

Review of general practice incentives: International evidence and literature review

Commissioned by:

Australian Department of Health and Aged Care

Conducted by:

Centre for Primary Health Care and Equity, Centre for Future Health Systems and School of Population Health, Faculty of Medicine and Health, University of New South Wales

Authors:

Prof Michael Kidd AO, A/Prof Joel Rhee, A/Prof Anurag Sharma, Dr Ben Harris-Roxas, Emeritus Prof Mark Harris AO, Cathelijne van Kemenade, Mary Burns, Kate Marshall, Dr Rula Ali, Dr Anthony Sunjaya, Rachael Kearns, Dr Michael Wright, Dr Shona Bates, Dr Nouhad El Hadad, A/Prof Frederic Sitas and A/Prof Margo Barr

April 2024

This report is copyright and confidential and has been prepared for the Australian Department of Health and Aged Care as a piece of commissioned work. Information provided in the report is confidential and may not be reproduced in part or whole or quoted without approval.

Suggested citation: Kidd M, Rhee J, Sharma A, Harris-Roxas B, Harris M, van Kemenade C, Burns M, Marshall K, Ali R, Sunjaya A, Kearns R, Wright M, Bates S, El Hadad N, Sitas F and Barr M (2024) Review of General Practice Incentives - International Evidence and Literature Review Report. Faculty of Medicine and Health, UNSW Sydney.

For further information, please contact Centre for Primary Health Care and Equity, Centre for Future Health Systems or the School of Population Health at the Faculty of Medicine and Health, UNSW.

Contents

Summary of Findings3
Acronyms and Glossary of Terms4
Background5
Objectives6
Methods7
Criteria for considering studies for this review7
Search methods for identification of studies8
Data collection and analysis8
Results
Description of studies11
Results for the specific research questions13
- RQ1: Evidence from the literature on the impact of blended funding models on quality outcomes in multidisciplinary settings
- RQ2: Evidence from the literature on the effectiveness of different payment models on preventive care for people with complex chronic disease
- RQ3: Benefits of funding primary care using different payment models
- RQ4: Evidence from the literature on what drives behavioural change in primary care providers 30
Discussion
Conclusions
References
Appendices47
Appendix 1: Database search strategies
Appendix 2: Inclusion and exclusion criteria for screening63
Appendix 3: Data extraction fields64
Appendix 4: Included studies – Description67
Appendix 5: Included reviews – Quality assessment

Summary of Findings

In Australia, the general practice sector faces significant challenges, prompting a re-evaluation of its funding models. This rapid literature review aims to inform the design of future payment models and performance incentives in primary health care. It focuses on systematic reviews related to funding models in general practice, emphasising multidisciplinary care, preventive services for complex chronic conditions, efficacy of payment models and drivers of behavioural change among providers.

Key databases were searched using PRISMA guidelines, supplemented by broader searches and citation tracking, yielding insights from 30 systematic reviews and 12 additional publications. The evidence suggests that some blended payment models could enhance care quality, particularly in multidisciplinary and preventive settings. However, improvements were mostly observed in incentivised activities, with a noted ceiling effect.

In conclusion, while blended payment models show promise, evidence on the cost-effectiveness of payfor-performance or capitation models is lacking. Cost-benefit analysis should be considered before implementing specific incentives.

Acronym / Term	Description	
ACE	Angiotensin-Converting Enzyme	
AMI	Acute Myocardial Infarction	
CKD	Chronic kidney disease	
CAPs	Capitation Payments	
EFFS	Enhanced Fee-for-Service (with P4P)	
FFS	Fee-for-Service	
GP	General Practitioner	
HbA1c	Haemoglobin A1C	
ITS	Interrupted Time Series	
NICE	National Institute for Health and Care Excellence	
NRT	Nicotine Replacement Therapy	
P4P	Pay for Performance	
РСР	Primary Care Provider	
PREM	Patient Reported Experience Measure	
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses	
PROM	Patient Reported Outcome Measure	
QALY	Quality Adjusted Life Years	
QOF	Quality and Outcomes Framework	
RCT	Randomised Controlled Trial	
RoB	Risk of Bias	
ROBIS	Risk Of Bias In Systematic reviews	
SMD	Standardised mean difference	
SR	Systematic Review	
WHO	World Health Organisation	

Acronyms and Glossary of Terms

Background

There are several critical issues facing general practice in Australia, not the least of which is the ageing of the health workforce and predicted shortages in the medical, nursing and allied health workforce. This workforce and the quality of care they provide are influenced by the way general practice is funded.

Blended models of funding general practice in countries around the world involve a combination of payment structures or funding sources to support primary care services. This approach often incorporates multiple elements, such as:

- 1. Fee-for-Service (FFS): Primary care service receives payments for each service provided or procedure performed. This could include consultations, specific medical procedures or certain tests conducted.
- 2. Capitation Payments (CAPs): Primary care service receives a set amount of funding per patient registered with their practice. This encourages ongoing care, preventive measures and overall patient management.
- 3. Performance-Based Payments (P4P): Incentives are tied to achieving specific healthcare targets or quality indicators. For instance, primary care service might receive bonuses for meeting certain clinical outcome goals, patient satisfaction levels or participation in preventive health programs.
- 4. Grants or Lump Sum Payments: Additional funding may be allocated for specific initiatives, equipment purchases or for participation in research or community health projects.
- 5. Risk-Sharing Models: Some models share financial risk between healthcare providers and payers, where primary care services might receive bonuses for cost-effective care or might bear some financial responsibility for excessive healthcare spending.¹

A blended funding model aims to create a more comprehensive and flexible approach to support primary care services in delivering quality care while considering various aspects of patient needs, preventive care and healthcare system goals. It is often designed to incentivise efficient, patient-centred care and to encourage better health outcomes.²

In Australia, despite the introduction of some blended payments, the current funding model is still predominantly based on FFS, and this may be less suited to supporting multidisciplinary prevention and management of long-term conditions. Another important issue affecting Australian general practice is the increasing proportion of services being funded by a combination of public (Medicare) and patient copayment.

These issues are not new and over the past 30 years, members of the research team have been involved in advising on or evaluating the development of policies and strategies to address them (including the original Future of General Practice Strategy of 1992, the Enhanced Primary Care Program in 2006, the National Primary Health Care Policy in 2010, the Medicare Review in 2018, the Health Care Homes Trials between 2017 and 2021, and the Primary Care Responses to the COVID-19 pandemic). These reports were reviewed and a revised strategy for general practice was developed by the Commonwealth's Strengthening Medicare Taskforce, which included representations from stakeholder groups involved in general practice.

MyMedicare was launched at the end of 2023 (<u>https://www.health.gov.au/our-work/mymedicare</u>), and introduced voluntary patient registration to strengthen the relationship between patients, their general practice, general practitioner (GP) and primary care teams. This blended funding model has been introduced alongside additional payments for longer consultations and telehealth consultations. The following new blended funding payments will be available to registered practices and providers:

- 1. The General Practice in Aged Care Incentive from 1 August 2024, which will support regular health assessments, care plans and regular GP visits for people in residential aged care homes.
- 2. New blended funding payments to support better care in the community for people with complex, chronic disease who frequently attend hospitals. These arrangements will roll out progressively across Australia over three years from FY2024–25.
- Chronic Disease Management items linked to a patient's registration in MyMedicare from November 2024, to support continuity of care for people with chronic and complex conditions. Patients who are not registered in MyMedicare will still be able to receive Chronic Disease Management items from their usual GP.

Objectives

The aim of this rapid literature review was to inform the future design of *payment models and performance incentives for primary health care*. Specifically, it focused on systematic literature reviews related to primary care and general practice incentives, and on funding arrangements using health provider funding mechanisms (including financial incentives) in blended funding models. It documented how these funding models drive quality care and access to care, and how they promote multi-disciplinary team care arrangements across providers.

The emphasis was on the impact of blended funding models on both individuals and teams working under an integrated model framework. The funding models were based on a combination of several payment types: salaried, capitation, FFS, outcome based and block funding models. Three main types of studies capture the effectiveness of such models: i) Randomised Control Trials (RCTs); ii) studies looking at the pre- and post-effects of an intervention (over multiple treatment and control sites) involving changes in the funding model/financial incentives using methods such as difference in difference estimation; and iii) studies capturing temporal variation using interrupted time series (ITS) methods to quantify effects of change in funding models.

The research questions were as follows:

- RQ 1 What is the evidence from the literature on the effectiveness of different payment models on quality outcomes in multidisciplinary settings?
- RQ 2 What is the evidence from the literature on the effectiveness of different payment models on preventive care for people with complex chronic disease?
- RQ 3 What are the benefits of funding primary care using different payment models (e.g. block funding, FFS, outcomes based, incentives, salaried models and others), including the interactions between different funding models in blended systems and the outcomes they produce for people and the health system?
- RQ 4 What is the evidence on what drives behavioural change in primary care providers with regard to providing quality, patient-centred care?

The focus of this review was to identify:

- 1. International blended funding models for primary care and/or general practice, considered in the context of a country's health system.
- 2. Evidence of payment types, including financial incentives in international funding models, that impact on multidisciplinary, person-centred quality care in primary care and general practice.
- 3. How financial incentives impact on quality of care in primary care, including for patients with complex needs.

While the focus was on public funding models, mixed public-private (i.e. Medicare and a patient fee, or insurance as in the US) funding models and their impact on the quality of care were also examined. Scott et al. (2011) defined quality of care as "the degree to which changes in physician behaviour improve the well-being of patients" that is assessed by "patient reported outcome measures, clinical behaviours and intermediate clinical and physiological measures".³ The paper also referenced Donabedian (1966) who included structures and systems of care (e.g. Information Technology), processes (e.g. recording of assessment) and outcomes (e.g. clinical indicators and health status of patients).⁴ Our review largely followed this definition, although this was explicit in only a minority of reviews and did not include some of the dimensions of quality primary care identified by WHO (2024), such as integration and efficiency.⁵

Methods

This rapid review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.⁶

Criteria for considering studies for this review

Scott et al. (2011), in their Cochrane review on the effect of financial incentives on the quality of health care provided by primary care physicians, looked at literature between 2000 and 2009, specifically examining: i) different types of financial incentives that have improved quality; ii) characteristics of patient populations for whom quality of care has been improved by financial incentives; and iii) characteristics of primary care providers (PCPs) who have responded to financial incentives.³ The present review followed the same approach but focused on systematic reviews and meta-analyses, and limited the timeframe to between January 2010 and November 2023. To illustrate, Box 1 describes the search inclusions that were used, with a complete list provided in Appendix 1.

Types of studies 1. Systematic reviews 2. Meta-analyses 3. Key citations included in systematic reviews and/or meta-analyses **Types of participants** Individuals and/or settings offering or receiving primary health care: 1. General practice 2. Primary health care **Types of interventions** Primary care and general practice incentives and funding models: 1. Capitated funding 2. Performance incentives 3. Blended funding Types of outcome measures Primary outcome measures include the quality of care provided by primary care providers that are related to patients' health and well-being: 1. Access to care 2. Quality of care 3. Preventive care 4. Multidisciplinary care 5. Health outcomes 6. Cost of health care 7. Provider behaviour

Box 1: Description of search inclusions used.

Search methods for identification of studies

The following databases were searched in November 2023: PubMed, Cochrane, Embase, CINAHL, PsycINFO, PAIS and Web of Science. Search terms covered funding mechanisms, primary care and general practice, and systematic review/meta-analysis (2010-2023). See Appendix 1 for detailed search terms for each database. Additional wider searches were conducted by the researchers, as well as citation searches of the included systematic reviews/meta-analyses, to ensure that all relevant literature was identified.

Data collection and analysis

Selection of studies

Publications identified from the database searches were entered into Covidence (a screening and data extraction tool for conducting systematic reviews https://www.covidence.org/). Duplicates were identified and removed. Eligible papers needed to be systematic reviews or meta-analyses assessing health outcomes of funding mechanisms in primary care settings in high-income countries. Two reviewers independently screened the titles and abstracts based on the preset search criteria provided in Appendix 2.

Once the title and abstract screening were completed, the full text screening was conducted by two members of the team. Any discrepancies were discussed and resolved by the reviewers. If disagreements could not be resolved, a third reviewer was asked to resolve. For excluded papers, a specific reason was entered into Covidence.

Data extraction

Data abstraction was conducted using a modified version of the PRISMA checklist for systematic reviews. Items extracted covered objectives, methods of screening, data extraction, quality assessment and synthesis, results, limitations and interpretation. Data extraction was checked by a second reviewer. The specifics of data extraction items can be found in Appendix 3.

Quality assessment

The quality of each included review paper was evaluated using the ROBIS (Risk Of Bias In Systematic reviews) tool.^{7,8} ROBIS is designed to assess the risk of bias (RoB) in reviews with questions relating to interventions, aetiology, diagnosis and prognosis, and is completed in three phases: i) assess relevance (optional), ii) identify concerns with the review process (specifically Study eligibility criteria, Identification and selection of studies, Data collection and study appraisal and Synthesis and findings), and iii) judge the RoB.

Data synthesis

For each publication, the researchers examined heterogeneity, methods, funding models and settings. Outcomes analysed included access to care, quality of care, preventive care, multidisciplinary care, health outcomes, cost of health care and provider behaviour. Specific synthesis was undertaken for each of the four research questions. The availability of evidence in the literature was categorised as sufficient or insufficient by the authors, with sufficient evidence relying on a consideration of the reliability, relevance and sufficiency of the information obtained from various sources.

Results

The systematic search revealed 754 publications. Removing duplicates left 448 publications, of which the title and abstract were screened. Ninety-four articles were full-text assessed; 30 fulfilled the criteria for inclusion. The PRISMA flowchart can be seen in Figure 1 below. The search of other literature and specific citations resulted in the identification of 12 additional publications.



Figure 1: PRISMA flowchart of systematic review selection process.

Description of studies

The systematic search resulted in the inclusion of 22 systematic reviews,^{9–30} four meta-analyses,^{31–34} two reviews of reviews,^{35,36} one review³⁷ and one qualitative synthesis.³⁸

Additional searches identified 12 cohort and observational studies. Eleven were selected from economic literature^{39–49} and one specifically focused on multidisciplinary team work.⁵⁰

Characteristics of countries and settings

The systematic reviews assessed interventions in several countries. The majority of studies evaluated schemes in the UK (25), US (19), Europe (a total of 15 across Denmark, Finland, Norway, Sweden, Germany, Netherlands, Belgium, Italy, Spain, France, Ireland, Poland) and Canada (14). Nine included studies conducted in Australia. A smaller number of studies included in the reviews originated in Asia-Pacific (New Zealand, Taiwan, Japan, China, Philippines, South Korea, India), Africa (South Africa, Rwanda, Tanzania, Uganda, Zambia, Burundi, Democratic Republic of Congo), South America (Uruguay, Argentina) and Middle East (Türkiye, Israel) (Figure 2).



Figure 2: Geographical representation of origin of studies included in systematic reviews.

The focus of the systematic search was on general practice, but some of the eligible studies were set in other related primary health care settings including dentistry, allied health, nursing and mental and community health. The health areas studies targeted covered chronic disease, diabetes, cardiovascular disease, respiratory disease, mental illness, lower back pain, smoking cessation, cancer screening and laboratory tests and processes.

Additional searches identified two studies that focused on multidisciplinary settings in the US.^{50,51}

More detailed information on included studies can be found in Appendix 4.

Characteristics of the financial interventions and outcomes

The main payment scheme assessed across the reviews was P4P (n=22), but also included were FFS, CAPs, bonus payment, mixed and other models. The aim of the interventions was to improve the quantity and/or quality of health service provision through, for example, the implementation of indicators, changing provider behaviour or retention of medical work force.

Eleven studies provided further information particularly on FFS and capitation models.^{39–49}

Quality assessment of included systematic reviews

The overall RoB outcome assessments for the included articles were rated as low, high or unclear, with the majority (50% n=15) categorised as low RoB overall.

RoB were also provided for each of the Phase 2 concerns (1. Study eligibility criteria, 2. Identification and selection of studies, 3. Data collection and study appraisal and 4. Synthesis and findings). As shown in Figure 3, the majority were categorised as low RoB (90% (n=27) for study eligibility criteria, 63% (n=19) for identification and selection of studies, 63% (n=19) for data collection and study appraisal and 60% (n=18) for synthesis and findings). The ROBIS assessments for each article are provided in Appendix 5.



Figure 3: Overall assessment of risk of bias assessment.

Results for the specific research questions

The specific results for each of the research questions are provided in the next sections.

- RQ 1 What is the evidence from the literature on the effectiveness of different payment models on quality outcomes in multidisciplinary settings?
- RQ 2 What is the evidence from the literature on the effectiveness of different payment models on preventive care for people with complex chronic disease?
- RQ 3 What are the benefits of funding primary care using different payment models (e.g. block funding, FFS, outcomes based, incentives, salaried models and others), including the interactions between different funding models in blended systems and the outcomes they produce for people and the health system?
- RQ 4 What is the evidence on what drives behavioural change in primary care providers with regard to providing quality, patient-centred care?

RQ1: Evidence from the literature on the impact of blended funding models on quality outcomes in multidisciplinary settings

Mitchell et al. (2008) defined multidisciplinary care as "professionals from a range of disciplines work together to deliver comprehensive care that addresses as many of the patient's needs as possible".⁵² Thus multidisciplinary primary care settings are those in which this occurs. This can include settings (as in UK general practice) which involve professionals employed by different organisations or in different ways in the same service.

In 2010, to inform policy development in Australia, Scott et al. (2011) conducted a review to evaluate the effect of blended payment schemes for primary care that involved financial incentives for quality of care.³ In the seven studies included in the review, there was insufficient evidence to either support or not support their use. A key concern of this review was the lack of control over the selection of primary care providers into or out of the incentive schemes, with observed improvements in performance possibly being due to selection rather than actual changes in quality of care.

Since 2000 in Ontario, Canada, there have been new primary care funding schemes that have dramatically changed the mix of blended payments.⁵³ This saw a decrease in FFS and increases in CAPs and P4P.⁵⁴ The introduction of these schemes was based on the assumption that it would better suit multidisciplinary team care and improve access.

In the UK in 2004, the P4P "Quality and Outcomes Framework" (QOF) was introduced to fund up to 25% of GPs' income. Funding was linked to the achievement of quality targets for several chronic conditions for practice-based (rather than practitioner-based) incentive schemes (see Table 1) and blended with existing CAPs.⁵⁵ It resulted in improvements in quality of care for those conditions/aspects of care at the expense of other care that was not incentivised.⁵⁶

Since 2001, primary care in New Zealand has been funded through a mix of copayments and government funding (largely based on capitation at the level of the primary care organisation, and aiming to improve access and reduce inequalities).⁵⁷ In Europe there is considerable variation, but primary care physicians have been historically paid through a mix of CAPs and FFS. However, more recently P4P has been introduced.⁵⁸ In the US there have been efforts to expand the role of primary care in patient centred medical homes and accountable care organisations through a mix of private and public funding (from Medicaid for disadvantaged populations and Medicare for the older population).⁵⁹ Primary care is still funded predominantly through FFS with an increasing proportion of CAPs.⁶⁰ There have been some trials of P4P.

DOMAINS			
Clinical	Organisational		
Secondary prevention of coronary heart disease	Records and information		
Cardiovascular disease: primary prevention	Information for patients		
Heart failure	Education and training		
Stroke and transient ischemic attack	Practice management		
Hypertension	Medicines management		
Diabetes mellitus			
Chronic obstructive pulmonary disease	Patient Experience		
Epilepsy	Length of consultations		
Hypothyroidism	Patient survey (access)		
Cancer			
Palliative care	Additional Services		
Mental health	Cervical screening		
Asthma	Child health surveillance		
Dementia	Maternity services		
Depression	Contraception		
Chronic kidney disease			
Atrial fibrillation			
Obesity			
Learning disabilities			
Smoking			

Table 1: Domains of the Quality and Outcomes Framework.²⁷

Results

Overall description

There was sufficient evidence that blended payments involving P4P in combination with CAPs or FFS were likely to bring about small improvements in quality of care in multidisciplinary primary health care settings. This was more pronounced if the performance measure was a process of care measure (e.g. recording of risk factors) rather than outcome measures. There was some evidence that P4P had greater impacts on quality of care when it was combined with non-financial incentives (e.g. reputational) and quality improvement. There was insufficient evidence for improvement in health outcome measures.

There was insufficient evidence that increasing CAPs relative to FFS improved quality, however there was some evidence that service provision tended to decrease. There was insufficient evidence that either capitation or P4P reduced health inequalities. There were some mixed qualitative and quantitative findings suggesting that care which was not incentivised did not improve, and that targets may be achieved at the expense of holistic person-centred care. There was some evidence that P4P and CAPs, especially where funds were pooled at the practice level, supported teamwork and collaboration between primary care providers and disciplines.

The quality of the reviews was variable, with those with low RoB being less likely to show positive effects. Most of the research on quality of care in multidisciplinary primary care settings was from the UK, US and Canada.

Main findings

UK (QOF)

Gillam et al. (2012) reviewed 94 quantitative and qualitive studies on P4P between 2004 and 2011 as part of the QOF in UK primary care, implementing P4P in combination with CAPs.⁶¹ This low RoB review concluded that there were modest improvements in quality of care for chronic conditions and improvements in data recording and teamwork (especially for nurses), but that the impact of costs and patient experience was uncertain. The amount of change was not quantified.

Mandavia et al. (2017) reviewed the impact of the introduction of P4P on chronic disease quality of care.⁶² It included 28 studies in the UK, with 21 in primary care. Most reported positive results, including improvement in process outcomes (e.g. recording rates, referrals), or intermediate results (e.g. quality indicators). This review had uncertain RoB. The amount of change was not quantified.

Ahmed et al. (2021) reviewed 24 quantitative and ten qualitative papers, published between 2009 and 2019, on the impact of P4P vs CAPs in UK general practice under the implementation of the QOF.⁶³ This review of uncertain RoB reported mixed results, with some improvement in personalised care at the expense of holistic personalised care. The amount of change was not quantified.

