 2015	CBA	0.6	0.5	1.0	0.1000	

Intervention	Outcome	Study	Type of study	Absolute level effect (95% CI)	Relative change at 3 months (95% CI)	Relative change at 6 months (95% CI)	Relative change at 9 months (95% CI)	Relative change at 12 months (95% CI)	Relative change at 24 months (95% CI)
Repeated public release of hospital caesarean section rates	Caesarean section rate	Jang 2011	ITS	-0.52 (-0.77 to -0.26)	-0.04 (-1.23 to 1.18)	-1.49 (-2.55 to -0.40)	-2.92 (-4.50 to 1.30)	-4.34 (-6.61 to -1.95)	-
Mandatory public reporting of healthcare-associated infections	PICU blood cultures	Flett 2015	ITS	7.48 (1.09 to 13.87)	6.21 (-2.84 to 17.10)	9.90 (-0.45 to 22.64)	13.87 (1.42 to 29.82)	18.17 (2.90 to 38.77)	22.87 (4.11 to 49.86)
	PICU antibiotics			7.29 (4.46 to 10.12)	-0.11 (-2.03 to 1.89)	1.61 (-0.45 to 3.75)	3.36 (0.96 to 5.87)	5.15 (2.26 to 8.20)	6.98 (2.50 to 10.70)
	NICU antibiotics			-5.79 (-9.17 to -2.42)	8.12 (4.11-12.46)	6.06 (2.08 to 10.35)	4.05 (-0.35 to 8.85)	1.90 (-3.17 to 7.53)	-0.36 (-6.25 to 6.33)
	NICU blood cultures			-1.14 (-1.90 to -0.39)	2.49 (-0.51 to 5.67)	1.06 (-2.07 to 4.39)	-0.42 (-3.93 to 3.36)	-1.95 (-6.02 to 2.49)	-3.53 (-8.26 to 1.72)

cluster-randomised trial (cRT); controlled before-after (CBA) study; 95% confidence interval (95% CI); interrupted time series (ITS) study; neonatal intensive care unit (NICU); paediatric intensive care unit (PICU)

Table 4. Changes in provider performance

Intervention	Outcome	Study	Type of study	Absolute post-intervention difference	Absolute pre-intervention difference	Postintervention level in control group	Relative effect

Table 4. Changes in provider performance *(Continued)*

Public release of a range of quality indicators	All AMI processes	Tu 2009	cRT	2.0	0.9	65.6	0.0168
	Use of standard admission orders			6.1	0.7	72.5	0.0745
	Left ventricular function assessment			2.9	6.3	49.8	-0.0683
	Lipid test < 24 hours arrival			3.8	1.6	51.1	0.0431
	Fibrinolytics < 30 mins after arrival			2.6	3.1	45.7	-0.0109
	Fibrinolytics decided by ED physician			2.0	4.4	84.3	-0.0285
	Fibrinolytics prior to transfer to CCU			3.8	2.9	95.7	0.0094
	Aspirin < 6 hours arrival			5.5	3.1	82.6	0.0291
	B blockers < 12 hours arrival			2.4	3.9	73.7	-0.0204
	Aspirin at discharge			0.9	0.0	84.0	0.0107
	B blockers at discharge			0.6	0.0	85.6	0.0070
	ACEi, ARB for LV dysfunction			4.7	3.4	81.7	0.0159
	Statin at discharge			0.3	0.2	85.5	0.0012
	All CHF processes			1.0	3.0	54.6	-0.0366

Table 4. Changes in provider performance *(Continued)*

			LVF assessment	2.7	4.5	55.2	-0.0326
			Daily weights recorded	1.3	0.3	24.0	0.0417
			Counselling on > 1 aspect of CHF	0.9	1.7	55.3	-0.0145
			ACEi, ARB for LV dysfunction	6.3	1.7	92.4	0.0498
			B blockers for LV dysfunction	4.0	1.7	71.7	0.0321
			Warfarin for AF	0.6	3.1	64.2	-0.0389

atrial fibrillation (AF); acute myocardial infarction (AMI); angiotensin-converting enzyme inhibitor (ACEi); angiotensin-2 receptor blockers (ARB); beta-adrenergic blocking agents (B blockers); cluster-randomised trial (cRT); coronary care unit (CCU); congestive heart failure (CHF); emergency department (ED); left ventricular (LV); left ventricular failure (LVF); minutes (mins)

Table 5. Changes in patient outcome

Intervention	Outcome	Study	Type of study	Absolute postintervention difference	Absolute pre-intervention difference	Postintervention level in control group	Relative effect	
Public release of a range of quality indicators	AMI 30-day mortality	Tu 2009	cRT	2.4	0.5	9.8	0.1939	
	AMI 1-year mortality			3.1	1	19.4	0.1082	
	STEMI 30-day mortality			3.1	0.4	8.3	0.3253	
	STEMI 1-year mortality			3.9	1.2	13.5	0.2000	

Table 5. Changes in patient outcome   (*Continued*)

Intervention	Outcome	Study*	Type of study	Absolute level effect (95% CI)	Relative change at 4 months (95% CI)	Relative change at 8 months (95% CI)	Relative change at 12 months (95% CI)	Relative change at 24 months (95% CI)
	NSTEMI 30-day mortality			2.3	0.3	10.5	0.1905	
	NSTEMI 1-year mortality			3	0.9	22.6	0.0929	
	CHF 30-day mortality			1	0.9	9.6	0.0104	
	CHF 1-year mortality			2.6	0.6	30.3	0.0660	
	CHF and LV dysfunction 30-day mortality			0.9	0.6	8.5	0.0353	
	CHF and LV dysfunction 1-year mortality			6.3	1.8	26.3	0.1711	
Mandatory reporting of healthcare-associated infections	Pediatric quality indicator per 1000 eligible discharges	Rinke 2015	CBA	0.6	0.5	1	0.1000	
Hospital quality process and outcome metrics reported on a public website	30-day risk-adjusted mortality	Joynt 2016	ITS	0.12 (0.03 to 0.21)	1.57 (-4.28 to 8.18)	-2.47 (-8.20 to 4.03)	3.71 (-3.25 to 11.74)	7.18 (-1.91 to 18.13)
Public reporting	30-day readmission	DeVore 2016	ITS	0.00 (0.00 to 0.00)	-2.04 (-8.56 to 5.48)	-1.36 (-7.92 to 6.20)	-0.69 (-7.34 to 7.00)	0.72 (-6.32 to 8.90)

Table 5. Changes in patient outcome (*Continued*)

of risk-standardised hospital re-admission rates	(AMI)							
	30-day re-admission (heart failure)			0.00 (0.00 to 0.00)	-1.39 (-4.17 to 1.56)	-1.84 (-4.59 to 1.08)	-1.88 (-4.68 to 1.10)	-2.78 (-6.42 to 1.15)
	30-day re-admission (pneumonia)			0.00 (0.00 to 0.00)	-4.44 (-13.61 to 6.91)	-5.07 (-14.17 to 6.20)	-5.69 (-14.71 to 5.47)	-7.45 (-18.10 to 6.37)
	30-day re-admission (COPD)			0.00 (0.00 to 0.00)	-6.66 (-11.42 to -1.37)	-0.76 (-6.11 to 5.23)	-7.64 (-12.31 to -2.44)	-9.06 (-13.62 to -4.00)
	30-day re-admission (diabetes)			0.00 (-0.00 to 0.01)	-0.65 (-13.66 to 16.96)	0.00 (-13.13 to 17.81)	0.65 (-12.44 to 18.35)	1.98 (-13.57 to 24.36)
	30-day mortality (AMI)			0.00 (0.00 to 0.00)	34.38 (2.71 to 94.32)	35.83 (2.79 to 100.17)	37.38 (2.88 to 106.67)	43.06 (3.20 to 133.08)
	30-day mortality (heart failure)			0.00 (0.00 to 0.00)	6.04 (-5.86 to 21.37)	13.78 (-0.56 to 32.94)	9.98 (-3.46 to 27.77)	13.31 (-0.54 to 31.64)
	30-day mortality (pneumonia)			0.00 (0.00 to 0.00)	-3.96 (-23.10 to 27.85)	-3.72 (-16.70 to 14.05)	2.94 (-18.04 to 19.00)	-3.84 (-22.51 to 26.69)
	30-day mortality (COPD)			0.00 (0.00 to 0.00)	20.89 (5.51 to 41.52)	21.63 (5.68 to 43.24)	20.99 (5.54 to 41.75)	22.00 (5.77 to 44.13)
	30-day mortality (diabetes)			0.00 (0.00 to 0.00)	-14.73 (-34.83 to 23.29)	-15.10 (-35.48 to 24.12)	-14.78 (-34.92 to 23.40)	-19.39 (-42.65 to 35.66)