Canada

Carter et al. (2016) reviewed 14 studies as part of primary care reforms in Canada (Ontario, Quebec and Alberta) with the introduction of models of blended CAPs and P4P in the context of multidisciplinary team based primary care.⁶⁴ This low RoB review found that blended payments led to improvements in processes of care (especially screening, prevention and chronic disease management), while reducing the number of patients seen each day. In several studies there were significant reductions in Emergency Department visits and avoidable admissions (eg RR between 0.75 and 0.82, p < 0.05).

Multiple jurisdictions

Jia et al. (2021) reviewed 27 studies, published pre-2020, on the impact of blended payment methods on providers working in outpatient healthcare settings in US (16), Canada (3), Denmark (3), Australia (2), Taiwan (1), Germany (1) and India (1).³⁴ Interventions applied at the practice level (such as the UK QOF) were excluded. This low RoB review found uncertain evidence that P4P combined with CAPs, salary or FFS increased the quality of care and quantity of some services (e.g. immunisations RR 1.27, 95% Cl 1.19 - 1.36; prescribing to guidelines RR 1.07, 95% Cl 1.02 - 1.12). Both P4P and FFS funding at a higher fee induced service providers to provide more services.

Ogundeji et al. (2016) reviewed 96 studies, published pre-2016. All studies were included in a logistic regression and 37 in a meta-analysis and regression. Sixty-eight P4P schemes were evaluated, with 270 health outcomes in a variety of healthcare settings.³¹ The majority of studies were based in the UK (e.g. QOF) and US, with at least one study based in China, Rwanda, Tanzania, Türkiye, Zambia, Israel, Taiwan, Uganda, Australia, the Netherlands, Canada, Italy, Haiti and Malawi. The review had uncertain RoB. There was a statistically positive overall standardised mean difference (SMD) in outcomes of 0.13, (95% CI: 0.02 - 0.24, p < 0.0001), in favour of P4P. The multivariate logistic regression, adjusting for design features and evaluation type, showed that higher P4P incentives (e.g. > 5% of the individual or group's income) were three times as likely to produce a positive outcome (95% CI: 1.07 - 10.64, p = 0.04). There were non-statistically significant increases in the odds of success if incentives were paid to individuals versus groups/organisations (OR 2.01, 95% CI: 0.62 - 6.56, p = 0.37), and when there was low perceived

risk of not achieving the incentive (OR 0.61, 95% CI: 0.22 - 1.75, p = 0.37). The meta-analysis found no effect for P4P in improving health outcomes (SMD 0.61, 95% CI: 0.22 - 1.75, p = 0.37) such as hospital mortality rates or smoking cessation. However, larger effects were found for intermediate outcomes (SMD 0.07, 95% CI: -0.01 - 0.15) such as blood pressure or cholesterol reduction, and process measures (SMD 0.18, 95% CI: 0.06 - 0.31) such as cancer screening or smoking cessation advice. P4P schemes evaluated using less rigorous designs (e.g. non-controlled and quasi-experimental designs) were reported to have 24 times the odds of having positive estimates of effect than that found with RCTs.

Scott et al. (2018) reviewed 80 studies published between 2010 and 2015, evaluating 44 P4P schemes in the examination of value-based purchasing in healthcare.⁶⁵ Forty-two studies were based in the US, 16 in the UK, nine in Taiwan and the remainder in China, Canada, Italy, Australia, France, the Philippines and Rwanda. About half the studies showed positive impacts, and where P4P was used together with other programs, such as quality improvement, there was a higher likelihood of a positive effect. Where the P4P was used to increase physician income alone, there was a lower likelihood of positive effect. The size of incentive payments, relative to revenue, was not associated with the likelihood of positive outcomes (regression analysis showed that 1% increase in the size of the incentive led to a 1.2% increase in the proportion of positive outcomes, p > 0.5). The combination of rewards for P4P, with rewards for reducing costs, was no better than P4P alone in achieving positive outcomes.

Tao et al. (2016) reviewed 27 studies between 1980 and 2013, focusing on health inequities and comparing blended P4P and CAPs with FFS alone.⁶⁶ Fifteen studies were based in the UK, with the remainder in the US and Canada. This review was of uncertain quality and found no impacts for either scheme on socioeconomic or racial inequalities in access, utilisation or quality of primary care.

Van Herck et al. (2010) reviewed 128 studies (124 in primary care settings) between 1990 and 2009, focusing on the impact of P4P in improving healthcare.⁶⁷ Sixty-three studies were based in the US, 57 were based in the UK, and the remaining six were based in Australia, Germany, Argentina and Italy. The addition of P4P improved quality of care by about 5%, but there was a lot of variation especially by condition. Process measures (such as recording of smoking status) were more likely to improve than outcome measures (e.g. complications of diabetes) but this was not tested statistically. One study found a positive relationship between P4P and a multidisciplinary team approach, the use of clinical pathways and having adequate human resources for quality improvement projects. This low RoB review made the following six recommendations:

- 1. Select and define P4P targets based on baseline room for improvement (to avoid payments for performance that has already reached a ceiling)
- Make use of process and (intermediary) outcome indicators as target measures (such as recorded assessment of blood pressure and achieving blood pressure control associated with improved quality of life and mortality)
- 3. Involve stakeholders and communicate the program thoroughly and directly throughout development, implementation and evaluation (in order to increase commitment and engagement)
- 4. Implement a uniform P4P design across payers (to reduce complexity for providers)
- 5. Focus on quality improvement and achievement for both low and high performing providers (to both improve effectiveness and reduce inequities)
- 6. Distribute incentives at the individual level and/or at the team level (to motivate individual providers and increase engagement of non-physician members of the care team).

Wranik et al. (2019) reviewed 77 qualitative and quantitative studies from 2003 to 2016, examining a range of factors influencing multidisciplinary teamwork in primary care and identifying absence of adequate funding as an obstacle to teamwork.¹⁶ This review was assessed to have a low RoB. The QOF in the UK and pooling of funds at team level in Canada were assessed to have increased teamwork and collaboration in comparison with FFS. Quantitative findings were not synthesised across studies.

Reviews of Reviews

Two reviews of reviews were identified. However, there was some duplication as these included some reviews already cited above. Flodgren et al. (2011) reviewed four systematic reviews up to 2010, examining the effects of blended combinations of FFS and CAPs in the US, UK, Finland, Canada, Denmark, Germany and Ireland.⁶⁸ This low RoB review of reviews concluded that increasing financial incentives was associated with improvements in processes of care (e.g. recording of assessment) but was ineffective in improving adherence to clinical guidelines. Heider et al. (2020) reviewed 21 reviews up to 2020, on the effects of monetary incentives on primary healthcare services.⁶⁹ Seven reviews focused on the US and the remainder consisted of a combination of high-income countries. This high RoB review of reviews found that P4P and accountable care organisations had positive impacts on process quality measures. Quantitative effects or significance levels were not reported in these two reviews of reviews.

Additional review

We performed a review of individual studies conducted since 2010, focusing specifically on multidisciplinary teams in primary care. This identified one qualitative study by Bitton et al. (2012) from the US and not included in the reviews. The study evaluated a payment-linked patient-centred medical home in 2012.⁵⁰ It observed that P4P payment reform provided capacity to engage in team change (away from the "hamster wheel" of FFS) rather than necessarily motivating individual practitioners to change their practices (especially given the finding that clinical staff were often unaware of incentive formulas).

Key findings

- 1. Sufficient evidence that blended payments which included P4P improved quality of care in multidisciplinary primary care settings. There was more evidence of change in process rather than outcome measures of quality of care.
- 2. Insufficient evidence that CAPs improve quality of care.
- 3. Effects of payments on quality of care may be enhanced by providing incentives at both the individual and team level or linking incentives to other activities such as training and quality improvement.
- 4. Insufficient evidence for or against the impacts of P4P or CAPs on health equity.
- 5. Evidence from qualitative research suggesting that P4P and capitation may improve teamwork if funding is pooled at the level of teams.

Limitations

The reviews were of variable quality (RoB) and did not include metanalyses because of the heterogeneity of included studies. Relatively few studies (12/30) specifically focused on evaluating the impacts on multidisciplinary teams within primary care.

Conclusions

There was some support for the influence of P4P models on quality of care. However, when considering the benefits of blended payment models in team-based primary care, several factors should be taken into account, including: i) how benefits are distributed between organisations and providers and between individual providers and teams; and ii) the ability of primary care providers (and groups of their patients) to include or exclude some patients from enrolment (this effect minimised by encouraging rostering to ensure continuity of care), as this may distort the effects on performance and thus funding.³ We were unable to identify sufficient evidence for the impact of other funding models on quality of care in primary care.

There was also variation between healthcare systems, suggesting the importance of the overall context, including workforce supply. Finally, qualitative research suggested possible unintended consequences that need careful monitoring, such as a potential reduction in holistic patient-centred care, especially for care that was not incentivised. Figure 4 provides a summary.



Figure 4: Impacts of P4P in Quality Outcomes (adapted from Markovitz and Ryan 2017¹²).

RQ2: Evidence from the literature on the effectiveness of different payment models on preventive care for people with complex chronic disease

The increasing burden of complex chronic disease presents a major challenge in healthcare. In Australia, the 2022 National Health Survey reported that 81.4% of Australians had at least one long-term health condition, with 49.9% living with a chronic condition.⁷⁰ Multimorbidity, where an individual has two or more chronic conditions,⁷¹ affects around 40% of Australians aged over 44 years, increasing to 50% among those aged 65 to 74 and 70% for those over 85 years.⁷² Multimorbidity is associated with higher mortality, increased healthcare utilisation and a greater symptom burden, posing complex challenges for healthcare management and overall quality of life.^{73–75}

As multimorbidity becomes more prevalent, the complexity of care required also rises. For example, in Australia, the rise in complex care needs is shown by an increase in extended Medicare consultations and chronic disease management provided by GPs. From 2010 to 2019, long consultations (20 to 40 minutes) increased by 104%, and those over 40 minutes by 125%, while chronic disease and mental health claims rose by 182% and 202%, respectively.⁷⁶

Effective preventative care is designed to minimise preventable morbidity and mortality, as well as improve wellbeing and reduce the overall need for medical services. Preventative strategies can be implemented at primary, secondary and tertiary levels: primary prevention focuses on health promotion and preventing disease; secondary prevention includes screening and diagnosis of early disease and risk factors; and tertiary prevention aims to monitor and prevent the progression of disease complications. Evidence suggests patients with multimorbidity, especially those with mental illness, are less likely to use primary care preventative services.⁷⁷

People with complex chronic illness and multimorbidity often require frequent and coordinated health services. At present, single-disease approaches to care are not designed to address multimorbidity, resulting in fragmented care and poorer health outcomes.⁷⁸ Existing international funding models may exacerbate these challenges; for example, the UK's QOF, a P4P initiative incentivising primary care practices for specific activities in chronic disease management, does not account for multimorbidity⁷⁹ and GPs are not incentivised to focus on multimorbidity. Team-based approaches to primary care have demonstrated benefit for people with complex chronic illness;⁸⁰ however, they may not be suited to typical FFS payment models. Evidence on funding models which best support the unique needs of patients with complex chronic illness and multimorbidity is currently limited.

Results

Overall description

Ten reviews were eligible for inclusion.^{14,17–19,21,23,26,29,32} Of these, one review (a Cochrane systematic review) included meta-analytic data.³² Reviews were published between 2010 and 2021, and the number of studies included ranged between 11 to 128. Six reviews reported the types of complex chronic illnesses in patients, including diabetes, stroke, asthma, hypertension, severe mental illness and cardiovascular and respiratory diseases.^{14,18,19,23,25,32} Four reviews included studies reporting on preventative care services (e.g. cancer screening, immunisations or smoking cessation assistance) for the general patient population.^{17,26,29,32} All reviews included studies based in primary care settings. One

review included studies reporting on general practice care within a mental health service,³² and one review included studies based in hospital settings.²⁵

Reviews included various study types, including RCTs (6/10), ITS (6/10), observational (5/10) and cohort studies (4/10). One review lacked details on its 35 studies and did not specify study types, making it difficult to determine the types of studies included.²¹ Studies included in the reviews were primarily based in the UK. Other countries included the US, Canada, Australia, France, the Netherlands, Germany, Ireland, Spain, Italy, Argentina, China and Taiwan. One review reported findings from low- and middle-income countries (e.g. Afghanistan, Burundi, Democratic Republic of Congo, Rwanda and Tanzania); ³⁰ however, for the purpose of this report, findings related to high-income countries were included.

All reviews included studies reporting on P4P models. Three reviews compared payment models (e.g. P4P versus CAPs, FFS versus CAPs) across studies.^{17,27,30} Two reviews focused primarily on studies examining the QOF P4P scheme in the UK.^{18,23} The majority of reviews examined payments at the practice/clinic/organisation level^{14,17–19,23,25,29,32}; some included individual clinician payments as well as practice-level payments but did not make comparisons or analysed them separately^{14,17,29}. One review limited its analysis to financial incentives provided to individual clinicians.²⁶ One meta-analysis compared the impact of incentive payments to groups of clinicians/practices with payments made to individual clinicians and found no significant difference in outcomes (p=0.37).³¹

Main findings

This section presents findings on outcome measures for quality preventative care in complex chronic diseases. It covers primary health outcomes (e.g. mortality, morbidity and hospitalisation rates), intermediate clinical parameters (e.g. blood glucose, HbA1c monitoring and blood pressure), preventative care activities (e.g. cancer screening and smoking cessation advice), patient-reported outcomes and experiences, and concludes with equity measures.

Primary health outcomes

Few studies among the reviews assessed the impact of financial incentives on primary health outcomes in complex chronic illness. For example, Houle et al. (2012), a low RoB review, included a hypertension study that examined end-organ complications such as heart failure, AMI, stroke and renal failure.²⁴ Notably, one review by Langdown and Peckham (2014) highlighted that among the UK's QOF five types of clinical indicators, there was only one health outcome indicator, this being the seizure-free status of patients with epilepsy over the past 15 months.¹⁸ The RoB of this review was rated as unclear. The paucity of studies utilising primary health outcomes may reflect their limited use in P4P and other financial incentive programs.

Ahmed et al. (2021) found in cross-sectional studies that the introduction of the UK's QOF was associated with a reduction in mortality rates and increases in quality-adjusted life years (QALYs) across the general population.²¹ Unfortunately, major limitations were identified regarding these findings, due to the use of a large number of indicators (leading to the risk of multiple hypothesis testing) and to trends showing improvements in some of these parameters before the introduction of QOF.²¹ The RoB was rated as unclear.

Intermediate health outcomes

A multitude of intermediate health outcomes were used in the studies to measure the quality of preventative care provided to people with complex chronic diseases. Examples included health behaviours (e.g. smoking status or quit rates) and clinical parameters such as blood glucose, HbA1c and blood pressure levels.^{26,29}

Overall, P4P programs and general practice financial incentives showed varied and mixed effects on intermediate health outcomes in chronic disease patients, with the strongest evidence observed in diabetes care. One meta-analysis of low RoB by Yuan et al. (2017) and consisting of one RCT, showed little or no impact of P4P programs on blood pressure or cholesterol control.³² Other reviews found that P4P programs, particularly the UK's QOF, led to small improvements in intermediate health outcomes (e.g. blood glucose, blood pressure, lipid profile) among patients with diabetes and possibly asthma, but there was mixed evidence for coronary heart disease and hypertension.^{14,18,21,25,26} A review by Hamilton et al. (2013) examining the impact of primary care financial incentives on smoking cessation, found it difficult to ascertain whether improvements in quit rates were due to these incentives or other health system factors.²⁹ This review had a high RoB.

Preventative care activities and process measures

Most studies investigated preventative care activities and process measures. Studies reviewed focused on measuring health behaviours, such as vaccinations (including seasonal influenza), smoking cessation advice, cancer screening (especially breast, cervical and colorectal cancers), regular retinal examinations and monitoring of blood pressure, HbA1c, lipids and renal function. Prescription of medications [such as Angiotensin-converting enzyme (ACE) inhibitors and nicotine replacement therapy (NRT)] and referrals to other health services were also examined.^{17,25,26,29}

Preventative care activities and process measures examined in many of the reviews (e.g. vaccination rates, mammography cancer screening rates and smoking cessation activities) were not restricted to people with complex chronic diseases.²⁹

Overall, reviews provided sufficient evidence that financial incentives and P4P programs improve preventative activities and care processes especially for patients with diabetes. However, there is insufficient evidence for other preventive outcomes such as reducing smoking or weight. One metaanalysis by Yuan et al. (2017) of low RoB and involving a small number of studies, showed a small positive impact of P4P programs on the use of some tests and treatments; however, no impact was found for adherence to quality assurance criteria.³² Other reviews found mixed evidence that financial incentives result in positive changes to preventative care activities and chronic disease care processes. For example, one low RoB review by Van Herck et al. (2010), found a mixture of positive and neutral impact of P4P incentives on the performance of preventative care activities for people with diabetes, heart failure and asthma.²⁵ Another low RoB review by Mauro et al. (2019), found no changes in the breast cancer screening rates due to P4P, although results were mixed regarding colorectal and cervical cancer screening rates.¹⁷ Financial incentives improved the recording of smoking status and increased smoking cessation advice and referrals; however, evidence for prescribing NRT was sufficient, but inconsistent.²⁹

One review by Houle et al. (2012) and of low RoB, noted that studies with weaker study designs tended to report more positive results, with uncontrolled studies showing improvement in the adherence to

quality-of-care process indicators such as vaccination rates, colorectal cancer screening and the frequency of HbA1c monitoring.²⁶ Studies employing more robust designs (such as adjusted uncontrolled studies, non-randomised controlled studies, controlled before-and-after studies and randomised controlled studies) did not show improvements or only minor gains in some indicators.²⁶

Patient reported outcomes and experience measures

Three reviews examined the impact of financial incentives on patient satisfaction with care; however, analyses were not limited to patients with chronic diseases.^{14,21,25} Overall, the impact of P4P programs and financial incentives on patient satisfaction and experience measures was mixed, with a combination of positive, neutral and negative impact.^{14,21,25} The UK's QOF improved patient experience in some areas,²¹ yet also resulted in negative impacts, such as consultations being less likely to meet patient needs and concerns because practices focused on more measurable aspects of care at the expense of less quantifiable factors like patient-doctor relationships.²¹ No studies reported on other common patient-reported outcomes, such as health status and quality of life.

Equity

Five reviews examined the impact of quality-based payment incentives on equity of chronic disease care.^{14,19,21,23,25} Overall, findings were mixed and variable, with some reviews reporting positive impact on health equity, while others showed a neutral or negative impact. One review found that any improvements in care tended to favour older male patients and less deprived individuals; while, outcomes did not differ according to ethnicity.²³ Another review concluded that P4P programs had little or no impact on socioeconomic and racial inequity in people with diabetes, cardiovascular diseases, chronic obstructive pulmonary disease and preventative healthcare.¹⁹ Likewise, Van Herck et al. (2010) did not find any negative impact of P4P incentives on healthcare equity.²⁵ On the other hand, Ahmed et al. (2021) concluded that the UK QOF reduced the gap in quality care between the most and least deprived practices.²¹ The extent to which factors such as age, gender, ethnicity and deprivation influence intermediate health outcomes following the implementation of P4P remains unclear.²³

Key findings

- 1. Insufficient evidence was found as to the impact of P4P or financial incentives in general practices on primary health outcomes (e.g. mortality rates), particularly among patients with complex chronic disease.
- 2. Insufficient evidence was attributed to the scarcity of research focusing on primary health outcome measures, possibly due to many P4P programs emphasising other quality indicators.
- Some studies indicated positive effects of P4P or financial incentives, preventative care activities, process measures and intermediate health outcomes, particularly in patients with diabetes; however, these findings varied widely across studies, disease types, outcome measures and program designs.
- 4. Positive impacts, where present, tended to be small, and improvements were mostly confined to incentivised activities, sometimes negatively affecting non-incentivised areas.
- 5. There was an observed ceiling effect in improvements, plateauing after reaching the maximum thresholds set for financial incentives in program implementation.

Limitations

Outcome measures were largely focused on preventative activities, process measures and intermediate health outcome measures, rather than on primary health outcome measures. This may reflect the use of quality indicators in programs such as the UK's QOF, which focus heavily on care activities, process measures and intermediate health outcomes.¹⁸ Use of intermediate health outcome indicators requires pre-defined thresholds as a marker of optimal and sub-optimal care. This can lead to potential problems if the threshold values differ from the clinical practice guideline recommendations, as was the case in the UK's QOF with spirometry and blood pressure targets.¹⁸

Heavy reliance on process measures could also inadvertently change the focus of health providers from quality of care to quality of record-keeping, potentially introducing unintended consequences such as exception reporting (i.e. exclusion of some patients from the denominator to improve target rates) or even the avoidance of complex patients whose treatment targets may be difficult to meet.²⁶

There was no evidence of an immediate impact of P4P programs on non-incentivised activities; certainly no improvements were noted.^{18,26} However, a trend of decline in non-incentivised activities emerged over a longer term.¹⁸

The effect size of the improvements with P4P programs and other general practice financial incentives, appeared to be small and variable. A meta-analysis involving a small number of studies, showed little to no impact on intermediate health outcomes (blood pressure or cholesterol control) and a small positive impact of P4P programs on the use of some tests and treatments; however no impact was found for adherence to quality assurance criteria.³² Similar conclusions were drawn in another systematic review which quantified the improvements as around 5%, but with a substantial variability depending on the program design, the indicators used to measure outcomes and practice characteristics.²⁵

In the UK QOF it was noted that the quality of care indicators improved in the first few years after the introduction of the P4P program, before reaching a plateau.^{18,19,25} This may be due to a ceiling effect caused by the practices reaching the maximum threshold levels set for the financial incentives.^{18,21,25} The thresholds for financial incentives were set between 50% and 90%, which provided little incentive once practices reached the maximum threshold, and did not encourage practices to improve care for more challenging patient groups.¹⁸ On the other hand, raising the maximum threshold to 100% could inadvertently introduce inequities by penalising practices serving more deprived areas.¹⁸

Conclusions

Evidence on the impact of P4P and general practice financial incentives on primary health outcomes like mortality and morbidity, is insufficient. There is some evidence, though variable, of their positive influence on preventative care activities, process measures and some intermediate health outcomes, particularly in patients with diabetes. There is mixed evidence of the impact of these incentives on patient-reported outcome and experience measures, as well as equity. Findings should be treated with caution due to the generally low to moderate quality of the studies. Moreover, there was considerable variability in results across disease types, outcome measures and program design. Where there was positive impact, the effect size was often small and changes were often limited to incentivised activities with neutral or even negative impact on non-incentivised activities. Some programs also showed a ceiling effect over time as practices reached the maximum threshold set for financial incentives.