Acute Myocardial Infarction (AMI); ST-Elevation Myocardial Infarction (STEMI); Non-ST-Elevation Myocardial Infarction (NSTEMI); Congestive Heart Failure (CHF); Left Ventricular (LV); Chronic Obstructive Pulmonary Disease (COPD); Cluster Randomised Trial (cRT); Controlled Before-After (CBA) study; Interrupted Time Series (ITS) study; 95% Confidence Interval (95% CI)

* Joynt 2016 and DeVore 2016 provided outcomes in quarters rather than months and so have been presented as 4- and 8-months rather than the pre-specified 3- and 6-months.

APPENDICES

Appendix 1. Search strategies

The Cochrane Library

No.	Search terms	Results
#1	[mh "quality indicators, health care"]	430
#2	(performance next outcome):ti,ab	27
#3	(quality near/2 indicator*)	774
#4	(quality next (criteria or criterion or standard* or norm*))	7150
#5	(performance next (indicator* or measure* or data or rating* or information))	1580
#6	{or #1-#5}	9538
#7	[mh "patient satisfaction"]	11006
#8	[mh "consumer behavior"]	704
#9	[mh "consumer participation"]	1328
#10	[mh "patient acceptance of health care"]	25536
#11	[mh "decision making"]	3661
#12	[mh "choice behavior"]	1181
#13	(patient next (satisfaction or preference*)):ti,ab	7724
#14	(consumer next report*):ti,ab	3
#15	(decision next making):ti,ab	5100
#16	(choice next behavio*):ti,ab	54
#17	(provider next profiling):ti,ab	0
#18	{or #7-#17}	38104
#19	#6 and #18	531

#20	("Consumer Assessment of Healthcare Providers and Systems" or CAHPS):ti,ab	44
#21	(public next (disclosure or release)):ti,ab	26
#22	((public or publically or publicly) near/3 (report* or report-card*)):ti,ab	168
#23	((public or publically or publicly or publication* or publish* or release) near/3 quality near/2 (information or data or report* or criteria or criterion or standard* or norm* or indicator*)):ti,ab	58
#24	((public or publically or publicly or publication* or publish* or release) near/3 (performance or hospital) next (data or indicator* or measure* or rating* or information or outcome)):ti,ab	8
#25	((publication* or publish*) near/3 (report card* or reportcard*)):ti,ab	0
#26	((public or publically or publicly or release) near/3 (waiting time* or waiting list* or outcome* or mortality or certification or accreditation)):ti,ab	154
#27	{or #19-#26}	962

MEDLINE OVID

including Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Versions

No.	Search terms	Results
1	quality indicators, health care/	13425
2	performance outcome.ti,ab.	312
3	(quality adj2 indicator?).tw.	7698
4	(quality adj (criteria or criterion or standard? or norm*)).tw	8898
5	(performance adj (indicator? or measure? or data or rating? or information)).tw	13724
6	or/1-5	39819

7	exp patient satisfaction/	77553
8	consumer behavior/	19767
9	exp consumer participation/	37908
10	exp "patient acceptance of health care"/	211344
11	decision making/	82537
12	choice behavior/	27969
13	(patient adj (satisfaction or preference?)).ti,ab.	34638
14	consumer report*.ti,ab.	171
15	decision making.ti,ab.	99376
16	choice behavio?r.ti,ab.	1225
17	provider profiling.ti,ab.	68
18	or/7-17	420568
19	6 and 18	3693
20	("Consumer Assessment of Healthcare Providers and Systems" or CAHPS).ti,ab	536
21	(public adj (disclosure or release)).ti,ab.	465
22	((public or publicly or publicly) adj3 (report* or reportcard*)).ti,ab	7360
23	((public or publically or publicly or publication* or publish* or release) adj3 quality adj2 (information or data or report* or criteria or criterion or standard? or norm* or indicator?)).ti,ab	727
24	((public or publically or publicly or publication* or publish* or release) adj3 (performance or hospital) adj (data or indicator? or measure? or rating? or information or outcome)).ti,ab	210
25	((publication* or publish*) adj3 (report card* or reportcard*)).ti,ab	20
26	((public or publically or publicly or release) adj3 (waiting time* or waiting list* or outcome* or mortality or certification or accreditation)).ti,ab	2092

27	or/19-26	14336
28	intervention?.ti. or (intervention? adj6 (clinician? or collaborat* or community or complex or design* or doctor? or educational or family doctor? or family physician? or family practitioner? or financial or GP or general practice? or hospital? or impact? or improv* or individuali?e? or individuali?ing or interdisciplin* or multicomponent or multi-component or multidisciplin* or multi-disciplin* or multifacet* or multi-facet* or multimodal* or multi-modal* or personali?e? or personali?ing or pharmacies or pharmacist? or pharmacy or physician? or practitioner? or prescrib* or prescription? or primary care or professional* or provider? or regulatory or tailor* or target* or team* or usual care)).ab	242637
29	(pre-intervention? or preintervention? or "pre intervention?" or post-intervention? or postintervention? or "post intervention?").ti,ab	17765
30	(hospital* or patient?).hw. and (study or studies or care or health* or practitioner? or provider? or physician? or nurse? or nursing or doctor?).ti,hw	860541
31	demonstration project?.ti,ab.	2366
32	(pre-post or "pre test*" or pretest* or posttest* or "post test*" or (pre adj5 post)).ti,ab	96293
33	(pre-workshop or post-workshop or (before adj3 workshop) or (after adj3 workshop)).ti,ab	929
34	trial.ti. or ((study adj3 aim?) or "our study").ab.	938722
35	(before adj10 (after or during)).ti,ab.	441395
36	("quasi-experiment*" or quasiexperiment* or "quasi random*" or quasirandom* or "quasi control*" or quasicontrol* or ((quasi* or experimental) adj3 (method* or study or trial or design*))).ti,ab	131481
37	non-randomized controlled trials as topic/	169
38	pilot projects/	104980
39	pilot.ti. or (pilot adj (project? or study or trial)).ab.	93665

40	(time points adj3 (over or multiple or three or four or five or six or seven or eight or nine or ten or eleven or twelve or month* or hour? or day? or "more than")).ab	14152
41	("time series" adj2 interrupt*).ti,ab.	1859
42	interrupted time series analysis/	298
43	controlled before-after studies/	256
44	historically controlled study/	121
45	(multicentre or multicenter or multi-centre or multi-center).ti	42227
46	(control adj3 (area or cohort? or compare? or condition or design or group? or intervention? or participant? or study)).ab	626900
47	random*.ti,ab. or controlled.ti.	1014934
48	(control year? or experimental year? or (control period? or experimental period?)).ti,ab	15687
49	(utili?ation or programme or programmes).ti.	68459
50	(during adj5 period).ti,ab.	365550
51	((strategy or strategies) adj2 (improv* or education*)).ti,ab	28012
52	(clinical trial or multicenter study).pt.	698767
53	evaluation studies as topic/ or prospective studies/ or retrospective studies/	1219000
54	((evaluation or prospective or retrospective) adj study).ti,ab	237086
55	or/28-54	5017158
56	"comment on".cm. or review.pt. or (review not "peer review*").ti. or randomized controlled trial.pt	3590916
57	(rat or rats or cow or cows or chicken? or horse or horses or mice or mouse or bovine or animal?).ti,hw. or veterinar*.ti,ab, hw	6283639
58	exp animals/ not humans.sh.	4421689
59	or/56-58	9133273