RQ3: Benefits of funding primary care using different payment models

Physician remuneration could potentially be an important efficiency enhancing policy tool in publicly funded health systems, particularly for improving access, quality of care and overall patient outcomes. Economic theory provides a framework to formalize a relationship between how physicians are paid or compensated, and volume and quality of services provided. For example, a utility maximisation framework is used to model the trade-off physicians face between work and leisure. When combined with the profit maximization approach, targeted income hypothesis and physician's altruism, this could lead to a theoretical determination of the optimal number of services a physician will provide under a certain type of payment model and how this behaviour will change by altering the payment structure. For example, an FFS model might encourage low-value care or unnecessary services when patient numbers are low (to reach a certain target income), whereas an alternative payment model (such as CAPs) could potentially lead to under provision of services and primary care physicians preferring to enrol less complex and relatively healthy individuals in their practice (cream skimming). From a policy point of view, it is imperative to minimise the above extreme effects of both payment models while ensuring: i) multiple objectives are met, such as medical graduates finding primary care specialization lucrative (to avoid any skill shortages at the level of primary care); and ii) continuity of quality care at primary level, especially for complex patients with chronic conditions or multimorbidity (to reduce potentially preventable hospitalisations).

Blended payment models (a mix of FFS and CAPs) have recently gained prominence in several health settings to meet the objectives discussed above. Their aim is to combine a general overall payment with specific incremental targeted payments, thereby incentivising general practices or physicians to provide specific types of service delivery. For example, a CAP can be combined with a performance-based payment structure where general practices/physicians get paid extra for meeting a preset criteria linked to patient outcomes. This can be useful for preventive care objectives such as regular cancer screenings.

The main aim of this section is to review evidence of the effectiveness of such blended payment models in changing general practice or physician behaviour and achieving patient outcomes. Further, any policy with such blended payment structure can be very costly from the funder's perspective, as it requires budget allocation for incentive payments. Thus, an analysis of comparing the cost of blended payment models to the potential benefit is critical. As a secondary objective, we discuss potential methods for economic evaluation of blended payment models and review select studies that have evaluated such models in a real setting.

Results

Overall description

The focus of this review is on the impact of switching to blended models from FFS or CAP only payment systems, rather than comparing FFS and CAP models. These reforms involved tweaks to practice/physician payments in stages, and enabled us to isolate and quantify the incremental effects of changes in payments via the econometric modelling framework. The reforms themselves were triggered by the need to provide continuity of care to patients, reduce waiting times to see a physician and make

primary care more desirable for medical graduates. In the early 2000s, the Ontario government, in consultation with the Ontario Medical Association, developed several payment models as an alternative to the traditional FFS model. Practices were allowed to voluntarily switch from the FFS only model to these blended models. For example, one model involved CAPs only if patients were formally rostered with the practice, so as to ensure regular care. Physicians moving to blended models were encouraged to practice in groups and provide 24/7 access. They were given P4P incentives for services such as diabetes management. A large number of physicians switched from FFS only payment models to these blended models, and this came at a cost to the government. For example, payments to physicians increased by 58% (after adjusting for inflation) between 2003 and 2009.

Main findings

Zhang and Sweetman (2018) analysed responses of physicians to one such blended model, where physicians were paid on a capitation basis for some fee codes (services which were part of the capitation basket) and paid on an FFS basis for services outside the capitation basket.³⁹ The main result showed patients experienced a nine to 14% reduction in capitated services and a simultaneous increase of ten to 22% in FFS services from their rostering physicians (changes statistically significant p < 0.05). This indicated a behavioural change from physicians and a preference to provide more services through FFS fee codes. A study by Ding and Liu (2021) further confirmed this behaviour by focusing on treatment decisions of physicians among patients experiencing lower back pain.⁴⁰ They found that patients under the capitated model were 2.5% less likely to use any therapy and 4% less likely to have diagnostic tests. Also these patients received almost 30% less treatment in the diagnostic testing services and had 17% less therapy treatment overall. In contrast, these differences did not exist among more severe patients requiring surgeries. Thus, CAPs played a critical role in reducing usage of services. One argument could be that such reduction in use of services may be a positive outcome if these services were not required or were unnecessary.

Chami and Sweetman (2019) focused on this issue by analysing the changes in ordering diabetic related laboratory tests once physicians switch to a blended model.⁴¹ They found that after joining the blended model, physicians ordered 3% fewer inappropriate/redundant tests, and if the patient was treated by an interdisciplinary team, the reduction was about 9%. Furthermore, there was no decline in testing levels for HbA1c and authors claimed that these trends reflected improvement in continuity of care. However, the share of CAPs in blended models might also make a difference to service provision. A study from Denmark used a natural experiment where the share of CAPs in a blended model changed significantly from 20% to 80% (p < 0.001).⁴⁴ The study found that such a switch resulted in an overall reduction in provision of services related to quality of care among type 2 diabetic patients, particularly those with multiple co-morbidities. Similar trends were reported by Rudoler et al. (2015), who explored, using data from Ontario, the factors associated with primary care physicians self-selecting into payment models.⁴⁸ They found that physicians with more complex patient populations were less likely to switch into capitation-based payment models if higher levels of effort were not financially rewarded.

In Ontario, some blended models explicitly provided financial incentives for after-hours care (20% premium of FFS fees) and comprehensive care (10% premium of FFS fees). The impact of such incentives on service provision was the main focus of a study by Somé et al. (2019), where they found that

compared to the FFS only model, the blended model increased comprehensive care, after-hours care and non-incentivised services by 3%, 15% and 4% per annum respectively.⁴² They also found that blended FFS physicians provided more services by working additional days as well as a number of days during holidays and weekends. A switch between blended FFS and blended CAPs models, where both these models incentivised after hours and comprehensive care but differed in payment structure (i.e. blended capitation allowed CAPs adjusted for age and sex of enrolled patients, whereas blended FFS had a fee for service payment along with incentives), was analysed by Somé et al. (2020).⁴³ They found that switching from blended FFS to blended CAPs reduced the production of capitated services to enrolled patients and non-enrolled patients by 15% and 5% per annum respectively, and increased the production of after-hours and non-incentivised services by 8% and 15% per annum respectively. Thus, though blended CAPs reduced usage of services overall, services linked to financial incentives still showed an increase in the capitation models. One implication of this result is that if preventative care is linked to financial incentives, their provision should not decrease even in a blended capitation model.

Economic evaluation

It is imperative that any blended model that includes financial incentives, such as P4P, only makes payments for evidence-based quality improvement initiatives based on specific indicators. For example, an incentive payment based on a hypertension indicator can be made to a general practice if the percentage of patients under 80 years old with hypertension, and in whom the last recorded blood pressure (measured in the preceding nine months) was 140/90 or less, was above a certain threshold. Further, such payments needed to be cost effective, i.e. the monetised health benefit from such an incentive structure net of delivery cost and the performance payments should be positive. Thus, one key criterion to choose indicators for P4P is their feasibility to be evaluated for cost effectiveness. This is possible only if these indicators lead to a specific treatment or therapy, have clinically significant quantifying outcomes (e.g. reduction in blood pressure in hypertension patients) and are likely to have robust data on costs and benefits (such as lower cholesterol levels in diabetic individuals).

The National Institute for Health and Care Excellence (NICE) has developed a set of indicators for QOF and has provided guidelines on cost effectiveness that help the development of new potential indicators.⁴⁷ As discussed above, the key to cost effectiveness is the collection of cost and benefits data for each indicator. Main costs of an indicator could be costs associated with additional GP and nurse consultations for regular monitoring of patients' health status, initial setup costs of implementing the intervention, and any additional usage of services such as diagnostic services. The health benefits could be quantified based on the change in QALYs post intervention compared to the standard practice without the indicator. An intervention or indicator is considered as cost effective by NICE if it generates QALYs valued between GBP20,000 and GBP30,000.

Pandya et al. (2018) undertook cost effectiveness analysis of QOF using a lifetime simulation model to estimate QALYs and costs for a UK population cohort aged 40 to 74 years (n = 27,070,862) exposed to the QOF and for a counterfactual scenario without exposure.⁴⁹ They found that continuing QOF was not cost effective as it increased population-level QALYs and health-care costs with one additional QALY costing around GBP50,000. They also simulated that for QOF to be cost effective, the performance payments should be reduced by 36% and a substantial redesign of the program was recommended. On

the other hand, evaluation of P4P payments in the Canadian province of British Columbia, found that incentive payments might reduce overall costs and hospitalisations.⁴⁵ The study found that after controlling for patient age, sex, service needs level and continuity of care (defined as attachment to a general practice), the incentives reduced the net annual health care costs for patients with hypertension by approximately CAD308 per patient, for chronic obstructive pulmonary disease by CAD496 per patient and for congestive heart disease by CAD96 per patients, however costs for diabetes incentives increased by about CAD148 per patient. The incentives were also associated with fewer hospital days, fewer admissions and readmissions, and shorter lengths of hospital stays for all four groups.

It is imperative to undertake cost effectiveness analysis for P4P payments before including them in the blended payment models, as not all indicators might be cost effective. Another alternative is to have a ceiling on these payments to ensure cost effectiveness. A study from Taiwan found that a P4P programme for caring for patients with only diabetes had a return on investment of 2.60:1, and this was even higher in patients with diabetes and hypertension (3.48:1).⁴⁶

Key findings

- 1. There was insufficient evidence due to the limited number of relevant studies in this area.
- 2. Switching from FFS only to blended models might lead to general practices or physicians self-selecting these models.
- 3. Even in blended models, general practices and physicians focused more on the FFS component and could potentially increase services related to this payment structure. Any P4P incentives offered under FFS might be more successful in meeting physician objectives.
- 4. CAPs were related to lower service usage and could lead to a reduction in unnecessary services for low complexity patients.
- 5. An economic evaluation of P4P services is crucial.
- 6. Not all P4P programs were cost effective. An indicator level cost effectiveness analysis is recommended.
- 7. Implementation of blended models could prove very costly to the funder. Thus, an analysis of the overall benefits of such a model is critical pre-implementation.

Limitations

There were a limited number of studies that met this review's criteria, and almost all of them used payment reform implemented in Ontario, Canada as a setting. Most of the above studies used administrative data that lacked information on duration and complexity of GP visits, and health attitudes and behaviour of patients. Further, they included superficial information on patient demographic characteristics, socioeconomic status, geographic locations and healthcare needs. These were important unobservable factors critical in capturing the effect of P4P incentives on patient outcomes. Most of the studies used cross-sectional data and lacked randomisation. Though the propensity score matching approach was used as an alternative by the majority of studies, the results could not be interpreted as causal inference and ideally longitudinal data was required for robust causal inference. Due to the administrative burden, CAP billing was often incomplete (and not as accurate) as FFS billing, and this might be reflected in the under or overestimation of quantitative effects of that intervention. Most of the reviewed studies used data from Ontario, with most practices located in metro regions and working as large groups of interdisciplinary teams. These results thus might not be generalisable beyond such a setting, such as rural and remote areas with low GP density.

Conclusions

There was limited literature on the effectiveness of switching to blended models and their specific longterm effects on patient outcomes (e.g. access to care in a cost-effective manner). There was no evidence of a significant reduction in waiting times to see physicians or of reduced avoidable emergency department presentations after switching to blended models. This was further exacerbated by reduced billing once practices switched to CAP (to minimise administrative burden), and this made the tracking of service delivery very difficult. On the other hand, the majority of studies agreed on physicians' responses to blended models, and there was a clear preference for FFS based services even in a blended setting that could lead to reduction in capitation-based services. One implication is that physicians are more likely to respond to FFS-based P4P incentives, and any payments for preventative care indicators (e.g. cancer screening) are more likely to be effective if they are offered on an FFS basis.

From a funder's perspective, a switch to blended models could be very costly (costs doubled in Ontario in few years), indicating that a cost effectiveness analysis is imperative before implementation. Experience from QOF in UK also suggested that not all P4P incentives were cost effective, and thus an economic evaluation of each P4P incentive is desirable.

RQ4: Evidence from the literature on what drives behavioural change in primary care providers

This question focuses on change in behaviour of providers as opposed to change in behaviour of consumers, although the two are inevitably closely linked. Some additional drivers of behavioural change are briefly discussed towards the end of this section, but these were outside the scope of the literature reviewed.

Blended funding models aim to incentivise patient-centred care and improve patient health by encouraging better outcomes or discouraging poor practices and outcomes.² Scott et al. (2011) stated that financial payments can affect the behaviour of health providers both positively and negatively, and they found that there was insufficient evidence for the use of financial incentives to improve the quality of primary health care.³ Whilst several different changes in professional behaviour were reported in this review, there were no specific findings about what drove or motivated primary care providers towards behavioural change.^{3,68}

The evidence for the cost-effectiveness of financial incentives in relation to other behavioural change strategies was also insufficient.⁸¹ In the literature review on the impact of bundled payments^a in primary health care, Dawda (2015) identified some mechanisms which led to improvements in quality and costs savings.² These included adherence to protocols, service redesign and increased care coordination. However, these mechanisms referred to clinician behaviours rather than what motivated clinicians' behaviour change. Education and quality improvement have often been described as important for changing provider behaviour,⁸² but these were not the focus of the studies included in this review.

The effect of the context in any policy change is critical, and the context in which financial incentives are delivered on the behavioural responses of primary care providers is also important.³ In their commentary on financial incentives and the health workforce, Scott and Connelly (2011) stated that health care providers were motivated by a variety of factors, and that the effectiveness of incentives in achieving behavioural change depended on the relative weight given to these factors in different contexts.⁸¹ Sources of physician motivation can vary in different settings and be affected by individual preferences and practice styles.³ Spooner et al. (2001) found that the alignment of clinical and managerial agendas, and management priorities, with the values of health professionals, played a role in promoting behavioural change amongst primary care providers participating in a quality improvement program.⁸³

^a 'Bundled payment' covers a bundle of services including: (1) Payment for services, which are aggregated longitudinally. For example, it might include the pre-hospital elements of an elective procedure, the elective procedure itself and the post-hospital care elements for that procedure such as rehabilitation. (2) The pooling of funds for disparate group of providers. This, for example, will often include all the medical specialists required to deliver an episode of care. (3) The incorporation of a warranty e.g. includes the management of complications from a procedure.²

Results

Overall description

Relatively few of the included studies focused on behaviour change in primary care providers as a primary outcome. Instead, many focused on whether or not the behaviours of patients had changed. The studies included showed limited and mixed results in terms of their impact on provider behaviours.

A range of financial incentives for providers were used, but these yielded mixed results in terms of changing provider behaviour. The included studies predominantly described P4P measures including service targets and reimbursement rate adjustments.^{17,25,33,38} Despite limited and mixed evidence that P4P measures were effective, there was an even greater lack of certainty about:

- the level of payments required to affect provider behaviour (dose-response relationships)
- the mechanisms by which they should be administered.

This represented a critical evidence gap.²⁵ The following factors were described as enhancing the effectiveness of P4P measures in changing provider behaviour:

- perceptions about the adequacy of payments
- aligned education and training
- stakeholder involvement in design and delivery
- consistency in program design and communication
- tailoring to the broader health service context.

It is important to note that these appeared to be enablers rather than essential elements, and there was insufficient evidence to support these factors. Further, enablers such as the need to tailor P4P programs may have been in tension with the need for consistency in how these were designed and delivered across entire systems. CAPs had a mixed impact on behaviours that drove change with regard to quality and patient centredness.

Main findings

Several of the included studies looked at interprofessional and interdisciplinary care team changes. In general, these studies emphasised the need for interprofessional care teams to:

- Ensure role differentiation and clarity within teams, especially for any non-clinical team members
- Promote co-location, shared space and shared equipment, which were identified as factors that enhanced collaborative behaviours¹⁶ (N.B. insufficient evidence in support).

Most of the studies included a combination of interventions. The specific mix of interventions required to change provider behaviour remained unknown and were usually described as being developed or selected in response to contextual factors including system financing arrangements, population needs, workforce considerations and existing interventions and reforms.

Hamilton et al. (2013) reviewed 18 studies up to 2011 on the impact of P4P financial incentives for smoking cessation in the context of primary care teams.⁸⁴ This was a high RoB review. Ten were based in the UK, evaluating impacts of the introduction of the QOF in primary care, with the remainder in Germany, Taiwan and the US, examining FFS and bonus payments. Bonus payments (some of which

were per smoker) were for recording or referring patients, but in two studies it was awarded for each patient who stopped smoking. The studies showed improvements in processes (e.g. recording, providing advice or referral) but not in smoking cessation. There seemed to be synergistic effects when financial incentives were combined with GP training and free medication.

Key considerations: Factors affecting behaviour change

The literature emphasised that it was difficult to affect provider behaviour independent of the context in which the change occurred. Figure 5 synthesises the factors described in the reviewed literature that may impact the successful implementation of behavioural change interventions.



Figure 5: Key factors for successful behavioural change interventions with primary care providers.

The broad range of factors identified have been sorted into change characteristics, change processes, contextual factors (personal, inner setting and outer setting), impacts and outcomes. This approach draws on the Consolidated Framework for Implementation Research and the Framework for Partnership Working in Public Policy Provision Evaluation.^{85–87}

Key considerations: Success factors and design

Based on the literature reviewed, there was insufficient evidence to state that any specific factors were required for provider behavioural change interventions to be effective in all contexts. However, several design considerations were identified across several studies and reviews, including:

- **Sustaining interventions.** There was a need for changes to be sustained over the longer-term in order to bring about sustained behaviour change (insufficient evidence).
- **Influence rather than authority.** Some reviews that incorporated qualitative studies described the importance of convincing practitioners in order to change behaviour, rather than relying on authority to drive change (insufficient evidence).
- Behaviour change impacting process versus outcome. The literature described examples of positive changes in process measures related to behaviour where sustainable improvement in outcomes were mixed or inconclusive.

- **Communication.** Communication with providers and information management/information systems were identified as an enabler of provider behaviour (insufficient evidence).

A note on change management

None of the included studies referred to change management processes, and there was insufficient evidence about preferred change management processes. There was insufficient evidence in general about the comparative effectiveness of different change management approaches, which may be partly reflective of the different difficulties in doing comparative studies in this area. The use of specific models of change management, such as Kotter (*8-step model*), PROSCI (*3-Phase Process is a framework for organizational change*) and ADKAR (*Awareness, Desire, Knowledge, Ability and Reinforcement*), appear to be widespread in healthcare settings,⁸⁸ but this may be reflective of factors such as longstanding use and commercialisation, rather than effectiveness. As such, a specific change management approach cannot be recommended.

Related interventions

A number of practices have been discussed in the general literature on behavioural change, including:

- "nudging" and "digital nudging"
- interventions (primarily educational) to enhance cultural competence, cultural safety, diversity and inclusion, and holistic care
- feedback mechanisms and system listening.

However, these were not found in the studies in this review. As such they cannot be recommended but may be raised in stakeholder consultations.

Key findings

- Insufficient evidence of the impact of blended funding on behaviour change: Blended funding models showed insufficient evidence of affecting primary health care provider behaviour with regard to patient-centred care and improved health outcomes. There was a general lack of findings in the evidence reviewed regarding the motivations and factors driving primary care providers towards behaviour change.
- Insufficient evidence on cost-effectiveness of financial incentives used to drive behaviour change: There was insufficient evidence on the cost-effectiveness of financial incentives compared to other behaviour change strategies. Various financial incentives, including P4P measures and service targets, showed inconsistent impacts on modifying provider behaviours.
- 3. Importance of interprofessional teams: Studies emphasised the need for role differentiation and co-location in interprofessional care teams to enhance collaborative behaviours, although the evidence supporting this was low quality.
- 4. Unclear drivers of behavioural change: There was limited evidence on specific success factors associated with interventions that drove provider behaviour change across all contexts. However, the literature discussed sustaining interventions over time, focusing on influence rather than directing change, and on effective communication.

Limitations

In general, few of the included studies focused on provider behavioural change as a primary outcome rather than as an enabling factor of changes in patient outcomes. The included studies were of low-to-moderate levels of quality and assessed as having moderate to high RoB.

Relatively few controlled trials or natural experiments have been undertaken, due to the interventions being primarily related to policies and financing arrangements, which doesn't lend itself to trial, pre-post or comparative study designs.³³ As such, there was generally limited information on the comparative effectiveness of specific measures.

By the nature of this review's design, reviewing systematic reviews of previous studies creates a time lag between the date of an intervention and the evidence provided in this report. This review therefore was likely to exclude any studies relating to changes in provider behaviour associated with the COVID-19 pandemic, which has offered opportunities for pre-post style reviews of various reforms in primary health care practice.

Conclusions

There was insufficient, mixed quality evidence supporting the effectiveness of P4P models in influencing provider behaviour. Broader contextual factors influenced the impact of financial incentives of provider behaviour (at a practice and at an individual level), making it difficult to differentiate the impact of interventions from the features of the settings in which they were trialled or adopted. Most specific study findings were negative or mixed, and positive impacts, where reported, were modest. Role clarity and co-location were mentioned as factors that influenced provider behaviours, but the evidence supporting this was low quality. The varied quality of the studies and reviews included meant there were few factors we can be certain drove behavioural change among primary care providers. While financial incentives may have been a factor for many general practices, how they were designed and delivered (including value, duration, processes for claiming, information campaigns and considerations for impact on other services) was likely to affect their uptake.

Discussion

This rapid review, synthesising the results of 30 prior systematic reviews found insufficient evidence on the impact of performance and blended primary care funding models on clinical outcomes, although impacts on process measures were demonstrated. Other non-financial incentives, such as training, access to medication and quality improvement support, were required to synergistically improve outcomes. There was also insufficient and conflicting evidence on the equity impacts of the various funding models. Equity is an important issue that needs to be addressed in any consideration of appropriate funding models.

This review found insufficient evidence that CAPs improved quality of care and delivery of preventative care. Instead, we found limited evidence suggesting that different funding models resulted in a shift in quantity of care provided. There was also evidence suggesting that P4P and FFS may induce providers to provide more services for those covered/incentivised.³⁴ This may be at the expense of non-incentivised care and patient centred care, as providers may prioritise achieving the required process measures rather than patient contact/care.

We note that there was more sufficient evidence on the effectiveness of incentives in changing process measures for quality of care compared to health outcomes. This may partially relate to the quality and availability of data to assess performance using these measures. The impact of incentives linked to these performance indicators was found to vary depending on the indicators used. Studies reported that measures, such as those used in the UK QOF, were more focused on processes of care, such as recording of assessments, but were not effective in improving adherence to high value care, as per guidelines, or in optimising health outcomes.^{18,31,35,36} This focus on recording process measures for incentivised care may have contributed to a reduction in the interactions between providers and patients.²¹ This suggests a need to include patient-centred measures during the implementation of new performance frameworks so as to mitigate against any unintended consequence.²¹

We found evidence that when there were performance-based and blended models, physicians were more likely to self-select to deliver services that were FFS or incentivised.³⁴ There was also evidence that performance-based funding models may positively influence their provision of preventative care, both for primary and secondary prevention, though results varied across studies, diseases and program designs.²⁹ Considering this, in blended capitation models it can be beneficial to link preventative and other high value services to incentives so as to promote their delivery, while including low value services under capitation agreements to reduce their unnecessary utilisation. It is also important to have clear criteria for inclusion to prevent distortions due to selective inclusion of practitioners and patients.³

The variability in impacts on quality and cost-effectiveness of performance-based and blended funding models between settings and within the same system for different conditions, suggests that beyond improvements in designing performance indicators and incentives, economic evaluations need to be undertaken for each performance indicator introduced, especially prior to full implementation. Main costs of an indicator could be costs associated with additional GP and nurse consultations for regular monitoring of patients' health status, initial setup costs of implementing the intervention, and any additional usage of services such as diagnostic services.

For example, in Canada, adjusting for demographic and practice level variables per annum, incentives minimised net healthcare costs for hypertension and heart failure but not diabetes, despite reducing
patient hospital days, hospital admissions and readmissions for all these conditions.⁴⁹ Whereas a different performance-based funding model implementation for diabetes in Taiwan reported a significant (p < 0.001) return on investment.⁴⁶ These differences in impact may stem from the variation in health system profile, performance indicators, incentives and other design features of the programs. To support cost-effectiveness, fit for purpose indicators that lead to a specific treatment or therapy resulting in clinically significant, quantifiable outcomes, such as reduction in blood pressure, are needed. Otherwise, implementation of performance based and blended models could be very costly for funders, as was reported by a modelling study of the UK QOF.⁴⁹ Thus making it difficult or impossible to determine whether improvements as well as cost minimisations are due to incentives or other health system factors.

While there was evidence that UK's QOF reduced the gap in quality of care between the most and least deprived areas,²¹ there was insufficient evidence on the impact of performance related pay or CAPs on health equity. There was limited evidence for ceiling effects due to thresholds set by performance indicators, e.g. delivery of a particular service to between 50 to 90% of the eligible cohort. Though in return, raising thresholds to 100% may unfairly disadvantage facilities in deprived areas, as they would less likely reach targets.¹⁸ There is also a need for more studies reporting disaggregated results by gender and other equity groups (e.g. socioeconomic status, rurality).²³

While this review did not specifically examine the impact of various funding models on different members of the multidisciplinary team or on workforce supply and arrangements, historically the introduction of the UK QOF and the Ontario reforms were designed to increase payments received by GPs to maintain declining workforce numbers. There was some evidence that when performance related payments or CAPs were implemented at a practice level, there was greater collaboration between primary care providers and other disciplines. This also provided support for reform at the team level beyond a focus on only service provision as in FFS. Linking incentives to providers or teams as well as to other capacity development activities, such as training and quality improvement, may enhance the impact of these incentives. However, there was insufficient evidence on what drives behavioural change across primary care providers. Specifically, the level of payments required to affect provider behaviour (dose-response relationships) and the mechanisms by which they should be administered. Modelling should be undertaken to ascertain the impact of any shifts in funding models and overall investment in healthcare on workforce supply and interactions.

Recommendations to improve the development of indicators, especially when implementing new performance-based and blended funding models, include:

- establish the current baseline
- co-design indicators with multiple stakeholders
- implement indicators that are uniform across payers
- build support for quality improvement for both low and high performing providers; and
- distribute incentives at the individual provider and/or team level rather than higher up in the organisation.²⁵
- indicators need to be re-evaluated periodically as the program progresses (especially when ceiling
 effects occur a period of time after a new program is implemented^{18,21,25}) so as to confirm whether
 the improvements early on are a result of better reporting or a true increase in delivery of services.

Conclusions

- There was insufficient evidence that a shift to include more performance-based incentives in blended primary care funding models would result in improvements in health outcomes and minimise costs.
- There was sufficient evidence that linking incentives to not only performance indicators but also non-financial incentives, such as quality improvement programs, may synergistically improve outcomes.
- A key piece to implementation for performance-based or blended models, is the development of fit for purpose indicators that combine process and health outcome measures.
- A careful evaluation of each individual performance measure needs to be undertaken. This
 evaluation should be based on how the effects vary according to context, scope for improvement,
 benefit and cost of the measure, and should include both positive and negative effects.
- This review's findings support the proposition that to influence provider behaviours, incentives need to link to individual providers or teams, rather than being implemented only at an organisational level.
- In blended models, it can be beneficial to include low value services under the capitation agreement in order to reduce their unnecessary utilisation, and to link preventative and other high value services to incentives so as to promote their delivery. Indicators set should also include thresholds that account for a services patient mix.
- While various blended funding models have been implemented in other countries, it is important to tailor them for Australia because of the variability of impact for the same models across different settings.

References

- Oliver-Baxter, J. Blended funding models in primary health care. *PHCRIS RESEARCH ROUNDup* (2015).
- 2. Dawda, P. Bundled payments: Their role in Australian primary health care. (2015).
- 3. Scott, A. *et al.* The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrane Database Syst. Rev.* (2011) doi:10.1002/14651858.CD008451.pub2.
- 4. Donabedian, A. Evaluating the Quality of Medical Care. *Milbank Q.* 83, 691–729 (2005).
- 5. WHO. Quality of care. https://www.who.int/health-topics/quality-of-care#tab=tab_1 (2024).
- Page, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* n71 (2021) doi:10.1136/bmj.n71.
- Bristol Medical School: Population Health Sciences. ROBIS tool.
 https://www.bristol.ac.uk/population-health-sciences/projects/robis/robis-tool/.
- Whiting, P. *et al.* ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *J. Clin. Epidemiol.* 69, 225–234 (2016).
- Bowling, B., Newman, D., White, C., Wood, A. & Coustasse, A. Provider Reimbursement Following the Affordable Care Act. *Health Care Manag.* 37, 129–135 (2018).
- 10. Emmert, M., Eijkenaar, F., Kemter, H., Esslinger, A. S. & Schöffski, O. Economic evaluation of pay-forperformance in health care: a systematic review. *Eur. J. Health Econ.* **13**, 755–767 (2012).
- 11. Jackson, T. *et al.* The impact of financial incentives on the implementation of asthma or diabetes self-management: A systematic review. *PLOS ONE* **12**, e0187478 (2017).
- Markovitz, A. A. & Ryan, A. M. Pay-for-Performance: Disappointing Results or Masked Heterogeneity? *Med. Care Res. Rev.* 74, 3–78 (2017).

- Vahidi, R. G., Mojahed, F., Jafarabadi, M. A., Gholipour, K. & Rasi, V. A Systematic Review of the Effect of Payment Mechanisms on Family Physicians Service Provision and Referral Rate Behavior. *JPMS* 3, 54–60 (2013).
- 14. Lin, Y., Yin, S., Huang, J. & Du, L. Impact of pay for performance on behavior of primary care physicians and patient outcomes. *J. Evid.-Based Med.* **9**, 8–23 (2016).
- Tildy, B. E., McNeill, A., Perman-Howe, P. R. & Brose, L. S. Implementation strategies to increase smoking cessation treatment provision in primary care: a systematic review of observational studies. *BMC Prim. Care* 24, 32 (2023).
- Wranik, W. D. *et al.* Implications of interprofessional primary care team characteristics for health services and patient health outcomes: A systematic review with narrative synthesis. *Health Policy* 123, 550–563 (2019).
- Mauro, M., Rotundo, G. & Giancotti, M. Effect of financial incentives on breast, cervical and colorectal cancer screening delivery rates: Results from a systematic literature review. *Health Policy* 123, 1210–1220 (2019).
- Langdown, C. & Peckham, S. The use of financial incentives to help improve health outcomes: is the quality and outcomes framework fit for purpose? A systematic review. *J. Public Health* 36, 251–258 (2014).
- 19. Tao, W., Agerholm, J. & Burström, B. The impact of reimbursement systems on equity in access and quality of primary care: A systematic literature review. *BMC Health Serv. Res.* **16**, 542 (2016).
- 20. Mendelson, A. *et al.* The Effects of Pay-for-Performance Programs on Health, Health Care Use, and Processes of Care: A Systematic Review. *Ann. Intern. Med.* **166**, 341 (2017).
- 21. Ahmed, K. *et al.* What drives general practitioners in the UK to improve the quality of care? A systematic literature review. *BMJ Open Qual.* **10**, e001127 (2021).

- 22. Bes, Flinterman, GonzÃilez, & Batenburg. Recruitment and retention of general practitioners in European medical deserts: a systematic review. *Rural Remote Health* (2023) doi:10.22605/RRH7477.
- Boeckxstaens, P., Smedt, D. D., Maeseneer, J. D., Annemans, L. & Willems, S. The equity dimension in evaluations of the quality and outcomes framework: A systematic review. *BMC Health Serv. Res.* 11, 209 (2011).
- 24. Brocklehurst, P. *et al.* The effect of different methods of remuneration on the behaviour of primary care dentists. *Cochrane Database Syst. Rev.* (2013) doi:10.1002/14651858.CD009853.pub2.
- 25. Van Herck, P. *et al.* Systematic review: Effects, design choices, and context of pay-for-performance in health care. *BMC Health Serv. Res.* **10**, 247 (2010).
- 26. Houle, S. K. D., McAlister, F. A., Jackevicius, C. A., Chuck, A. W. & Tsuyuki, R. T. Does Performance-Based Remuneration for Individual Health Care Practitioners Affect Patient Care?: A Systematic Review. *Ann. Intern. Med.* **157**, 889 (2012).
- 27. Gillam, S. J., Siriwardena, A. N. & Steel, N. Pay-for-Performance in the United Kingdom: Impact of the Quality and Outcomes Framework--A Systematic Review. *Ann. Fam. Med.* **10**, 461–468 (2012).
- Carter, R., Riverin, B., Levesque, J.-F., Gariepy, G. & Quesnel-Vallée, A. The impact of primary care reform on health system performance in Canada: a systematic review. *BMC Health Serv. Res.* 16, 324 (2016).
- Hamilton, F. L., Greaves, F., Majeed, A. & Millett, C. Effectiveness of providing financial incentives to healthcare professionals for smoking cessation activities: systematic review. *Tob. Control* 22, 3–8 (2013).
- 30. Mandavia, R., Mehta, N., Schilder, A. & Mossialos, E. Effectiveness of UK provider financial incentives on quality of care: a systematic review. *Br. J. Gen. Pract.* **67**, e800–e815 (2017).

- Ogundeji, Y. K., Bland, J. M. & Sheldon, T. A. The effectiveness of payment for performance in health care: A meta-analysis and exploration of variation in outcomes. *Health Policy* **120**, 1141–1150 (2016).
- 32. Yuan, B., He, L., Meng, Q. & Jia, L. Payment methods for outpatient care facilities. *Cochrane Database Syst. Rev.* **2017**, (2017).
- 33. Rashidian, A., Omidvari, A.-H., Vali, Y., Sturm, H. & Oxman, A. D. Pharmaceutical policies: effects of financial incentives for prescribers. *Cochrane Database Syst. Rev.* **2015**, (2015).
- 34. Jia L, Meng Q, Scott A, Yuan B, & Zhang L. Payment methods for healthcare providers working in outpatient healthcare settings. *Cochrane Database Syst Rev* **1**, CD011865 (2021).
- Flodgren, G. *et al.* An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. *Cochrane Database Syst. Rev.* (2011) doi:10.1002/14651858.CD009255.
- Heider, A.-K. & Mang, H. Effects of Monetary Incentives in Physician Groups: A Systematic Review of Reviews. *Appl. Health Econ. Health Policy.* 18, 655–667 (2020).
- Scott, A., Liu, M. & Yong, J. Financial Incentives to Encourage Value-Based Health Care. *Med. Care Res. Rev.* 75, 3–32 (2018).
- 38. Khan, N., Rudoler, D., McDiarmid, M. & Peckham, S. A pay for performance scheme in primary care: Meta-synthesis of qualitative studies on the provider experiences of the quality and outcomes framework in the UK. *BMC Fam. Pract.* **21**, 142 (2020).
- 39. Zhang, X. & Sweetman, A. Blended capitation and incentives: Fee codes inside and outside the capitated basket. *J. Health Econ.* **60**, 16–29 (2018).
- 40. Ding, Y. & Liu, C. Alternative payment models and physician treatment decisions: Evidence from lower back pain. *J. Health Econ.* **80**, 102548 (2021).

- 41. Chami, N. & Sweetman, A. Payment models in primary health care: A driver of the quantity and quality of medical laboratory utilization. *Health Econ.* **28**, 1166–1178 (2019).
- Somé, N. H. *et al.* Production of physician services under fee-for-service and blended fee-for-service: Evidence from Ontario, Canada. *Health Econ.* 28, 1418–1434 (2019).
- Somé, N. H., Devlin, R. A., Mehta, N., Zaric, G. S. & Sarma, S. Stirring the pot: Switching from blended fee-for-service to blended capitation models of physician remuneration. *Health Econ.* 29, 1435– 1455 (2020).
- Skovsgaard, C. V., Kristensen, T., Pulleyblank, R. & Olsen, K. R. Increasing capitation in mixed remuneration schemes: Effects on service provision and process quality of care. *Health Econ.* 32, 2477–2498 (2023).
- 45. Hollander, M. J. & Kadlec, H. Incentive-Based Primary Care: Cost and Utilization Analysis. *Perm. J.* **19**, 46–56 (2015).
- 46. Hsieh, H.-M. *et al.* Cost-Effectiveness of a Diabetes Pay-For-Performance Program in Diabetes Patients with Multiple Chronic Conditions. *PLOS ONE* **10**, e0133163 (2015).
- Qureshi, N., Weng, S. & Hex, N. The role of cost-effectiveness analysis in the development of indicators to support incentive-based behaviour in primary care in England. *J. Health Serv. Res. Policy* 21, 263–271 (2016).
- Rudoler, D. *et al.* Paying for Primary Care: The Factors Associated with Physician Self-selection into Payment Models. *Health Econ.* 24, 1229–1242 (2015).
- Pandya, A. *et al.* Modelling the cost-effectiveness of pay-for-performance in primary care in the UK.
 BMC Med. 16, 135 (2018).
- 50. Bitton, A. *et al.* Off the Hamster Wheel? Qualitative Evaluation of a Payment-Linked Patient-Centered Medical Home (PCMH) Pilot. *Milbank Q.* **90**, 484–515 (2012).

- 51. Petersen, L. A. *et al.* Effects of Individual Physician-Level and Practice-Level Financial Incentives on Hypertension Care: A Randomized Trial. *JAMA* **310**, 1042 (2013).
- 52. Mitchell, G. K., Tieman, J. J. & Shelby-James, T. M. Multidisciplinary care planning and teamwork in primary care. *Med. J. Aust.* **188**, (2008).
- Hutchison, B. & Glazier, R. Ontario's Primary Care Reforms Have Transformed The Local Care Landscape, But A Plan Is Needed For Ongoing Improvement. *Health Aff. (Millwood)* 32, 695–703 (2013).
- Hutchison, B. & Glazier, R. Ontario's Primary Care Reforms Have Transformed The Local Care Landscape, But A Plan Is Needed For Ongoing Improvement. *Health Aff. (Millwood)* 32, 695–703 (2013).
- 55. Martin Roland, D. Linking physicians' pay to the quality of care—a major experiment in the United Kingdom. *N Engl J Med* **351**, 1448–54 (2004).
- 56. Doran, T. *et al.* Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. *Bmj* **342**, (2011).
- 57. New Zealand Government. Cabinet material: Achieving Pae Ora through primary care. (2023).
- 58. Building Primary Care in a Changing Europe. (European Observatory on Health Systems and Policies, Copenhagen (Denmark), 2015).
- 59. Shi, L. & Singh, D. Essentials of the US Health Care System. (Jones & Bartlett Learning, 2022).
- Tikkanen, R., Osborn, R., Mossialos, E., Djordjevic, A. & Wharton, G. International health Care System Profiles: United States June 2020, The Commonwealth Fund. https://www.commonwealthfund.org/international-health-policy-center/countries/united-states (2020).
- 61. Gillam SJ, Siriwardena AN, & Steel N. Pay-for-performance in the United Kingdom: impact of the quality and outcomes framework: a systematic review. *Ann Fam Med* **10**, 461–8 (2012).

- 62. Mandavia R, Mehta N, Schilder A, & Mossialos E. Effectiveness of UK provider financial incentives on quality of care: a systematic review. *Br J Gen Pr.* **67**, e800–e815 (2017).
- 63. Ahmed K *et al.* What drives general practitioners in the UK to improve the quality of care? A systematic literature review. *BMJ Open Qual* **10**, (2021).
- 64. Carter R, Riverin B, Levesque JF, Gariepy G, & Quesnel-Vallée A. The impact of primary care reform on health system performance in Canada: a systematic review. *BMC Health Serv Res* **16**, 324 (2016).
- 65. Scott, A., Liu, M. & Yong, J. Financial incentives to encourage value-based health care. *Med. Care Res. Rev.* **75**, 3–32 (2018).
- 66. Tao W, Agerholm J, & Burström B. The impact of reimbursement systems on equity in access and quality of primary care: A systematic literature review. *BMC Health Serv Res* **16**, 542 (2016).
- 67. Van Herck P *et al.* Systematic review: Effects, design choices, and context of pay-for-performance in health care. *BMC Health Serv Res* **10**, 247 (2010).
- Flodgren, G. *et al.* An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. *COCHRANE DATABASE Syst. Rev.* (2011) doi:10.1002/14651858.CD009255.
- 69. Heider A.-K. & Mang H. Effects of Monetary Incentives in Physician Groups: A Systematic Review of Reviews. *Appl. Health Econ. Health Policy.* **18**, 655–667 (2020).
- 70. Australian Bureau of Statistics. National Health Survey.
 https://www.abs.gov.au/statistics/health/health-conditions-and-risks/national-health-survey/2022
 (2022).
- Van Den Akker, M., Buntinx, F. & Knottnerus, J. A. Comorbidity or multimorbidity: what's in a name?
 A review of literature. *Eur. J. Gen. Pract.* 2, 65–70 (1996).
- 72. Australian Institute of Health and Welfare. *Chronic Condition Multimorbidity*. https://www.aihw.gov.au/reports/chronic-disease/chronic-condition-multimorbidity.

- Hajat, C. & Stein, E. The global burden of multiple chronic conditions: A narrative review. *Prev. Med. Rep.* 12, 284–293 (2018).
- 74. Kanesarajah, J., Waller, M., Whitty, J. A. & Mishra, G. D. Multimorbidity and quality of life at midlife: a systematic review of general population studies. *Maturitas* **109**, 53–62 (2018).
- 75. Nunes, B. P., Flores, T. R., Mielke, G. I., Thumé, E. & Facchini, L. A. Multimorbidity and mortality in older adults: a systematic review and meta-analysis. *Arch. Gerontol. Geriatr.* **67**, 130–138 (2016).
- 76. The Royal Australian College of General Practitioners. *General Practice: Health of the Nation 2021*.(2021).
- Ukhanova, M. A. *et al.* Uptake of Preventive Services Among Patients With and Without Multimorbidity. *Am. J. Prev. Med.* 59, 621–629 (2020).
- Cassell, A. *et al.* The epidemiology of multimorbidity in primary care: a retrospective cohort study.
 Br. J. Gen. Pract. 68, e245–e251 (2018).
- 79. Dixon, A. *et al.* The Quality and Outcomes Framework: does it reduce health inequalities. *Final Rep. NIHR Serv. Deliv. Organ. Programme* (2010).
- Meyers, D. J. *et al.* Association of Team-Based Primary Care With Health Care Utilization and Costs Among Chronically III Patients. *JAMA Intern. Med.* **179**, 54–61 (2019).
- Scott, A. & Connelly, L. B. Financial incentives and the health workforce. *Aust. Health Rev.* 35, 273–277 (2011).
- Gupta, B., Li, D., Dong, P. & Acri, M. C. From intention to action: A systematic literature review of provider behaviour change-focused interventions in physical health and behavioural health settings.
 J. Eval. Clin. Pract. 27, 1429–1445 (2021).
- 83. Spooner, A., Chapple, A. & Roland, M. What makes British general practitioners take part in a quality improvement scheme? *J. Health Serv. Res. Policy* **6**, 145–150 (2001).

- Hamilton FL, Greaves F, Majeed A, & Millett C. Effectiveness of providing financial incentives to healthcare professionals for smoking cessation activities: systematic review. *Tob Control* 22, 3–8 (2013).
- Asthana, S., Richardson, S. & Halliday, J. Partnership Working in Public Policy Provision: A Framework for Evaluation. *Soc. Policy Adm.* 36, 780–795 (2002).
- 86. Breimaier, H. E., Heckemann, B., Halfens, R. J. G. & Lohrmann, C. The Consolidated Framework for Implementation Research (CFIR): a useful theoretical framework for guiding and evaluating a guideline implementation process in a hospital-based nursing practice. *BMC Nurs.* 14, 43 (2015).
- 87. CFIR Research Team. The Consolidated Framework for Implementation Research. *Center for Clinical Management Research* https://cfirguide.org/ (2024).
- Harrison, R. *et al.* Where do models for change management, improvement and implementation meet? A systematic review of the applications of change management models in healthcare. *J. Healthc. Leadersh.* 85–108 (2021) doi:10.2147/JHL.S289176.
- 89. Jia, L., Meng, Q., Scott, A., Yuan, B. & Zhang, L. Payment methods for healthcare providers working in outpatient healthcare settings. *Cochrane Database Syst. Rev.* **2021**, (2021).