60	55 not 59	3575771
61	exp randomized controlled trial/	467146
62	controlled clinical trial.pt.	94240
63	randomi#ed.ti,ab.	523398
64	placebo.ab.	190671
65	drug therapy.fs.	2009566
66	randomly.ti,ab.	284535
67	trial.ab.	428718
68	groups.ab.	1746590
69	or/61-68	4163334
70	Clinical Trials as topic.sh.	186965
71	trial.ti.	183461
72	or/61-64,66,70-71	1182798
73	exp animals/ not humans/	4421689
74	72 not 73	1092643
75	60 or 74	4276001
76	27 and 75	5296

Embase OVID

1974 to 2017 June 23

No.	Search terms	Results
1	*health care quality/	68159
2	performance outcome.ti,ab.	342
3	(quality adj2 indicator?).tw.	10919

(Continued)

4	(quality adj (criteria or criterion or standard? or norm*)).tw	12871
5	(performance adj (indicator? or measure? or data or rating? or information)).tw	17408
6	or/1-5	104870
7	*patient satisfaction/	19552
8	*consumer attitude/	1238
9	*consumer/	13586
10	exp *patient attitude/	77273
11	exp *decision making/	64896
12	(patient adj (satisfaction or preference?)).ti,ab.	47403
13	consumer report*.ti,ab.	224
14	decision making.ti,ab.	125401
15	choice behavio?r.ti,ab.	1259
16	provider profiling.ti,ab.	78
17	or/7-16	289262
18	6 and 17	6609
19	("Consumer Assessment of Healthcare Providers and Systems" or CAHPS).ti,ab	688
20	(public adj (disclosure or release)).ti,ab.	539
21	((public or publically or publicly) adj3 (report* or reportcard*)).ti,ab	4867
22	((public or publically or publicly or publication* or publish* or release) adj3 quality adj2 (information or data or report* or criteria or criterion or standard? or norm* or indicator?)).ti,ab	925
23	((public or publically or publicly or publication* or publish* or release) adj3 (performance or hospital) adj (data or indicator? or measure? or rating? or information or outcome)).ti,ab	260

24	((publication* or publish*) adj3 (report card* or reportcard*)).ti,ab	25
25	((public or publically or publicly or release) adj3 (waiting time* or waiting list* or outcome* or mortality or certification or accreditation)).ti,ab	2577
26	or/18-25	15596
27	intervention?.ti. or (intervention? adj6 (clinician? or collaborat* or community or complex or design* or doctor? or educational or family doctor? or family physician? or family practitioner? or financial or GP or general practice? or hospital? or impact? or improv* or individuali?e? or individuali?ing or interdisciplin* or multicomponent or multi-component or multidisciplin* or multi-disciplin* or multifacet* or multi-facet* or multimodal* or multi-modal* or personali?e? or personali?ing or pharmacies or pharmacist? or pharmacy or physician? or practitioner? or prescrib* or prescription? or primary care or professional* or provider? or regulatory or tailor* or target* or team* or usual care)).ab	311084
28	(pre-intervention? or preintervention? or "pre intervention?" or post-intervention? or postintervention? or "post intervention?").ti,ab	23998
29	(hospital* or patient?).hw. and (study or studies or care or health* or practitioner? or provider? or physician? or nurse? or nursing or doctor?).ti,hw	2331637
30	demonstration project?.ti,ab.	2749
31	(pre-post or "pre test*" or pretest* or posttest* or "post test*" or (pre adj5 post)).ti,ab	150055
32	(pre-workshop or post-workshop or (before adj3 workshop) or (after adj3 workshop)).ti,ab	1380
33	trial.ti. or ((study adj3 aim?) or "our study").ab.	1333514
34	(before adj10 (after or during)).ti,ab.	576316
35	("quasi-experiment*" or quasiexperiment* or "quasi random*" or quasirandom* or "quasi control*" or quasicontrol* or ((quasi* or experimental) adj3 (method* or study or trial or design*))).ti,ab	148541
36	quasi experimental study/	3847