Appendices

Appendix 1: Database search strategies

CINAHL

No	Query	
S72	S71 and (S70 OR S50 OR S15)	
S71	(S57 OR TI physician*) AND (S65 OR S58)	
S70	S1 or S2 or S3 or S4 or S5 or S6 or S7 or S8 or S9 or S10 or S66 or S67 or S68 or S69	
S69	TI (("follow up" N2 design) or ("follow up" N2 study) or ("follow up" N2 studies) or (crossover N2 design) or (crossover N2 study) or (crossover N2 studies) or (crossover N2 trials) or (cross-over N2 design) or (cross-over N2 study) or (cross-over N2 studies) or (cross-over N2 trials)) or AB (("follow up" N2 design) or ("follow up" N2 study) or ("follow up" N2 studies) or (crossover N2 trials) or (crossover N2 study) or (crossover N2 study) or (crossover N2 trials) or (crossover N2 design) or (crossover N2 study) or (crossover N2 studies) or (crossover N2 design) or (crossover N2 study) or (crossover N2 trials) or (cross-over N2 trials) or (cross-over N2 study) or (crossover N2 study) or (crossover N2 trials) or (cross-over N2 study) or (cross-over N2 study) or (cross-over N2 trials) or (cross-over N2 study) or (cross-over N2 stu	
S68	TI ((multicase N2 design*) or (multicase N2 study) or (multicase N2 trial*) or (multi-case N2 design*) or (multi-case N2 study) or (multi-case N2 trial*)) or AB ((multicase N2 design*) or (multicase N2 study) or (multi-case N2 trial*) or (multi-case N2 design*) or (multi-case N2 study) or (multi-case N2 trial*)	
S67	TI (("case control " N2 design*) or ("case control " N2 study) or ("case control " N2 trial*)) or AB (("case control " N2 design*) or ("case control " N2 study) or ("case control " N2 trial*))	
S66	TI ((cross-sectional N2 design) or (cross-sectional N2 study) or (cross-sectional N2 trial)) or AB ((cross- sectional N2 design) or (cross-sectional N2 study) or (cross-sectional N2 trial))	
S65	S59 or S60 or S61 or S62 or S63 or S64	
S64	TI (incentive N2 plan or incentive N2 plans or Incentive N2 program* or physician N2 compens*) or AB (incentive N2 plan or incentive N2 plans or Incentive N2 program* or physician N2 compens*)	
S63	TI (blend\$ N2 payment* or bonus\$ N2 payment* or capped N2 payment* or capitate* N2 payment* or fixed N2 payment* or linear payment* mixed N2 payment* or non-linear payment or "performance pay*" or prospective payment* or target N2 payment* or capitate or AB (blend\$ N2 payment* or bonus\$ N2 payment* or capped N2 payment* or capitate* N2 payment* or fixed N2 payment* or linear payment* mixed N2 payment* or fixed N2 payment* or linear payment* mixed N2 payment* or fixed N2 payment* or linear payment* mixed N2 payment* or capitate* N2 payment* or fixed N2 payment* or linear payment* mixed N2 payment* or capitate* N2 payment* or capitate* N2 payment* or payment* or target N2 payment* or capitation)	
S62	TI ("episode of care" N2 payment* or payment* N2 "per-patient" or payment* N2 "per-episode" or payment* N2 "per-visit" or retroactiv\$ N2 payment* or retrospectiv\$ N2 payment* or threshold* N2 payment* or variable N2 payment*) or AB ("episode of care" N2 payment* or payment* N2 "per-patient" or payment* N2 "per-episode" or payment* N2 "per-visit" or retroactiv\$ N2 payment* or retrospectiv\$ N2 payment* or threshold* N2 payment* or variable N2 payment*)	
S61		

	TI ("fee for service" OR "pay for performance" or "performance based payment*" or "value based
	payment*") or AB ("fee for service" OR "pay for performance" or "performance based payment*" or
	"value based payment*")
S60	TI ("physician reimburs" OR "practitioner reimburs" OR "reimbursement mechanism" or reimbursement plan* or reimbursement program or reimbursement scheme*) or AB ("physician reimburs" OR "practitioner reimburs" OR "reimbursement mechanism" or reimbursement plan* or reimbursement
	program or roimbursoment scheme*)
	TI (Einancial Incontive* OB physician incontive* or "financial reward" OB "monotony reward" OB "economic
550	reward" OP "financial honus" OP "monotary honus" OP "economic honus" OP "colory honus" OP "financial
555	honus" OR "monetary bonus" OR doctor N1 incentiv*) or AB (Financial Incentive* OR physician incentive*
	or "financial reward" OP "monetary reward" OP "economic reward" OP "financial honus" OP "monetary
	bonus" OR "economic bonus" OR "salary bonus" OR "financial bonus" OR "monetary bonus" OR doctor N1
	incentiv*)
65.0	
558	MH "Physician Incentive Plans"
S57	S51 or S52 or S53 or S54 or S55 or S56
S56	TI (Primary care or primary healthcare or "primary health care") or AB (Primary care or primary healthcare or "primary health care")
	TI (Family N2 practice or Family N2 Practitioner* or General practice* or general practitioner* or family
S55	physician* or family doctor* or general* physician* or group practice* or private practice*) or AB (Family
	N2 practice or Family N2 Practitioner* or General practice* or general practitioner* or family physician* or
	family doctor* or general* physician* or group practice* or private practice*)
S54	MH "group practice+"
S53	MH "family practice"
S52	MH "primary health care"
S51	MH "Physicians+"
650	
550	516 of 517 of 518 of 519 of 520 of 521 of 522 of 523 of 524 of 525 of 526 of 527 of 528 of 529 of 530 of 531
	01 352 01 355 01 354 01 355 01 356 01 357 01 358 01 359 01 340 01 341 01 342 01 345 01 344 01 345 01 346 01
	TI ("navment* N2 hlend*" or "navment* N2 hlue cross" or "navment* N2 honus*" or "navment* N2
549	capitat*" or "navment* N2 capped" or "navment* N2 enisode of care" or "navment* N2 fixed" or
545	"navment* N2 government*" or "navment* N2 insur*" or "navment* N2 level*" or "navment* N2 linear"
	or "payment* N2 medicaid" or "payment* N2 medicare" or "payment* N2 non-linear" or "payment* N2
	per-patient" or "payment* N2 per-episode" or "payment* N2 per-visit" or "payment* N2 performance" or
	"payment* N2 prospectiv*" or "payment* N2 retroactiv*" or "payment* N2 retrospectiv*" or "payment*
	N2 reward*" or "payment* N2 schedule*" or "payment* N2 system*" or "payment* N2 target*" or
	"payment* N2 third-part*" or "payment* N2 threshold*" or "payment* N2 uncap*" or "payment* N2
	shared" or "payment* N2 variable" or "payment* N2 per-visit*") or AB ("payment* N2 blend*" or
	"payment* N2 blue cross" or "payment* N2 bonus*" or "payment* N2 capitat*" or "payment* N2 capped"
	or "payment* N2 episode of care" or "payment* N2 fixed" or "payment* N2 government*" or "payment*
	N2 insur*" or "payment* N2 level*" or "payment* N2 linear" or "payment* N2 medicaid" or "payment* N2
	medicare" or "payment* N2 non-linear" or "payment* N2 per-patient" or "payment* N2 per-episode" or
	"payment* N2 per- visit" or "payment* N2 performance" or "payment* N2 prospectiv*" or "payment* N2
	retroactiv*" or "payment* N2 retrospectiv*" or "payment* N2 reward*" or "payment* N2 schedule*" or
	"payment* N2 system*" or "payment* N2 target*" or "payment* N2 third-part*" or "payment* N2

	threshold*" or "payment* N2 uncap*" or "payment* N2 shared" or "payment* N2 variable" or "payment* N2 per-visit*")
S48	TI ("financial N2 bonus*" or "financial N2 reward*" or "incentive* N2 bonus*" or "incentive* N2 reward*" or "monetary* N2 bonus*" or "monetary* N2 reward*" or "physician* N2 bonus*" or "physician* N2 reward*" or "provider* N2 bonus*" or "provider* N2 reward*" or "practitioner* N2 bonus*" or "practitioner* N2 reward*" or "salar* N2 bonus*" or "salar* N2 reward*") or AB ("financial N2 bonus*" or "financial N2 reward*" or "incentive* N2 bonus*" or "incentive* N2 reward*" or "monetary* N2 bonus*" or "monetary* N2 reward*" or "physician* N2 bonus*" or "incentive* N2 reward*" or "monetary* N2 bonus*" or "monetary* N2 reward*" or "physician* N2 bonus*" or "physician* N2 reward*" or "provider* N2 bonus*" or "provider* N2 reward*" or "physician* N2 bonus*" or "physician* N2 reward*" or "salar* N2 bonus*" or "provider* N2 reward*" or "practitioner* N2 bonus*" or "practitioner* N2 reward*" or "salar* N2 bonus*" or "salar* N2 reward*" or "salar* N2 bonus*" or "practitioner* N2 reward*" or "salar* N2 bonus*" or
S47	TI ("performance* N2 pay*" or P4P or "pay for quality improvement*" or P4QI or "fee-for service*") or AB ("performance* N2 pay*" or P4P or "pay for quality improvement*" or P4QI or "fee-for service*")
S46	TI ("physician* N4 bonus*" or "physician* N4 incentive*" or "physician* N4 financial" or "physician* N4 monetar*" or "physician* N4 payment*" or "physician* N4 pay*" or "physician* N4 plan*" or "physician* N4 salar*" or "practitioner* N4 bonus*" or "practitioner* N4 incentive*" or "practitioner* N4 financial" or "practitioner* N4 monetar*" or "practitioner* N4 payment*" or "practitioner* N4 pay*" or "practitioner* N4 plan*" or "practitioner* N4 pay*" or "practitioner* N4 pay*" or "doctor* N4 plan*" or "doctor* N4 pay*" or "physician* N4 financial" or "physician* N4 financial" or "physician* N4 payment*" or "doctor* N4 pay*" or "doctor* N4 pay*" or "doctor* N4 pay*" or "doctor* N4 pay*" or "physician* N4 pay*" or "physician* N4 pay*" or "doctor* N4 pay*" or "doctor* N4 pay*" or "physician* N4 pay*" or "practitioner* N4 pay*" or "doctor* N4 reward*" or "practitioner* N4 pay*" or "doctor* N4 pay*" or "doctor* N4 pay*" o
S45	TI ("change* N2 patient outcome*" or "change* N2 policy" or "change* N2 policies" or "change* N2 practic*" or "change* N2 provider*" or "change* N2 treatment outcome*" or "changing N2 policies" or "changing N2 patient outcome*" or "changing N2 policy" or "changing N2 policies" or "changing N2 practic*" or "changing N2 provider*" or "changing N2 policy" or "changing N2 policies" or "changing N2 practic*" or "changing N2 provider*" or "changing N2 treatment outcome*" or "changing N2 disease management" or "improv* N2 patient outcome*" or "improv* N2 policy" or "improv* N2 policies" or "improv* N2 practic*" or "improv* N2 patient outcome*" or "improv* N2 policy" or "improv* N2 policies" or "improv* N2 practic*" or "improv* N2 patient outcome*" or "effect* N2 policy" or "effect* N2 policies" or "effect* N2 practic*" or "effect* N2 provider*" or "effect* N2 reatment outcome*" or "effect* N2 policies" or "influenc* N2 practic*" or "influenc* N2 provider*" or "influenc* N2 policy" or "effect* N2 policies" or "influenc* N2 provider*" or "influenc* N2 policy" or "influenc* N2 policies" or "influenc* N2 provider*" or "influenc* N2 policy" or "change* N2 policies" or "influenc* N2 practic*" or "influenc* N2 provider*" or "influenc* N2 policy" or "change* N2 policies" or "change* N2 practic*" or "changing N2 practic*" or "change* N2 policy" or "change* N2 policy" or "change* N2 policies" or "changing N2 practic*" or "changing N2 provider*" or "changing N2 policy" or "changing N2 practic*" or "changing N2 provider*" or "changing N2 policy" or "improv* N2 policies" or "changing N2 practic*" or "changing N2 practic*" or "changing N2 practic*" or "improv* N2 provider*" or "improv* N2 policy" or "improv* N2 policy" or "improv* N2 policy" or "improv* N2 policies" or "improv* N2 practic*" or "improv* N2 provider*" or

S44	TI "virtual communit*" or AB "virtual communit*"	
S43	MH Social marketing/ or "social marketing"	
S42	TI (internet.ti,ab,hw. or (intranet or LAN or WAN or blog* or (computer* adj2 network*) or online* or web* or wiki)) or AB (internet.ti,ab,hw. or (intranet or LAN or WAN or blog* or (computer* adj2 network*) or online* or web* or wiki))	
S41	TI (telephon*.ti,ab,hw. or (tele-health or tele-medicine or e-health)) or AB (telephon*.ti,ab,hw. or (tele- health or tele-medicine or e-health))	
S40	TI (computers, handheld/ or handheld*.ti,ab. or (PDA or "personal data assistant*" or blackberr*)) or AB (computers, handheld/ or handheld*.ti,ab. or (PDA or "personal data assistant*" or blackberr*))	
S39	TI (("user computer" or "computer user")) or AB (("user computer" or "computer user")) TI ("quality N2 assurance" or "quality N2 improvement" or "quality N2 initiativ*" or "quality N2 plan*" or "quality N2 program*" or "quality N2 review") or AB ("quality N2 assurance" or "quality N2 improvement"	
330	or "quality N2 initiativ*" or "quality N2 plan*" or "quality N2 program*" or "quality N2 review")	
S37	TI ("policy N2 chang*" or "policy N2 effect*" or "policy N2 impact*" or "policy N2 influenc*" or "policies N2 chang*" or "policies N2 effect*" or "policies N2 impact*" or "policies N2 influenc*") or AB ("policy N2 chang*" or "policy N2 effect*" or "policy N2 impact*" or "policy N2 influenc*" or "policies N2 chang*" or "policies N2 effect*" or "policy N2 impact*" or "policy N2 influenc*" or "policies N2 chang*" or "policies N2 impact*" or "policy N2 influenc*" or "policies N2 chang*" or "policies N2 effect*" or "policy N2 impact*" or "policy N2 influenc*" or "policies N2 chang*" or "policies N2 impact*" or "policies N2 impact*" or "policy N2 influenc*" or "policies N2 chang*" or "policies N2 impact*" or "policies N2 influenc*" or "policies N2 chang*" or "policies N2 impact*" or "policies N2 influenc*" or "policies N2 chang*" or "policies N2 impact*" or "policies N2 influenc*" or "policies N2 chang*" or "policies N2 impact*" or "policies N2 impact*" or "policies N2 influenc*" or "policies N2 chang*" or "policies N2 impact*" or "policies N2 impact*" or "policies N2 influenc*" or "policies N2 impact*" or "	
S36	TI (("physician behavio*r*" or "practice pattern*")) or AB (("physician behavio*r*" or "practice pattern*"))	
S35	TI community-base* or AB community-base*	
S34	TI ("outreach N2 communit*" or "outreach N2 plan*" or "outreach N2 program*" or "outreach N2 visit*") or AB ("outreach N2 communit*" or "outreach N2 plan*" or "outreach N2 program*" or "outreach N2 visit*")	
S33	TI "opinion leader*" or AB "opinion leader*"	
S32	TI ("nurse-led N2 nurse*" or "nurse-led N2 led" or "nurse-led N2 managed" or "nurse-led N2 coordinat*" or "nurse-led N2 co-ordinat*" or "nurse* N2 led" or "nurse* N2 managed" or "nurse* N2 coordinat*" or "nurse* N2 co-ordinat*") or AB ("nurse-led N2 nurse*" or "nurse-led N2 led" or "nurse-led N2 managed" or "nurse-led N2 coordinat*" or "nurse-led N2 co-ordinat*" or "nurse* N2 led" or "nurse* N2 managed" or "nurse* N2 coordinat*" or "nurse-led N2 co-ordinat*" or "nurse* N2 led" or "nurse* N2 managed" or "nurse* N2 coordinat*" or "nurse* N2 co-ordinat*" or "nurse* N2 led" or "nurse* N2 managed" or "nurse* N2 coordinat*" or "nurse* N2 co-ordinat*")	
S31	TI ("knowledge N2 synthesis" or "evidence N2 synthesis") or AB ("knowledge N2 synthesis" or "evidence N2 synthesis")	
S30	TI ("knowledge N2 gap*" or "knowledge N2 barrier*" or "evidence N2 gap*" or "evidence N2 barrier*" or "practice N2 gap*" or "practice N2 barrier*") or AB ("knowledge N2 gap*" or "knowledge N2 barrier*" or "evidence N2 gap*" or "evidence N2 barrier*" or "practice N2 gap*" or "practice N2 barrier*")	
S29	TI ("knowledge N2 transfer*" or "knowledge N2 translation" or KT") or AB ("knowledge N2 transfer*" or "knowledge N2 translation" or KT")	
S28	TI (collaborat* or "cross-profession*" or intraprofession* or intra-profession* or inter-profession* or interprofession* or teambase* or team-based "skill N2 mix*") or AB (collaborat* or "cross-profession*" or intraprofession* or intra-profession* or inter-profession* or interprofession* or teambase* or team-based "skill N2 mix*")	

S27	TI ("incentiv* N2 economic" or "incentiv* N2 employee*" or "incentiv* N2 financ*" or "incentiv* N2 insurer*" or "incentiv* N2 insurance" or "incentiv* N2 market*" or "incentiv* N2 monetar*" or "incentiv* N2 pay*" or "incentiv* N2 physician*" or "incentiv* N2 plan*" or "incentiv* N2 practitioner*" or "incentiv* N2 program*" or "incentiv* N2 provider*" or "incentiv* N2 reimburs*" or "incentiv* N2 salary" or "incentiv* N2 salarie*" or "incentiv* N2 staff" or "incentiv* N2 team*" or "incentiv* N2 value-based") or AB ("incentiv* N2 economic" or "incentiv* N2 employee*" or "incentiv* N2 financ*" or "incentiv* N2 insurer*" or "incentiv* N2 insurance" or "incentiv* N2 market*" or "incentiv* N2 monetar*" or "incentiv* N2 pay*" or "incentiv* N2 insurance" or "incentiv* N2 market*" or "incentiv* N2 monetar*" or "incentiv* N2 pay*" or "incentiv* N2 physician*" or "incentiv* N2 market*" or "incentiv* N2 monetar*" or "incentiv* N2 pay*" or "incentiv* N2 physician*" or "incentiv* N2 market*" or "incentiv* N2 monetar*" or "incentiv* N2 pay*" or "incentiv* N2 physician*" or "incentiv* N2 plan*" or "incentiv* N2 market*" or "incentiv* N2 salary" or "incentiv* N2 salarie*" or "incentiv* N2 plan*" or "incentiv* N2 practitioner*" or "incentiv* N2 program*" or "incentiv* N2 provider*" or "incentiv* N2 reimburs*" or "incentiv* N2 salary" or "incentiv* N2 salarie*" or "incentiv* N2 staff" or "incentiv* N2 reimburs*" or "incentiv* N2 salary" or
S26	MH Guideline Adherence
S25	TI ((feedback not (feedback loop*))) or AB ((feedback not (feedback loop*)))
S24	TI ("evidence based N3 algorithm*" or "evidence based N3 evaluat*" or "evidence based N3 guideline*" or "evidence based N3 healthcare" or "evidence based N3 implement*" or "evidence based N3 improv*" or "evidence based N3 intervention*" or "evidence based N3 management" or "evidence based N3 pathway*" or "evidence based N3 physician*" or "evidence based N3 plan*" or "evidence based N3 practic*" or "evidence based N3 prescrib*" or "evidence based N3 plan*" or "evidence based N3 quality" or "evidence based N3 treatment") or AB ("evidence based N3 algorithm*" or "evidence based N3 evaluat*" or "evidence based N3 treatment") or AB ("evidence based N3 algorithm*" or "evidence based N3 evaluat*" or "evidence based N3 guideline*" or "evidence based N3 healthcare" or "evidence based N3 implement*" or "evidence based N3 guideline*" or "evidence based N3 healthcare" or "evidence based N3 implement*" or "evidence based N3 improv*" or "evidence based N3 healthcare" or "evidence based N3 implement*" or "evidence based N3 improv*" or "evidence based N3 intervention*" or "evidence based N3 management" or "evidence based N3 pathway*" or "evidence based N3 physician*" or "evidence based N3 plan*" or "evidence based N3 practic*" or "evidence based N3 program*" or "evidence based N3 quality" or "evidence based N3 treatment") TX ("education* N3 continuing" or "education* N3 group*" or "education* N3 outreach" or "education* N3
S23	physician*" or "education* N3 plan*" or "education* N3 practitioner*" or "education* N3 program*" or "education* N3 staff*") or AB ("education* N3 continuing" or "education* N3 group*" or "education* N3 outreach" or "education* N3 physician*" or "education* N3 plan*" or "education* N3 practitioner*" or "education* N3 program*" or "education* N3 staff*")
S22	TI ("clincical N2 gap*" or "knowledge N2 gap*" or "evidence N2 gap*" or "quality N2 gap*" or "research N2 gap*" or "practice N2 gap*") or AB ("clincical N2 gap*" or "knowledge N2 gap*" or "evidence N2 gap*" or "quality N2 gap*" or "research N2 gap*" or "practice N2 gap*")
S21	MH decision making/ or decision mak*
S20	TI ("change* N2 patient outcome*" OR "change* N2 policy" OR "change* N2 policies" or "change* N2 process*" or "change* N2 practic*" or "change* N2 provider*" or "change* N2 treatment outcome*" or "change* N2 disease management" or "change* N2 policies" or "improv* N2 policies" or "improv* N2 policies" or "improv* N2 policy" or "improv* N2 policies" or "improv* N2 process*" or "improv* N2 practic*" or "improv* N2 provider*" or "improv* N2 policies" or "improv* N2 policies" or "improv* N2 policies" or "improv* N2 process*" or "improv* N2 process*" or "effect* N2 process*" or "effect* N2 policies" or "effect* N2 process*" or "effect* N2 process*" or "effect* N2 process*" or "influenc* N2 process*" or "influenc* N2 policies" or "influenc* N2 process*" or "influenc* N2 provider*" or "influenc* N2 process*" or "change* N2 process*" or "change* N2 provider*" or "change* N2 provider*" or "influenc* N2 process*" or "change* N2 process*" or "change* N2 process*" or "change* N2 provider*" or "influenc* N2 provider*" or "influenc* N2 process*" or "change* N2 process*" or "change* N2 provider*" or "influenc* N2 provider*" or "influenc* N2 process*" or "influenc* N2 provider*" or "change* N2 provider*" or "change* N2 process*" or "influe