37	*experimental design/ or *pilot study/	12503
38	pilot.ti. or (pilot adj (project? or study or trial)).ab.	127487
39	(time points adj3 (over or multiple or three or four or five or six or seven or eight or nine or ten or eleven or twelve or month* or hour? or day? or "more than")).ab	20189
40	("time series" adj2 interrupt*).ti,ab.	2195
41	(multicentre or multicenter or multi-centre or multi-center).ti	60391
42	(control adj3 (area or cohort? or compare? or condition or design or group? or intervention? or participant? or study)).ab	838775
43	random*.ti,ab. or controlled.ti.	1272455
44	(control year? or experimental year? or (control period? or experimental period?)).ti,ab	18126
45	(utili?ation or programme or programmes).ti.	84680
46	(during adj5 period).ti,ab.	479514
47	((strategy or strategies) adj2 (improv* or education*)).ti,ab	35217
48	*clinical trial/ or *multicenter study/	22311
49	*evaluation study/ or *prospective study/ or *retrospective study/	17721
50	((evaluation or prospective or retrospective) adj study).ti,ab	334438
51	or/27-50	6068150
52	(rat or rats or cow or cows or chicken? or horse or horses or mice or mouse or bovine or animal?).ti	1706244
53	(exp animal/ or exp invertebrate/ or animal experiment/ or animal model/ or animal tissue/ or animal cell/ or nonhuman/ or exp eperimental animal/) not (human/ or normal human/ or human cell/)	6203744
54	or/52-53	6406291
55	51 not 54	5294553

56	random*.ti,ab.	1208834
57	factorial*.ti,ab.	30632
58	(crossover* or cross over*).ti,ab.	88860
59	((doubl* or singl*) adj blind*).ti,ab.	197702
60	(assign* or allocat* or volunteer* or placebo*).ti,ab.	850967
61	crossover procedure/	52055
62	single blind procedure/	27932
63	randomized controlled trial/	458261
64	double blind procedure/	140193
65	or/56-64	1880362
66	exp animal/ not human/	4805839
67	65 not 66	1685252
68	55 or 67	5737909
69	26 and 68	9702

ClinicalTrials.gov

Search terms	Results
performance indicator AND public AND behaviour	6
quality indicator AND public AND behaviour	21
performance data AND public AND behaviour	31
performance measure AND public AND behaviour	52
performance rating AND public AND behaviour	36
performance information AND public AND behaviour	37

Search terms	Results
quality AND public AND behaviour	15
performance AND public AND behaviour	5

Appendix 2. Certainty of the evidence

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Certainty (overall score)
Outcome: Changes in healthcare utilisation by consumers							
3	1 RT 1 NRT 1 ITS	0	0	0	0	0	**Low**
Studies: Farley 2002a, Farley 2002b, Romano 2004	No cause to increase or decrease level of confidence.						
Outcome: Changes in healthcare decisions taken by healthcare providers (professionals and organisations)							
4	2 RT 2 ITS	Initial: 3 Final: 3	-1*	0	0	0	**Low**
Studies: Flett 2015; Ikkersheim 2013; Jang 2011; Zhang 2016	* -1 for inconsistency as Zhang 2016 showed a change in behaviour, which was not consistently observed throughout the other 3 studies						
Outcome: Changes in healthcare decisions taken by healthcare purchasers							
0	0	0	0	0	0	0	-
Studies: None							
Outcome: Changes in provider performance							
1	1 RT	-2*	0	0	0	0	**Low**
Studies: Tu 2009	* -2 for risk of bias as there was attrition of participating hospitals, evidence of contamination of the intervention across intervention and control hospitals, and blinding was not possible given the nature of the intervention						

Outcome: Changes in patient outcome							
5	1 RT 3 ITS 1 CBA	0	-2*	0	0	0	**Low**
Studies: DeVore 2016; Joynt 2016; Liu 2017; Rinke 2015; Tu 2009	* -2 for inconsistency as there was marked disagreement between studies with two showing improvements in patient outcome (Tu 2009, Liu 2017) and three showing no such improvements (DeVore 2016, Joynt 2016, Rinke 2015).						
Outcome: Changes in staff morale							
0	0	0	0	0	0	0	-
Studies: None							
Outcome: Adverse effects							
0	0	0	0	0	0	0	-
Studies: None							
Outcome: Impact on equity							
1	1 ITS	0	0	0	0	0	**Low**
Studies: Romano 2004	No cause to increase or decrease level of confidence.						

WHAT'S NEW

Last assessed as up-to-date: 26 June 2017.