	outcome*" or "effect* N2 policy" or "effect* N2 policies" or "effect* N2 process*" or "effect* N2 practic*" or "effect* N2 provider*" or "effect* N2 treatment outcome*" or "effect* N2 disease management" or "effect* N2 process*" or "influenc* N2 patient outcome*" or "influenc* N2 policy" or "influenc* N2 policies" or "influenc* N2 process*" or "influenc* N2 practic*" or "influenc* N2 provider*" or "influenc* N2 treatment outcome*" or "influenc* N2 disease management" or "influenc*")		
S19	TI (booklet* or brochure* or pamphlet* or paper-based or "printed material*") or AB (booklet* or brochure* or pamphlet* or paper-based or "printed material*")		
S18	TI (barrier* and facilitator) or AB (barrier* and facilitator)		
S17	TI (audit or self-audit) or AB (audit or self-audit)		
S16	TI academic detailing or AB academic detailing		
S15	S10 or S11 or S12 or S13 or S14		
S14	TI controlled or AB controlled		
S13	TI random* or AB random*		
S12	TI ("clinical study" or "clinical studies") or AB ("clinical study" or "clinical studies")		
S11	(MM "Clinical Trials+")		
S10	TI ((comparative N2 study) or (comparative N2 studies) or "evaluation study" or "evaluation studies") or AB ((comparative N2 study) or (comparative N2 studies) or "evaluation study" or "evaluation studies")		
S9	TI ((multicent* n2 design*) or (multicent* n2 study) or (multicent* n2 studies) or (multicent* n2 trial*)) or AB ((multicent* n2 design*) or (multicent* n2 study) or (multicent* n2 studies) or (multicent* n2 trial*))		
S8	TI ((cross-sectional N2 design) or (cross-sectional N2 study) or (cross-sectional N2 trial)) or AB ((cross- sectional N2 design) or (cross-sectional N2 study) or (cross-sectional N2 trial))		
S7	MH "Multiple Time Series"		
S6	MH "Time Series"		
S5	MH Experimental Studies or Community Trials or Community Trials or Pretest-Posttest Design + or Quasi- Experimental Studies + Pilot Studies or Policy Studies + Multicenter Studies		
S4	TI (quasi-experiment* or quasiexperiment* or quasi-random* or quasirandom* or quasi control* or quasicontrol* or "quasi* N3 method*" or "quasi* N3 study" or "quasi* N3 studies" or "quasi* N3 trial" or "quasi* N3 design*" or "experimental N3 method*" or "experimental N3 study" or "experimental N3 studies" or "experimental N3 trial" or "experimental N3 design*") or AB (quasi-experiment* or quasiexperiment* or quasi-random* or quasirandom* or quasi control* or quasicontrol* or "quasi* N3 method*" or "quasi* N3 study" or "quasi* N3 studies" or "quasi* N3 trial" or "quasi* N3 design*" or "experimental N3 method*" or "experimental N3 studies" or "quasi* N3 trial" or "quasi* N3 design*" or "trial" or "experimental N3 design*")		
S3	TI ("control* N2 before" or "control* N2 after") or AB ("control* N2 before" or "control* N2 after")		
S2	TI ("pre test*" or pretest* or posttest* or "post test*") or AB ("pre test*" or pretest* or posttest* or "post test*")		
S1	TI (intervention* or multiintervention* or multi-intervention* or postintervention* or post-intervention* or pre-intervention*) or AB (intervention* or multiintervention* or multi- intervention* or postintervention* or post-intervention* or preintervention* or pre-intervention*)		

Cochrane

No Query

#1 MeSH descriptor Physicians explode all trees

#2 (physician* OR doctor*):ti and (physician* OR doctor*):kw

#3 MeSH descriptor Primary Health Care explode all trees

#4 MeSH descriptor Family Practice, this term only

#5 MeSH descriptor Group Practice explode all trees

#6 MeSH descriptor Partnership Practice explode all trees

#7 MeSH descriptor Private Practice explode all trees

#8 "family practice" or "general practice" or "general practitioner*" or "family doctor" or "primary care" or "primary healthcare" or "primary health care" or "group practice*" or "group practitioner*":ti,ab,kw

#9 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8)

#10 MeSH descriptor Reimbursement Mechanisms explode all trees

#11 MeSH descriptor Insurance, Health, Reimbursement, this term only

#12 MeSH descriptor Medicare Payment Advisory Commission explode all trees

#13 MeSH descriptor Fees and Charges explode all trees

#14 MeSH descriptor Income explode all trees

#15 (capitation or c apitated or capitating or fundhold\$ or "fund-hold*"):ti and (capitation or capitated or capitating or fundhold\$ or "fund-hold*"):ab

#16 "Economic incentive*" OR "Financial Incentive*" OR "physician incentive*" OR "monetary incentive*" or "economic reward" or "financial reward" OR "monetary reward" OR "financial bonus" OR "monetary bonus" OR "economic bonus" OR "salary bonus" OR "financial bonus"-:ti or "Economic incentive*" OR "Financial Incentive*" OR "physician incentive*" OR "monetary incentive*" or "economic reward" or "financial reward" OR "monetary reward" OR "financial bonus" OR "monetary bonus" OR "economic bonus" OR "salary bonus" OR "monetary reward" OR "financial bonus" OR "monetary bonus" OR "economic bonus" OR "salary bonus" OR "financial bonus"-:ab or "physician reimburs" OR "practitioner reimburs" OR "reimbursement mechanism" or "reimbursement plan*" or "reimbursement program*" or "reimbursement scheme*":ti or "physician reimburs" OR "practitioner reimburs" OR "reimbursement mechanism" or "reimbursement plan*" or "reimbursement program*" or "reimbursement scheme*":ab

#17 "fee for service" OR "pay for performance" or "performance based payment*" or "value based payment*" :ti or "fee for service" OR "pay for performance" or "performance based payment*" or "value based payment*" :ab or "Blend* payment*" or "bonus* payment*" or "capped payment*" or "capitate* payment*" or "fixed payment*" or "linear payment*" OR "mixed payment*" or "nonlinear payment*" or "performance pay*" or "prospective payment*" or "target payment*" or capitation or "physician* payment*" or "physician* fee*" or "Payment threshold*":ti or "Blend* payment*" or "bonus* payment*" or "capped payment*" or "capitate* payment*" or "fixed payment*" or "linear payment*" or "bonus* payment*" or "non-linear payment*" or "performance pay*" or "prospective payment*" or "target payment*" or "target payment*" or "physician* payment*" or "physician* fee*" or "Payment threshold*":ab

#18 (("episode of care" OR "per-patient" OR "per-episode" OR "per-visit") AND payment*):ti or (("episode of care" OR "per-patient" OR "per-episode" OR "per-visit") AND payment*):ab

#19 (#10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18)

#20 (#9 AND #19), from 2010 to 2023

#21 MeSH descriptor Physician Incentive Plans, this term only

#22 (#20 AND ("financial incentive*" OR "physician incentive*"))

Embase

No	Query		
1	(Physician? adj2 Incentiv\$ adj (plan? or program or policy)).ti,ab.		
2	((physician? or practitioner? or doctor?) adj4 (bonus\$ or incentive? or financial or monetar\$ or payment? or "profit shar\$" or reward? or salar\$)).ti,ab.		
3	exp Physicians/ and (incentiv\$ adj (economic or financial or monetar\$ or payment? or reimburs\$)).ti,ab,hw.		
4	*Medical fee/		
5	*Income/ or *"Salary"/ or Physician Income/		
6	Reimbursement/ or reimburs\$.ti.		
7	*Income/ or *"Salary"/ or Physician Income/		
8	("pay for performance" or "pay for compliance" or "pay for participation" or "performance pay\$" or P4P or "pay for quality improvement?" or P4QI or "fee*for service?").ti,ab.		
9	((doctor? or physician? or "general practitioner?" or pa?ediatrician?) adj2 (fee? or income or salary or salaries)).ti,ab.		
10	((incentiv\$ or bonus\$ or reward?) adj (economic or employee? or financ\$ or insurer? or insurance or "managed care" or HMO or market\$ or monetar\$ or payment? or "performance based" or physician? or practitioner? or program\$ or provider? or reimburs\$ or salary or staff or team\$ or value based)).ti,ab.		
11	(reimburs\$ adj (disproportion\$ or health\$ or insurer? or mechanism? or plan\$ or physician? or practitioner? or program\$ or proportion\$ or provider? or relative or scale? or share? or sharing or valu based or performance base? or QI or quality or scheme?)).ab.		
12	((compensation or compensatory) adj2 (doctor? or physician? or plan? or practitioner? or system?)).ti,ab.		
13	(capitation or capitated or capitating or fundhold\$ or fund hold\$).ti,ab.		
14	("rate setting" or "rate review").ti,ab.		
15	(gainshar\$ or payer provider? or payer patient?).ti,ab.		
16	("pay for compliance" or "pay for participation" or "pay for performance" or "performance pay\$" or P4P or "pay for quality improvement?" or P4QI or "fee*for service?").ti,ab.		
17	(payment? adj (blend\$ or "blue cross" or bonus\$ or capped or "episode of care" or fixed or government\$ or insurance or insurer? or level? or linear or medicaid or medicare or non*linear or per*patient or per*episode or per*visit or performance or prospectiv\$ or retroactiv\$ or retrospectiv\$ or reward\$ or schedule? or system? or target\$ or third part\$ or threshold? or uncap\$ or shared or variable or per*visit?)).ti,ab.		
18	((copay\$ or co*pay\$ or cost*shar\$ or prepaid or pre*paid or prepay\$ or pre*pay\$) adj4 (physician? or practitioner? or performance)).ti,ab.		
19	primary health care/ or primary medical care/		
20	(primary adj2 care).ti,ab.		
21	General Practitioner/		
22	General Practice/		
23	((community or family or general or group) adj2 (doctor? or physician? or practice? or practitioner?)).ti,ab.		
24	randomized controlled trial/ or controlled clinical trial/ or clinical trial/ or controlled study/		
25	multicenter study/		
26	major clinical study/		
27	random\$.ti,ab.		

28	controlled.ti.		
29	(control\$ adj2 (clinical or group\$ or trial\$ or study or studies or design\$ or method\$)).ti,ab.		
30	((multicent\$ or multi*cent\$ or multisite? or multi*site?) adj (study or studies or trial\$)).ti,ab.		
31	((single or double or triple or treble) adj blind\$).ti,ab.		
32	single blind procedure/ or double blind procedure/		
33	(intervention? or multiintervention? or multi*intervention? or postintervention? or post*intervention? or pre*intervention?).ti.ab.		
34	(pre*post or "pre test\$" or pretest\$ or posttest\$ or "post test\$").ti.ab.		
35	(control\$ adj "before and after").ti,ab.		
36	before*after.ti,ab.		
37	("quasi*experiment\$" or quasiexperiment\$ or "quasi random\$" or quasirandom\$ or "quasi control\$" or quasicontrol\$ or ((quasi\$ or experimental) adj3 (method\$ or study or studies or trial or design\$))).ti,ab.		
38	("time series" adj2 interrupt\$).ti,ab,hw.		
39	experimental design/		
40	quasi experimental study/		
41	(knowledge adj2 (application or broke\$ or creation or diffus\$ or disseminat\$ or exchang\$ or implement\$ or management or mobili\$ or translat\$ or transfer\$ or uptake or utili\$)).ti,ab. or "knowledge translation"/		
42	(research\$ adj2 (diffus\$ or disseminat\$ or exchang\$ or transfer\$ or translation\$)).ti,ab.		
43	("research findings into action" or "research to action" or "research into action" or "evidence to action" or "evidence to practice").ti,ab.		
44	(diffusion adj2 innovation).ti,ab.		
45	research utili?ation.ti,ab.		
46	((knowledge or evidence) adj2 synthesis).ti,ab.		
47	(evidence\$ adj2 (exchang\$ or translat\$ or transfer\$)).ti,ab.		
48	(("systematic review\$" or "knowledge synthes\$") adj5 ("decision mak\$" or "policy mak\$" or policymak\$ or "policy decision?" or "health polic\$")).ti,ab.		
49	(("systematic review\$" or "knowledge synthes\$") adj2 (application or implement\$ or utili?ation or utilize? or utilise? or utili?ing)).ti,ab.		
50	"academic detailing".ti,ab.		
51	(audit or self*audit).ti,ab,hw.		
52	"barrier? and facilitator?".ti,ab.		
53	(booklet\$ or brochure? or pamphlet? or paper*based or "printed material?").ti,ab.		
54	decision making/ or decision mak\$.ti,ab,hw.		
55	((change? or changing or improv\$ or effect\$ or influenc\$) adj2 ("patient outcome?" or policy or policies or process\$ or practic\$ or provider? or "treatment outcome?" or "disease management" or process\$)).ti,ab.		
56	((clinical or knowledge or evidence or quality or research or practice) adj2 gap?).ti,ab.		
57	(education\$ adj3 (continuing or group? or outreach or physician? or plan\$ or practitioner? or program? or staff? or team?)).ti,ab,hw.		
58	("evidence based" adj3 (algorithm? or evaluat\$ or guideline? or healthcare or implement\$ or improv\$ or intervention\$ or management or medicine or pathway? or physician? or plan? or practic\$ or prescrib\$ or program? or quality or treatment)).ti,ab.		
59	(feedback not (feedback adj loop\$)).ti,hw.		
60	Guideline Adherence/		

61	(guideline? adj3 (adhere\$ or enforc\$ or influenc\$ or implement\$ or impact\$ or introduc\$ or uptake)).ti,ab.		
62	(incentiv\$ adj2 (economic or employee? or financ\$ or insurer? or insurance or market\$ or monetar\$ or pay\$ or physician? or plan? or practitioner? or program\$ or provider? or reimburs\$ or salary or salarie? or staff or team\$ or value*based)).ti,ab.		
63	(collaborat\$ or "cross*profession\$" or intraprofession\$ or intra*profession\$ or interprofession\$ or interprofession\$ or inter*profession\$ or (skill adj2 mix\$) or teambase? or team*based).ti,ab,hw.		
64	("nurse*led" or (nurse? adj2 (led or managed or coordinat\$ or co*ordinat\$))).ti,ab.		
65	((knowledge or evidence or practice) adj2 (gap? or barrier?)).ti,ab.		
66	"opinion leader?".ti,ab.		
67	(outreach adj2 (communit\$ or plan? or program? or visit?)).ti,ab.		
68	community*base?.ti,ab.		
69	("physician behavio?r?" or "practice pattern?").ti,ab,hw.		
70	((policy\$ or policies) adj2 (chang\$ or effect? or impact? or influenc\$)).ti,ab.		
71	(quality adj2 (assurance or improvement? or initiativ\$ or plan\$ or program\$ or review)).ti,ab.		
72	(QI adj (inititative? or intervention? or program\$ or plan\$)).ti,ab.		
73	("user computer" or "computer user").ti,ab.		
74	computers, handheld/ or handheld?.ti,ab. or (PDA or "personal data assistant?" or blackberr\$).ti,ab.		
75	telephon\$.ti,ab,hw. or (tele*health\$ or tele*medicine).ti,ab.		
76	(e*health\$ or e*medicine or e*practice).ti,ab.		
77	internet.ti,ab,hw. or (intranet or LAN or WAN or blog\$ or (computer\$ adj2 network\$) or online\$ or videoconferen\$ or web\$ or wiki).ti,ab.		
78	(video\$ adj5 (diagnos\$ or healthcare or "health care" or learning or "patient care" or teaching)).ti,ab.		
79	Social marketing/ or (marketing or "virtual communit\$" or facebook or twitter or "social networking").ti,ab.		
80	((change? or changing or improv\$ or effect\$ or influenc\$) adj2 ("healthcare delivery" or "health care delivery" or "delivery of health\$" or (organi? adj2 healthcare) or (organi? adj2 "health care") or "patient outcome?" or policy or policies or practic\$ or provider? or "treatment outcome?" or "disease management")).ti,ab.		
81	(effective adj2 (practice or healthcare or "health care")).ti,ab.		
82	((standard or usual) adj care).ti,ab.		
83	((doctor? or "healthcare provider?" or "health care provider?" or nurse? or nursing or physician? or practitioner?) adj4 (bonus\$ or incentive? or financial or monetar\$ or payment? or pay? or plan? or reward? or salar\$)).ti,ab.		
84	((performance\$ adj2 pay\$) or P4P or "pay for quality improvement?" or P4QI or "fee*for service?").ti,ab.		
85	("performance based" or value*based).ti,ab.		
86	((financial or incentive? or monetar\$ or physician? or provider? or practitioner? or salar\$) adj2 (bonus\$ or reward?)).ti,ab.		
87	(payment? adj2 (blend\$ or "blue cross" or bonus\$ or capitat\$ or capped or "episode of care" or fixed or government\$ or insur\$ or level? or linear or medicaid or medicare or non*linear or per*patient or per*episode or per*visit or performance or prospectiv\$ or retroactiv\$ retrospectiv\$ or reward\$ or schedule? or system? or target\$ or third*part\$ or threshold? or uncap\$ or shared or variable or per*visit?)).ti,ab.		
88	*cross*sectional study/		
89	(cross*sectional adj2 (design or study or studies or trial?)).ti,ab.		
90	*case control study/		

91	(("case control " or multicase or multi*case) adj2 (design? or study or studies or trial?)).ti,ab.	
92	*follow up/	
93	(("follow up" or follow*up) adj2 (design or study or studies)).ti,ab.	
94	*crossover procedure/	
95	((crossover or cross*over) adj2 (design or study or studies or trial)).ti,ab.	
96	*pilot study/	
97	(pilot\$ adj2 (project? or study or studies)).ti,ab.	
98	*comparative study/	
99	(comparative adj2 (study or studies)).ti,ab.	
100	*intervention study/	
101	("evaluation study" or "evaluation studies").ti,ab.	
102	1 or 2 or 3 [physician incentives]	
103	or/4-18 [financial terms]	
104	or/19-23 [physicians or primary care]	
105	exp primary medical care/	
106	or/24-32 [RCT filter]	
107	or/33-40 [epoc filter]	
108	or/41-87 [intervention filter]	
109	or/88-101 [alt studies]	
110	103 and 104 [fin incentives and physicians]	
111	106 or 107 [rct or epoc]	
112	108 and 109 [int and alt study filter]	
113	102 or 110 [phys. fin. incent. or fin. incent. and phys. or pc]	
114	113 and 105 [outcomes]	
115	111 or 112 [study design filters]	
116	114 and 115	
117	limit 116 to (abstracts and human and english language and yr="2010 - 2023")	
118	limit 117 to embase	
119	limit 118 to (clinical trial or randomized controlled trial or controlled clinical trial or multicenter study)	
120	limit 119 to (article and (journal or report))	
121	limit 118 to (meta analysis or outcomes research or "systematic review")	

(financial incentive*) OR (economic incentive*) OR (monetary incentive*) OR (target payment*) OR (blend* payment*) OR (non-linear payment*) OR (linear payment*) OR (payment threshold*) OR

(prospective payment ") OR "physician* compensation" OR "fee for service" OR "pay for performance" OR "pay for compliance" OR "pay for participation" OR "capitated payment*" OR capitation OR monetary

2. AND physician* OR "primary healthcare" OR "primary care" OR "general practitioner*" OR "family doctor*" OR "group practice*"

- 3. AND systematic review OR meta analysis
- 4. Filter 2010 2023

((noft(financial incentive*)) OR (noft(economic incentive*)) OR (noft(monetary incentive*)) OR (noft(target payment*)) OR (noft(blend* payment*)) OR (noft(non-linear payment*)) OR (noft(linear payment*)) OR (noft(payment threshold*)) OR (noft(prospective payment*)) OR noft("physician* compensation") OR noft("fee for service") OR noft("pay for performance") OR noft("pay for compliance") OR noft("fee for service") OR noft("capitated payment*") OR noft(capitation) OR noft(monetary)) AND noft(physician* OR "primary healthcare" OR "primary care" OR "general practitioner*" OR "family doctor*" OR "group practice*") AND noft(systematic review OR meta analysis)

PAIS

PsychINFO

No	Query		
1	monetary incentives/ or monetary rewards/		
2	incentives/		
3	salaries/ or bonuses/ or "income (economic)"/ or income level/ or professional fees/		
4	((incentiv\$ or bonus\$ or reward?) adj (economic or employee? or financ\$ or insurer? or insurance or market\$ or monetar\$ or payment? or physician? or practitioner? or program\$ or provider? or reimburs or salary or staff or team\$ or value*based)).ti,ab.		
5	(reimburs\$ adj (disproportion\$ or health\$ or insur\$ or mechanism? or plan\$ or physician? or practitioner? or program\$ or proportion\$ or provider? or relative or scale? or share? or sharing or value*based or performance*base? or QI or quality or scheme?)).ti,ab.		
6	("pay for compliance" or "pay for participation" or "pay for performance" or "performance pay\$" or P4P or "pay for quality improvement?" or P4QI or "fee*for service?").ti,ab.		
7	(payment? adj (blend\$ or "blue cross" or bonus\$ or capped or "episode of care" or fixed or government or insurance or insurer? or level? or linear or medicaid or medicare or non*linear or per*patient or per*episode or per*visit or performance or prospectiv\$ or retroactiv\$ or retrospectiv\$ or reward\$ or schedule? or system? or target\$ or third*part\$ or threshold? or uncap\$ or shared or variable or per*visit?)).ti,ab.		
8	((compensation or compensatory) adj (doctor? or physician? or plan? or practitioner? or system?)).ti,ab.		
9	(copay\$ or co*pay\$ or cost*shar\$ or prepaid or pre*paid or prepay\$ or pre*pay\$).ti,ab.		
10	(gainshar\$ or payer*provider? or payer*patient?).ti,ab.		
11	("rate setting" or "rate review").ti,ab.		
12	(capitation or capitated or capitating or fundhold\$ or fund*hold\$).ti,ab.		
13	fee for service/		
14	reimburs\$.ti.		
15	medical fee?.ti,ab.		
16	primary health care/		
17	private practice/		
18	(primary adj2 care).ti,ab.		
19	((community or family or general or group) adj2 (doctor? or physician? or practice? or practitioner?)).ti,ab.		
20	((partner\$ or private) adj (practice? or practitioner?)).ti,ab.		
21	exp physicians/		
22	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15		
23	16 or 17 or 18 or 19 or 20 or 21		
24	22 and 23		
25	limit 24 to (human and yr="2010 - 2023")		
26	limit 25 to (abstracts and ("0800 literature review" or "0830 systematic review" or 1200 meta analysis or 1300 metasynthesis)		

Pubmed

No	Query	Filters
8	(#3 OR #4) AND #2	Systematic Reviews
		and meta-analyses,
		2010 - 2023
_		
4	("Physicians"[Mesh] AND (incentiv*[Title/Abstract] AND (economic[Title/Abstract] OR	
	financial[litie/Abstract] OR monetary[litie/Abstract] OR payment*[litie/Abstract] OR	
	reimburs*[Title/Abstract]))) OR Physician Incentive Plans/ [MeSH Terms] OR	
	((physician*[Title/Abstract] OR practitioner*[Title/Abstract] OR	
	doctor [Inte/Abstract]) AND (bonus [Inte/Abstract] OR incentive [Inte/Abstract] OR	
	Inancial InterAdstract OR monetar [InterAdstract] OR payment [InterAdstract] OR	
	profit shar* [fitte/Abstract] OR reward*[fitte/Abstract] OR salar*[fitte/Abstract]))	
3	(consy*[Title/Abstract] OB co-nsy*[Title/Abstract] OB cost-shar*[Title/Abstract] OB	
5	nrenaid[Title/Abstract] OR pre-naid[Title/Abstract] OR prenav*[Title/Abstract] OR pre-	
	prepaul[nic/Abstract] or pre-paul[nic/Abstract] or prepay [nic/Abstract] or pre- nav*[Title/Abstract]) or ((compensation[Title/Abstract] OR	
	compensatory[Title/Abstract]) AND (doctor*[Title/Abstract] OR	
	nhysician*[Title/Abstract] OR nlan*[Title/Abstract] OR practitioner*[Title/Abstract] OR	
	system*[Title/Abstract])) or (novment*[Title/Abstract] AND (blend*[Title/Abstract]) OR	
	blue cross[Title/Abstract] OR bonus*[Title/Abstract] OR canned[Title/Abstract] OR	
	enisode of care[Title/Abstract] OR fived[Title/Abstract] OR	
	government*[Title/Abstract] OR insurance[Title/Abstract] OR insurer*[Title/Abstract]	
	OR level*[Title/Abstract] OR linear[Title/Abstract] OR medicaid[Title/Abstract] OR	
	medicare[Title/Abstract] OR non-linear[Title/Abstract] OR ner-nationt[Title/Abstract]	
	OR per-episode[Title/Abstract] OR per-visit*[Title/Abstract] OR	
	performance[Title/Abstract] OR prospectiv*[Title/Abstract] OR	
	retroactiv*[Title/Abstract] OR retrospectiv*[Title/Abstract] OR reward*[Title/Abstract]	
	OR schedule*[Title/Abstract] OR system*[Title/Abstract] OR target*[Title/Abstract] OR	
	third-part*[Title/Abstract] OR threshold*[Title/Abstract] OR uncap*[Title/Abstract] OR	
	shared[Title/Abstract] OR variable[Title/Abstract] OR per-visit*[Title/Abstract])) or	
	("pay for compliance"[Title/Abstract] OR "pay for participation"[Title/Abstract] OR	
	"pay for performance"[Title/Abstract] OR performance pay*[Title/Abstract] OR	
	P4P[Title/Abstract] OR pay for quality improvement*[Title/Abstract] OR	
	P4QI[Title/Abstract] OR fee-for service*[Title/Abstract]) or gainshar*[Title/Abstract] or	
	payer-provider*[Title/Abstract] or payer-patient*[Title/Abstract] or	
	(incentiv*[Title/Abstract] OR bonus*[Title/Abstract] OR reward*[Title/Abstract]) AND	
	(economic[Title/Abstract] OR employee*[Title/Abstract] OR financ*[Title/Abstract] OR	
	insurer*[Title/Abstract] OR insurance[Title/Abstract] OR market*[Title/Abstract] OR	
	monetar*[Title/Abstract] OR payment*[Title/Abstract] OR physician*[Title/Abstract]	
	OR practitioner*[Title/Abstract] OR program*[Title/Abstract] OR	
	provider*[Title/Abstract] OR reimburs*[Title/Abstract] OR salary[Title/Abstract] OR	
	staff[Title/Abstract] OR team*[Title/Abstract] OR value-based[Title/Abstract]) or	
	(reimburs*[Title/Abstract]) AND (disproportion*[Title/Abstract] OR	
	health*[Title/Abstract] OR insur*[Title/Abstract] OR mechanism*[Title/Abstract] OR	
	plan*[Title/Abstract] OR physician*[Title/Abstract] OR practitioner*[Title/Abstract] OR	
	program*[Title/Abstract] OR proportion*[Title/Abstract] OR provider*[Title/Abstract]	
	OR relative[Title/Abstract] OR scale*[Title/Abstract] OR share*[Title/Abstract] OR	
	sharing[Title/Abstract] OR value-based[Title/Abstract] OR performance-	
	base*[Title/Abstract] OR QI[Title/Abstract] OR quality[Title/Abstract] OR	

	scheme*[Title/Abstract]) or rate setting[Title/Abstract] or rate review[Title/Abstract]	
	or (capitation[Title/Abstract] OR capitated[Title/Abstract] OR capitating[Title/Abstract]	
	OR fundhold*[Title/Abstract] OR fund-hold*[Title/Abstract]) or fees and	1
	charges/[Mesh] or capitation fee/[Mesh] or fee-for-service plans/[Mesh] or fees,	1
	medical/[Mesh] or fees, pharmaceutical/[Mesh] or prescription fees/[Mesh] or rate	1
	setting and review/[Mesh] or insurance, health, reimbursement/[Mesh] or	1
	reimbursement mechanisms/[Mesh] or fee-for-service plans/[Mesh] or "physician	1
	payment review commission"/[Mesh] or medicare payment advisory	1
	commission/[Mesh] or reimbursement, disproportionate share/[Mesh] or	1
	reimbursement, incentive/[Mesh] or relative value scales/[Mesh] or Income/[Mesh] or	1
	Fees, Medical/[Mesh]	
2	"Primary Health Care"[Mesh] or (primary[Title/Abstract] AND care[Title/Abstract]) or	
	"Physicians, Family"[Mesh] or "Family Practice"[Mesh] or ((community[Title/Abstract]	1
	OR family[Title/Abstract] OR general[Title/Abstract]) AND (doctor*[Title/Abstract] OR	1
	physician*[Title/Abstract] OR practice*[Title/Abstract] OR	
	practitioner*[Title/Abstract]))	

Web of Science

1. (financial incentive*) OR (economic incentive*) OR (monetary incentive*) OR (target payment*) OR (blend* payment*) OR (non-linear payment*) OR (linear payment*) OR (payment threshold*) OR (prospective payment*) OR "physician* compensation" OR "fee for service" OR "pay for performance" OR "pay for compliance" OR "pay for participation" OR "capitated payment*" OR capitation OR monetary

2. AND physician* OR "primary healthcare" OR "primary care" OR "general practitioner*" OR "family doctor*" OR "group practice*"

3. AND systematic review OR meta analysis

4. Index Date 2010 – 2023

Appendix 2: Inclusion and exclusion criteria for screening

Parameter	Include	Exclude
Population	General practice, primary care	Specialised services Medical specialties Hospital services
Intervention / Exposure	Capitated funding Performance incentives Blended funding	
Comparator / Context	Fee for service	Patient payments only
Outcome	Access to care Quality of care Preventive care Multidisciplinary care Health outcomes Cost of health care Provider behaviour	
Study Characteristics	Systematic review Evaluation Economic analysis	Not in English
Other	New Zealand, UK, Europe, Canada, US, Singapore, Japan, Korea, Taiwan, Israel, Australia	Low and middle income countries

Table A2.1: Inclusion and exclusion criteria used for screening.

Appendix 3: Data extraction fields

Section	Items	Description/Coding
Paper	Author Year Full citation	
Type of review	Systematic review, Meta- analysis, Qualitative synthesis	
Search	Years of studies Information sources/ Data bases searched	
Included studies	Number by design Country/health system contexts Health focus Health service focus	RCTs, ITS, case control, cohort, observational before after, qualitative, comparison or control group Country/jurisdiction; predominant funding mechanism for primary care; reform context. Specific diseases or health issues Primary care/general practice, community programs, mixed
Population focus	Health provider	Medical, nursing, allied health, other
	Consumers/population	Age group group, socioeconomic status, ethnicity, disease characteristics, rurality.
Setting in which the intervention is delivered (Organisational/ practice	Service size/number of sites/service population size Practice type	Range of size, sites Range of different practices/services
	involvement (practice research network, etc)	
	Staffing profile Records, data and communication arrangements/ infrastructure	Range of staff (medical, nursing, allied health) Data sharing
Sample	Selection criteria Numbers in each category	
Risk of bias	Methods to assess Assessment	
Funding/incentive	Aim Type	Health or service impact, behavour P4P, capitation, salary, Blended, Other. Service eligibility, population/patient group.
	Criteria for funding Basis of funding	Performance measures, population coverage, measures

Table A3.1: Data extraction fields for systematic reviews and meta-analyses.

Section	Items	Description/Coding
	Funding Who receives funding Other Comparison	Funding amount, frequency Service (type), provider Other information on funding/incentive Funding for comparison
Impacts/outcomes (including definition, measurement tool/units, number of and period of follow up measures, type of analysis (eg intention to treat, quantitative/ qualitative)	Service use/Access Quality of care Hospitalisation Quality of life Costs/cost savings Patient experience Provider experience Unintended consequences	Access by specific populations: Type and frequency Preventive care, Chronic disease quality care Primary or readmissions Type of measure
Barriers and enablers to implementation		Describe
Change/ implementation		Implementation – is it evaluated and how Implementation – influences on implementation outcomes – identified/implied Translating/scaling up models – described? How? Stated change processes or models
Limitations (as	Authors	
identified) Interpretation (by authors)	Reviewer	
Relevance of SR to research questions	 Outcomes in MD settings Complex chronic diseases Comparison of payment models Behaviour change in primary care providers 	
Studies to highlight	Citations	

Section	ltem	Торіс
General Information	1	Identification features of the study: Record number, Author, Article title, Citation, Type of publication, Country of origin, Source of funding
Study Characteristics	2	Aim/objectives of the study, Study design, Study inclusion and exclusion criteria, Recruitment procedures used (e.g. details of randomization, blinding), Unit of allocation (e.g. participant, general practice, etc.)
Participant Characteristics	3	Characteristics of participants at the beginning of the study, such as: Age, Gender, ethnicity, Socioeconomic status, Disease characteristics, Comorbidities
	4	Number of participants in each characteristic category for intervention and comparison group(s) or mean/median characteristic values (record whether it is the number eligible, enrolled, or randomised that is reported in the study)
Intervention and Setting	5	Setting in which the intervention is delivered
Jan B	6	Description of the intervention(s) and control(s) (e.g. dose, route of administration, number of cycles, duration of cycle, care provider, how the intervention was developed, theoretical basis [where relevant])
	7	Description of cointerventions
Outcome Data/Results	8	Unit of assessment/analysis
	9	Statistical techniques used
	10	For each prespecified outcome: Whether reported, Definition used in study, Measurement tool or method used, Unit of measurement (if appropriate), Length of follow-up, number and/or times of follow-up measurements
	11	For all intervention group(s) and control group(s): Number of participants enrolled, Number of participants included in the analysis, Number of withdrawals and exclusions lost to follow-up, Summary outcome data
	12	Type of analysis used in study (e.g. intention to treat, per protocol)
	13	Results of study analysis, e.g. dichotomous (odds ratio, risk ratio and confidence intervals, p-value), continuous (mean difference, confidence intervals);
	14	If subgroup analysis is planned, the above information on outcome data or results will need to be extracted for each patient subgroup
	15	Additional outcomes, Costs, Resources, Adverse events

Table A3.2: Reporting of other literature and specific citations from within the systematic reviews or meta-analyses.

Appendix 4: Included studies – Description

Table A4.1: Overview of studies evaluated by included reviews.

Eirst author VEAR	lus	huk	Europe	Canada	Australia
Abmod 2021		UK Eladaran 2011, Backham 2014, Witter 2012, Scott 2011, Chauhan 2017, Michie 2011, Allen	Europe	Canada	Australia
Anneu 2021		Plogrefi 2011, Peckhain 2014, Witter 2015, Scott 2011, Chainar 2017, Michie 2011, Anen 2018, Hackett 2014, Roland 2014, Raleigh 2012, Fleetcroft 2010, Kontopantelis 2015, Fleetcroft 2012, Doran 2011, Mandavia 2017, Langdown 2014, Dixon 2011, Gillam 2012, Ackurach 2010, Fichers 2012			
Bec 2023		ASIWOTH 2010, Fichera 2017 Crampton 2016, Edwards 2015, Jones 2018, Lee 2019, MacVicar 2016, Roberts 2012	Aaras 2015 (Norway) Abelsen 2019 (Norway) Chevillard 2019		
2023			(France) Demmer 2021 (Germany) de Oliveira (Portugal) Dini		
			2012 (Germany) Dowling 2019 (Ireland) Ferste Kamer de Sater		
			Conoraal 2015 (Nothorlands) Elum 2016 (Cormany) Hoist 2015		
			(Cormany), Helta 2015 (Nervay), Karakelias 2017 (Crosse)		
			(Germany), Holte 2015 (Norway), Karakollas 2017 (Greece),		
			Keniet 2015 (Norway), Ozegowski 2013 (Europe), O Carroli 2015		
			(Ireland), Unal 2015 (Turkey), Vande Berg 2012 (Germany),		
			Wilhelmi 2018 (Germany)		
Boeckxstaens 2011		Millet 2007 a-c, Millet 2008 a-c, Mc Govern 2008 a-b, Simpson 2006, Crawley 2009, Ashworth	1		
		2007, Ashworth 2008, Doran 2008, Gulliford 2007, Saxena 2007, Strong 2006, Mc Lean 2006,			
		Doran 2006, Gray 2007			
Bowling 2018	Jones 2015, Rosenbaum 2014, Claffey 2012, Koh 2010, Bodenheimer 2006, Basu				
	2015, CMS 2016, Cleverley 2011, MPAC, 2018, Schroeder 2013, VanLare 2012,				
	Aroh 2015, Matthes 2010				
Brocklehurst 2013		Clarkson 2008, Coventry 1989			
Carter 2016				Heroux 2014, Levesque 2012, Feldman 2012, Manns 2012,	
				Campbell 2012, Kiran 2014, Li 2014, Kantarevic 2013, Kiran	
				2012, Jaakimainen 2011, Kantarevic, 2015, Kantarevic 2014,	
				Kralj 2013, Kantarevic 2011	
Emmert 2012	Nahra 2006, An 2008, Kouides 1998, Norton 1992, Rosenthal 2009, Rvan 2009,				
	Curtin 2006. Parke 2007				
Flodgren 2011	Davidson 1992. Hickson 1987. Kouides 1998	Baines 1997, Beaulieu 2005, Bradlow 1993, Burr 1992, Corney 1997, Coulter 1993, Harris	Linnala 2001 (Finland), Krasnik 1990, Walley 2000 (Ireland).	Hutchison 1996	
		1996 Kammerling 1996 Rafferty 1997 Ritchie 1992 Rosenthal 2005 Shen 2003 Whynes	Guether 1995 (Germany) Schoffski 1997 (Germany)		
		1997 Wilson 1995 Wilson 1990	Succiner 1995 (Sermany), Senonski 1997 (Sermany)		
Ciller 2012		Addink 2011 Alabhadi 2010 Ashwarth (2006, 2007, 2008, 2011) Battla 2008, Calvart 2000			
Ginam 2012		Adullik 2011, Alabbadi 2010, Ashwol (1 (2006, 2007, 2008, 2011), Boltie 2008, Calvert 2009,			
		Campbell (2007, 2008, 2009, 2010), Carey 2009, Checkland 2008, Coleman 2007, Crawley			
		2009, Cupples 2008, Dixon (2006, 2008, 2010, 2011), Downing 2007, Dowrick 2009, Fleetcrott			
		(2006, 2008, 2010), Gemmell 2009, Gulliford 2007, Guthrie 2006, Hippisley-Cox 2007, Jaiveer			
		2006, Khunti 2007, Kontopantelis 2010, Leydon 2011, MacBride-Stewart 2008, Magee 2010,			
		Maisey 2008, McDonald 2007, McElduff 2004, McGovern 2008, McGregor 2008, McLean			
		2006, McLean 2007, Millett (2007, 2008, 2009), Murray 2010, Oluwatowoju 2010, Purdy			
		2011, Roland 2009, Saxena 2007, Serumaga 2011, Shah 2011, Shohet 2007, Sigfrid 2006,			
		Simpson (2006, 2007), Smith 2008, Srirangalingam 2006, Steel 2007, Strong (2006, 2009),			
		Sutton 2009 Tabrani (2007, 2008) Tsimtsiou 2009, Vaghela 2009, Walker 2010, Whalley			
		2008 Wilkinson 2010 Williams 2006			
Hamilton 2013	An 2008, McMenamin 2003, Stevens 2005, Roski 2003	Millet (2007, 2008), Sutton 2010, Campbell (2007, 2009), Simpson (2006, 2010), Tahrani 2007,	, Salize 2009 (Germany)		
		Coleman (2001, 2007), Cupples 2008			
Heider 2020	Aviki 2018, Christianson 2008, Hamilton 2013, Hodgson 2001, Huang 2013,	Christianson 2008, Forbes 2017, Hamilton 2013, Huang 2013, Hussey 2012, Johri 2003,	Hamilton 2013 (Germany), Huang 2013 (Ireland), Hussey 2012	Johri 2003, Mendelson 2017, Wranik 2019	Mendelson 2017
	Hussey 2012, Johri 2003, Kaufman 2017, Keyhani 2013, Kondo 2018, Mendelson	Mendelson 2017. Sabatino 2012. Scott 2011. Van Herck 2010. Wranik 2019	(Belgium, Netherlanads, Italy, Sweden), Johri 2003 (Italy).		2010. Wranik 20
	2017 Nejati 2019 Sabatino 2012 Scott 2011 Steiner 1998 Van Herck 2010		Mendelson 2017 (France Netherlands Italy) Scott 2011		,
			(Germany) Van Herck 2010 (Germany, Snain, Italy)		
			(Germany), van herek 2010 (Germany, Spain, italy)		
Houle 2012	Enirbrother 2001, Grady 1997, Recentbal 2015, Eagan 2010, Reculicy 2005	Campbell 2000, Serumaga 2011, Vamer 2011, Alchamson 2012, MacBride Stowart 2008			
Houle 2012	Morrow 1995, Chung 2002, Armour 2004, Groops 2004, Kirap 2012, Young	Daran 2011, McGoyorn 2008a h. Millott 2007, Simpson (2006, 2011), Stool 2007			
	2007 Etterer 2006 Devret 2005 Concerns 2010 Colores 2007 Children 2006	Doran 2011, McGovern 2008a-b, Millett 2007, Simpson (2000, 2011), Steel 2007			
	2007, Ettner 2006, Pourat 2005, Gacagan 2010, Coleman 2007, St Jacques 2004				
Jackson 2017	Beck 2004, Chien 2012, Conrad 2013, Fagan 2010, Mandel 2007, Rosenthal	Gulliford 2007, Kontopantelis 2013, Pape 2015, Vamos 2011		LeBlanc 2017	
	2005, Young 2007				
Jia 2021	Christensen 2000, Chung 2010, Davidson 1992, Fairbrother (1999, 2001),	Clarkson 2008, Gosden 2003, Hickson 1987, Kouides 1998	Flierman 1992 (Denmark), Jensen 2014 (Denmark), Krasnik	Gray 2015, Houle 2016, Li 2013	Bilardi 2010, Gre
	Gleeson 2017, Lurie 1992, Petersen 2013,, Yesalis (1980, 1984), Young (2007,		1990 (Denmark)		
1	2012)				
Khan 2020		Mitchell 2011 Hannon 2011 Checkland 2010 Hackett 2014 Alderson 2014 McGrogor 2009	1		1
		Gill 2012 Cheraghi-Sohi 2012 Campbell 2011 Maywell 2012 Chow Crabon 2012 Loctor			
1		One 2012, Cheraghe-John 2012, Campbell 2011, Waxwell 2015, Chew-Grandini 2015, Lester	4		
1		2011, maisey 2006, Campbell 2008, Lester 2013, Cheragni-Soni 2013, NHS England, McDonald	4		
Langdown 2014		Guinford 2007, Millet (2007, 2008, 2009), Simpson 2011, Alshamsan 2012, Hippisley-Cox			
		2007, Doran 2011, Calvert 2009, Campbell 2009, Strong 2009			
Lin 2016	Beaulieu 2005, Chen 2010, Chien 2010, Coleman 2007, Fagan 2010, Kouides	Addink 2011, Anwar 2012, Campbell (2007, 2009, 2010), Crawley 2009, Cupples 2008, Dalton	Chauvel 2013 (France), Kirschner 2013 (Netherlands)		
	1998, Lester 2010, Levin-Scherz 2006, Wee 2001, Young 2007	2011, Doran 2010, Fleetcroft 2012, Gemmell 2009, Hamilton 2010, Kontopantelis 2012, Lee			
		2010, Millett (2007a-b, 2009a-c), Mindell 2011, Norbury 2011, Oluwatowojou 2010,			
1		Serumaga 2011, Simpson 2011, Tahrani (2007, 2008), Vamos (2011), Whalley (2008)			
1					
Mandavia 2017		Enver 2000 Kristonson 2014 Tabrasi (2007, 2000) Bura 2016 D	+	+	<u> </u>
		ramar 2009, Kristensen 2014, Fanrani (2007, 2008), Ryan 2016, Paper 2015, McDonald 2015,			
1		Lee 2011, Hamilton (2016, 2010), Fichera 2016, Allen 2016, Sutton (2010, 2012), Kreif 2016,			
1		Aisnamsan 2012, Vamos 2011, Millet (2007, 2009a-b), Guiliford 2007, Campbell (2007, 2009),			
1		Calvert 2009, Kontopantelis 2013, Serumaga 2011, Doran 2011, Steel 2007			
1					
Markovitz 2017	Rosenthal 2005, Mullen 2010, Chien (2010. 2011. 2012a-b. 2014a-b). Bardach	Doran (2006, 2008a-b,) Ryan 2012, Gravelle 2008. Vamos 2011. Grav 2007. Alshamsan 2012.		Li 2014, Kantarevic 2013	1
	2013 An 2008 Bhalla 2013 Dowd 2013 Chung 2010 Petersen 2013 Jba (2010	Millett 2007 Harrison 2014 Kontopantelis 2013 Hamilton 2014 Sutton 2012			
1	2012) Lindenauer 2007 Bhattachanaya 2000 Vina 2000 Shih 2014 Mahratra				
1	2007, Casalina 2007, Chan 2010, Calamar 2007, Mild 2009, Silili 2014, Wenfotta				
1	2007, Casalino 2007, Chen 2010, Coleman 2007, Damberg (2009, 2010), Felt-				
1	LISK 2007, Kyan (2010, 2011, 2012a-b, 2013), Nicholas 2011, Werner 2011,				
1	Epstein 2014, McHugh 2013, CMS 2012, MedPAC 2014, Berenson 2012, Joynt				
1	2013, Shih 2014, Gu 2014, Gilman 2014, Weinick 2010				
Mauro 2019	Hillman 1998, Wee 2001, Rosenthal (2005, 2007), Gilmore 2007, Pearson 2008,		Kirschner 2012 (Netherlands), Sicsic (2016, 2017), Constantinou	Kiran 2014, Morland 2017	Greene 2013
1	Gavagan 2010, Mullen 2010, Lester 2010, Song 2012		2017		1