Date	Event	Description
3 April 2018	New search has been performed	This is the first update of this review. We updated the searches to June 2017, and identified 8 new studies. The review now includes 12 studies. The 'Risk of bias' assessments, data extraction, and data synthesis were undertaken for all 12 studies, to bring the review into line with the latest Effective Practice and Organisation of Care (EPOC) guidelines. Four authors (Marjan Faber, Liv Rygh, Katherine Deane, and Martin Eccles) left the original co-author team, and were replaced by five others (David Metcalfe, Arturo Rios Diaz, Olubode Olufajo, Sofia Massa, and Daniel Perry)

(Continued)

29 March 2018	New citation required but conclusions have not changed	The conclusion that the evidence base is inadequate to directly inform policy and practice has not changed since the last version of this review

HISTORY

Protocol first published: Issue 4, 2003

Review first published: Issue 11, 2011

Date	Event	Description
21 August 2008	New citation required and minor changes	Comments on protocol.
4 April 2008	Amended	Converted to new review format.
12 August 2003	New citation required and major changes	Substantive amendment.

CONTRIBUTIONS OF AUTHORS

DM coordinated this update and drafted the review, with guidance from DP. NK, and SF were co-authors on an earlier version, and so helped develop the study protocol. DM, ARD, and OO undertook study selection for the updated review, extracted and re-analysed study data, and performed the 'Risk of bias' assessments. DM and SM undertook the statistical analysis. All authors made critical revisions to the manuscript.

DECLARATIONS OF INTEREST

DM: None known

NK: None known

ARD: None known

OO: None known

SF: None known

SM: None known

DP: None known

SOURCES OF SUPPORT

Internal sources

- University of Newcastle upon Tyne, UK.

External sources

- UK National Institute for Health Research Cochrane Programme Grant, UK.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

In the protocol, we listed types of participants as healthcare providers, which included hospitals, practices, and individual health professionals. Patients and other healthcare consumers and purchasers of health care were also target groups for the aims and scopes of performance measurements. We added these types to the list of participants, so it should be mentioned here, but should not be considered a change of protocol. We mentioned patients, other healthcare consumers, and purchasers in the protocol description of outcome measures, but they were missing in the types of participants. We solved this inconsistency in the review by adding these types of participants. We did not present funnel plots as there were not more than 10 studies contributing to any given analysis, as recommended by Higgins 2011. We had planned to report the findings of included studies that also included data on awareness, knowledge of performance data, attitude, and cost data. However, there were substantial difficulties in presenting such data from two studies due to inadequate survey response (Farley 2002b) and poor reporting (Ikkersheim 2013) and so these data were not presented.

Since the publication of the protocol, the Effective Practice and Organisation of Care (EPOC) Group has adjusted the definitions for the quality criteria. In the review, we used the latest version of the 'Risk of bias' tables to assess the included studies (EPOC 2013). We also revised our use of study design nomenclature and handling of data (including re-analysis), in line with the latest EPOC guidance. Similarly, we reported the certainty of evidence provided by each group of studies using the GRADE criteria, and we included a 'Summary of findings' table. We also updated the outcomes to include 'adverse events' and 'equity' as recommended by the EPOC Group. As two included studies presented data in quarter years rather than months, we reported outcomes at 4- and 8-months rather than 3- and 6-months for these studies in Table 5 (DeVore 2016; Joynt 2016).

The original version of this review was published by Nicole Ketelaar, Marjan Faber, Signe Flottorp, Liv Rygh, Katherine Deane, and Martin Eccles (Ketelaar 2011). MF, LR, KD, and ME have since left the author team, and been replaced by David Metcalfe, Arturo Rios Diaz, Olubode Olufajo, Sofia Massa, and Daniel Perry.

INDEX TERMS

Medical Subject Headings (MeSH)

*Information Dissemination; *Quality Improvement; Canada; Consumer Health Information [*methods]; Evaluation Studies as Topic; Health Maintenance Organizations [standards]; Hospitals [*standards]; Medicaid; Organizational Innovation; Quality Assurance, Health Care [*methods]; Randomized Controlled Trials as Topic; Reproducibility of Results; United States

MeSH check words

Humans