	N7	Other
		Flodgren 2011, Peckham 2014, Witter 2013, Scott 2011, Chauhan 2017, Michie 2011 (UK + other countries)
		Ono 2014 (OECD contries)
		Lee 2010 (Taiwan)
		Chang (2008, 2010; Taiwan)
017, Van Herck ‹ 2019	Wranik 2019	Chaix-Couturier 2000 (NR), Hamilton 2013 (Taiwan), Huang 2013 (Argentina, Taiwan), Hussey 2012 (Japan, Taiwan), Mendelson 2017 (Taiwan, South Korea), Nejati 2019 (South Korea, Taiwan), Petersen 2006 (NR), Schatz 2008 (NR), Van Herck 2010 (Argentina)
		Twardella 2007 (Europe, NS)
Greene 2013		Lee 2010 (Taiwan), Singh 2015 (India), Twardella 2007 (Europe, NS)
		Chen 2011 (Taiwan), Lee 2010 (Taiwan), Rubinstein 2009 (Argentina)
	<u> </u>	
		Kuo 2011 (Taiwan)

First author YEAR	US	ик	Europe	Canada	Australia	NZ
Mendelson 2017	Asch 2015, Petersen 2013, Esse 2013, Friedberg 2014, Krus 2013, Lemak 2015, Rosenthal 2016, Share and Mason 2012, Young 2012, Torchiana 2013	Alshamsan 2012, Arrowsmith 2014, Gallagher 2015, Kendrick 2015, Kontopantelis 2013, Macbride-Stewart 2008, Mclintock 2014, Vamos 2011, Doran 2011, Harrison 2014, Karunaratne 2013, Sutton 2010, Calvert 2009, Hamilton 2016, Kontopantelis 2015, Millett 2009, Murray 2010, Norbury 2011, Simpson 2011, Smith 2008, Szatkowski 2011, Taggar 2012, Tahrani (2007, 2008), Vaghela 2009, Mason 2015, Kalwij 2012	Martins 2007 (Netherlands), Pechlivanoglou 2015 (Netherlands), Kirschner 2013 (Netherlands), Michael-Lepage and Ventelou 2016 (France), Sicsic and Franc 2016 (France)	Kiran 2014, Li 2014	Greene 2013	
Ogundeji 2016	Kouides 1998, An 2009, Lindenauer 2007, Gilmore 2007, Chien 2010, McMenamin 2003, Kruse 2013, Chen 2010, Coleman 2007, Glickman 2007, Bischoff 2013, Fagan 2010, Larsen 2003, Gavagan 2010, Jha 2012	Campbell 2009, Kontopantelis 2012, Sutton 2012, Simpson 2011, Srirangalingam 2006, Vaghela 2008, Calvert 2009				
Rashidian 2015		Baines 1997, Bradlow 1993, Burr 1992, Corney 1997, Doran 2011, Harris 1996, Martens 2007, Rafferty 1997, Serumaga 2011, Whynes 1997, Wilson 1995, Wilson 1999	Granlund 2006 (Sweden), Guether 1996 (Germany), Schoffski 1997 (Germany), Walley 2000 (Ireland)			
Scott 2018	Song (2011, 2012, 2013, 2014), McWilliams 3024, Sharp 2013, Chien 2014, Afendulis 2014, Jha (2010, 2012), Kruse 2012, Ryan 2012, Shih 2013, Lee 2012, Kawai 2015, McWilliams (2014, 2015), Nyweide 2015, Collar (2012, 2013, 2014) Ryan 2015	Sutton 2012, Kristensen 2014, Meacock 2014, Allen 2016, McDonald 2012, Alshamsan 2012, Gallagher 2015, Serumaga 2011, Simpson 2011, Arrowsmith 2013, Lee 2011, Kontopantelis , (2012, 2014), Harris 2014, McLeod 2015, Hamilton 2014		Cheng 2013, Kiran 2014, Li 2014		
Tao 2016	Stepanikova 2004, Le Cook 2007, Balsa 2007, Bindman 2005, Bhalla 2013, Dowc 2011	Alshamsan 2012, Millett (2007a-c, 2009), McGovern 2008, Hamilton 2010, Kontopantelis 2013, Crawley 2009, Simpson (2006, 2011), Hammouche 2011, Lee 2011, Smith 2008, Norbury 2011		Muggah 2012		
Tildy 2023	Bailey (2016, 2017, 2020), Li (2018, 2020), Marino 2016, Miraldo 2018, Parnes 2002, Thorndike 2007, Peterson 2016, Shi 2017, Fortmann 2020, Mullins 2009	Szatowski 2011, Langley (2011a-b, 2012a-b), Alageel 2019, Forster 2016, Coleman 2007, Dhalwani (2013, 2014), Farley 2017, Fichera 2016, Hardy 2014, McGovern 2008, Millett 2017, Simpson 2006, Sutton 2010, Szatowski (2010, 2011, 2016, 2021), Taggar 2012, Tahrani 2007	Vasankari 2011 (Finland), Bennett 2008 (Ireland), Firzpatrick 2011 (Ireland), Frijling 2003 (Netherlands), Pajak 2010 (Poland) Tilson 2004 (Ireland), Williams 2004 (Ireland), Donner-Banzhoff 1996 (Germany), van Doorn-Klomberg 2014 (Netherlands), Akman (Turkey), Verbiest (Netherlands) 2013		Mullins 1999, Wright 2018	
Vahidi 2013	Badgett 1997	Godsen 2002	Kristianen 1993 (Norway), Krasnik 1990 (Denmark), Grytten 2001 (Norway), Sorensen 2003 (Norway)	Lee Susan 1996, Richard 2009, Hutchison 1996		
VanHerck 2010	Kouides 1993, Bennett 1994, Morrow 1995, Fairbrother 1997, Grady 1997, Hillman 1998, Kouides 1998, Fairbrother 1999, Hillman 1999, Lebaron 1999, Safran 2000, Fairbrother 2001, Shortell 2001, Amundson 2003, Casalino 2003, Chung 2003, Larsen 2003, McMenamin 2003, Roski 2003, Armour 2004, Greene 2004, Li 2004, McMenamin 2004, Schmittdiel 2004, Beaulieu 2005, Pourat 2005, Kosenthal 2005, Shenkman 2005, Curtin 2006, Ettner 2006, Grossbart 2006, Kahn 2006, Levin-Scherz 2006, Nahra 2006, Reiter 2006, Reschovsky 2006, Rittenhouse 2006, Williams 2006, Casale 2007, Coleman 2007, Gilmore 2007, Glickman 2007, Lindenauer 2007, Mehrotra 2007, O'Malley 2007, Simon 2007, Young 2007, An 2008, Bhattacharyya 2008, Greenberg 2008, Herrin 2008, Karve 2008, Pearson 2008, Rosenthal 2008, Weber 2008, Bhattacharyya 2009, De brantes 2009, Menachemi 2009, Mullen 2009, Pham 2009, Robinson 2009, Rodriguez 2009, Ryan 2009, Vina 2009	McLean 2007, MacBride-Stewart 2008, McGovern 2008a, Langham 1995, Hippisley-cox 2004, McElduff 2004, Ashworth 2005, Fleetcroft 2006, Jaiveer 2006, Sigfrid 2006, Srirangalingam 2006, Strong 2006, Williams 2006b, Wright 2006, Ashworth 2007, Campbell 2007, Coleman 2007b, Downing 2007, Gray 2007, Gulliford 2007, Mc lean 2007, Millett 2007a-c, Saxena 2007, Shohet 2007, Simpson 2007, Tahrani 2007, Ashworth 2008, Fleetcroft 2008, Mason 2008a-d, Steel 2008, Tahrani 2008, Calvert 2009, Campbell 2009, Crawley 2009, Gemmell 2009, Strong 2009, Doran 2006, Ashworth 2007, Steel 2007, Doran 2008a-b, Vaghela 2008, Ritchie 1992, Lynch 1995, McLean 2006, Simpson 2006, Sutton 2006, Gravelle 2008, McGovern 2008b, Sutton 2009, Wang 2006	Twardella 2007 (Germany), Salize 2009 (Germany), Cattaneo 2001 (Italy), Cupples 2008 (Ireland), Gene badia 2007 (Spain), Pedros 2009 (Spain)			
Wranik 2019		Allan 2014, Charles-Jones 2003, De Stampa 2013, Maisey 2008, Shaw 2005, Afzali 2003, Coventry 2015, Davey 2005, Hann 2007, Larkin 2005, O'Neill 2008, Richards 2013 , Bradley 2008, Jesson 2003		Al Sayah 2014, Asselin 2016, Donnelly 2014, Dufour 2014, Gaboury 2009, Goldman 2010, Gucciardi 2008, Gucciardi 2015, Howard 2003, Bareil, Kotecha, Legault 2012, MacNaughton 2013, Oelke 2014, Reay 2006, Sargeant 2008, Suter 2009, Wranik 2017, Beaulieu 2013, Carter 2016, Devin 2013, Dieleman 2004, Dobson 2006, Farris 2004, Fletcher 2012, Graham 2006, Gray 2010, Haggerty 2008, Hogg 2009, Johnson 2014, Laubscher 2012, Lawson 2012, McLean 2008, O'Reilly 2007, Sellors 2003, Simpson 2011, Taylor 2005, Johnson 2011, Martin-Misener 2009, Mitton 2007, Russel 2009	Bailey 2006, Bajorek 2015, Cioff 2010, Dey 2011, Ehrlich 2011, Freeman 2012, Grimmer- Somers 2008, McDonald 2012, Merrick 2014, Nacarella 2009, Patterson 2007, Pearce 2011, Van 2011, Afzali 2003, Happell 2013, Harris 2011, McRae 2008, Proudfoot 2007, Zwar 2007, Karnon 2013	Pullon (2009, 2011), Finlayson 2012
VanHerck 2010	An 2008, Bardach 2013, Catalano (2000, 2005), Chien 2012, Hillman (1998, 1999), Petersen 2013, Roski 2003	Alshamsan 2012, Lee 2011, McLinktock 2014, Serumaga 2011				

	Australia	N7	Other
	Greene 2013		Unici
			Lee 2010 (Taiwan) Basinga 2011 (Bwanda) Kua 2011
			(Taiwan) Cupples 2008 (Ireland) Tsai 2010 (Taiwan) Chang
			2008 (Taiwan), Salize 2009 (Germany), Twardella 2007
			(Germany), Li 2010 (Taiwan)
			Chou 2008 (Taiwan), Chu 2008 (Taiwan)
			Heigh 2014 Yu 2014 Chen 2012 Cheng 2012 Lai 2013 Kuo
			2011 Li 2010 Tsai 2010 (Taiwan) Powell-lackson 2014 Yin
			2014 (China)
	Mullins 1999 Wright 2018		
	Walling 1999, Wright 2018		
96			Broomberg 1990 (S Africa), Triunfo 2009 (Uruguay)
			Rubinstein 2009 (Argentina)
Dufour 2014	Pailoy 2006 Paiorok 2015 Cioffi	Bullon (2000, 2011)	
08 Gucciardi 2015	2010 Dev 2011 Ehrlich 2011	Finlayson 2012	
, MacNaughton	Freeman 2012, Grimmer-	,	
3, Suter 2009,	Somers 2008, McDonald 2012,		
evlin 2013,	Merrick 2014, Nacarella 2009,		
letcher 2012,	Patterson 2007, Pearce 2011,		
ogg 2009, Johnson	Van 2011, Afzali 2003, Happell		
n 2008, O'Reilly	2013, Harris 2011, McRae 2008,		
05, Johnson 2011, 009	Karnon 2013		
			Rasinga 2011 (Rwanda), Bonfrer 2014 (Rurundi), Canavan
			2008 (Tanzania), Engineer 2016 (Afghanistan). Rudasingwa
			2015 (Burundi), Soeters 2008 (Burundi), Soeters 2011 (DRC),
			Yip 2014 (China)

Table A4.2: Characteristics of included reviews.

Author, year and title	Aims and Objectives	Type of review (years searched): Countries	Information sources/data bases searched	Main funding for primary care	Health, service and provider focus	Selection criteria	Aim - health or service impact, behaviour	Type - P4P, capitation, salary, blended	Basis for funding - performance measures, population coverage	Funding - funding amount, frequency	Who receives funding - service (type), provider	Service use/access - access by specific populations incl type and frequency	Quality of care - preventative care, chronic disease quality care	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of assessment	Relevance of the paper to research questions
Ahmed 2021: What drives general practitioners in the UK to improve the quality of care? A systematic literature review.	1) Explored current incentive schemes available in general practice in the UK, their impact and effectiveness in improving quality of care and (2) Identified other types of incentives in the literature	Systematic review (2009- 2019)	Pubmed/ Medline; Embase; Cochrane Library; Other: NICE Evidence, Health Management Information Consortium	Capitation	Broad range - mostly chronic disease; Primary care/general practice; Medical and Nursing	All health-related evaluation studies in UK excluding disease specific studies, opinion, non- healthcare related	Quality of care	P4P	Combination: performance, population coverage	A cross sectional study found relationship between the size of the financial incentive and expected health gain. (Fleetcroft R, Parekh-Bhurke S, Howe A, et al. The UK pay-for- performance programme in primary care: estimation of population mortality reduction. Br J Gen Pract 2010;Sep;60(578):e 345-52)	Service	Population coverage	Chronic disease, less on preventive care,	QOF found mixed results. Positive effects such as improved chronic condition management, increased patient satisfaction and improved sociodemographic inequalities were identified. Decreased quality of care in non- incentivised conditions, some dissatisfaction with tick box care.	UK, not economic databases,	Mixed Methods Appraisal tool; Nil	Outcomes in multidisciplin ary settings; Complex chronic diseases; Comparison of payment models
Bes 2023: Recruitment and retention of general practitioners in European medical deserts: a systematic review.	Compared interventions and their potential effectiveness of GP recruitment and retention in under- served areas	Systematic review (2011- 2021)	Pubmed/ Medline; Embase	Mix of capitation and fee for service	None; Primary care/general practice; Medical	Rural, remove or under-served in EU countries In English or Dutch	Retention of medical workforce	Increased or guaranteed salary; Bonus or scholarship for working in area; co- funding of additional staff; funding for transport or research; building new group practice;	Direct grant		Provider			A number of publications showed that financial and regulatory interventions (financial bonus, a mandatory placement law or capitation remuneration) had a pronounced effect on a more equal geographical distribution of GPs.	Only two databases used for search. Only English and Dutch	Joanna Briggs Institute checklists for study design; All met the JBI quality criteria for the appropriate study design, but most were of low methodologi cal quality to assess the effectiveness of the outcome	Behaviour change in primary care providers
Boeckxstaens 2011: The equity dimension in evaluations of the quality and outcomes framework: a systematic review.	To describe the evolution of pre-existing (in)equity in health care in the period after the introduction of QOF.	Systematic review (2004 to 2009)	Pubmed/ Medline; EconLit	P4P	Studies reported on a range of primary care patients (e.g., smoking cessation, diabetes, stroke). ; Primary care/general practice; Medical	P4P, QOF, contract, primary care, UK		P4P (QOF)				[Suggest: Populations: by SES, social deprivation indicators, ethnicity, national (England, UK), after introduction of QOF.	Equity in access to care evaluated by comparing users vs. non-users. None of the included studies compared the profile of users vs. non- users. No studies measured dimensions of access care, availability of equal services, equal quality of services for people in equal need, variations in personal inconvenience, cost or availability. Other outcomes included: comparing differences in treatment (e.g., treatment outcomes), or QOF scores, or 'inequity in exception'. None of the selected studies take the concept of need (defined as normative need, felt need or expressed need) into consideration.	Studies focused on equity in treatment and treatment outcomes. Studies found after introduction of QOF, quality of care improved, and for the majority of indicators, all citizens benefited from this improvement - although there was variation depending on the type of indicator, patient group, study design and detail provided on the indicator.	Heterogeneity of studies meant only systematic description possible. Selected publications use databases at the practice level and not at the patient level. Also, reported studies often use the practice as the level of analysis and/or use area level scores of deprivation as a proxy for the socioeconomic status of the patient - might not be true at individual level ('ecological fallacy'). Studies only report on data 2 years post-QOF implementation. No RCTs.	NR; NR	Complex chronic diseases; Behaviour change in primary care providers

Author, year and title	Aims and Objectives	Type of review (years searched): Countries	Information sources/data bases searched	Main funding for primary care	Health, service and provider focus	Selection criteria	Aim - health or service impact, behaviour	Type - P4P, capitation, salary, blended	Basis for funding - performance measures, population coverage	Funding - funding amount, frequency	Who receives funding - service (type), provider	Service use/access - access by specific populations incl type and frequency	Quality of care - preventative care, chronic disease quality care	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of assessment	Relevance of the paper to research questions
Bowling 2018: Provider Reimburseme nt Following the Affordable Care Act.	To study the effects of the ACA on physician reimbursemen t rates from CMS to determine the most cost- effective method of delivering health care services.	Systematic review (2009 to 2016)	Pubmed/ Medline; CINAHL; Other: Google Scholar and Google	Medicaid, FFS	None; Primary care/general practice; Medical	NA		Physician reimburseme nt rates from CMS.			Primary care provider	Increase in Medicaid-insured encounters after ACA.	Change from FFS to value-based purchasing has made it imperative for physicians to provide high-quality care and focus on decreasing readmission rates.		Potential for publication bias, research bias in synthesis, ongoing changes to healthcare associated with ACA may not have been captured.	NR.	Comparison of payment models; Behaviour change in primary care providers
Brocklehurst 2013: The effect of different methods of remuneration on the behaviour of primary care dentists.	Evaluate the effects of different methods of remuneration on the level and mix of activities provided by primary care dentists and the impact this has on patient outcomes.	Systematic review (1947- 2013)	Embase; EconLit	Departme nt of Health Chief Scientist Office, Scottish Executive and Scottish Higher Education Funding Council	Dental; Other: Primary care dentists; Dentists	Inclusion: primary care dentists providing routine dental care. RCTs, NRCTs, CBA, and ITS studies. Different payment methods (fee-for- service, fixed salary, capitation). Outcomes included clinical activity, health service use, patient outcomes, oral health inequalities, costs, and unintended effects of payment methods on practice profitability and patient satisfaction.		1. Fee-for- service remuneration: GBP6.80 for each second permanent molar fissure sealed during a six-month period and for preventive sealing of third permanent molars (the fee did not affect National Health Service capitation payments which were GBP2.76 per month to age 12 and GBP4.01 thereafter) 2. Education regarding evidence- based practice (1-day workshop) 3. Both fee- for-service and education					The first study found an increase in clinical activity related to fee- for-service payments. In the second study, dentists working under capitation arrangements restored carious teeth at a later stage in the disease process than fee-for- service controls. In the capitation arm, the dentists tended to see their patients less frequently and tended to carry out fewer fillings and extractions but tended to give more preventive advice.	Introduction of QOF benefited the aged and males. Ethnicity difficult to draw conclusion; at the level of total QOF score, ethnicity appeared to be of no influence. Small, but significant differences in deprivation with introduction of QOF favouring less deprived groups. After adjusting for practice characteristics, the impact of SES deprivation on healthcare differences was no longer evident, suggesting these differences are largely due to variations at the practice level.	Exclusion of newer econometric studies that draw causal inferences from non- random assignments, such as Chalkley 2008. Potential improvement by considering non- experimental data sources in future updates for a comprehensive evidence base.	GRADE; The number of studies using an experimental design was very low. Both included studies had a high risk of bias and the quality of the evidence from the two included studies was low/very low for all outcomes,	Comparison of payment models; Behaviour change in primary care providers

Author, year	Aims and	Type of	Information	Main	Health,	Selection criteria	Aim - health	Type - P4P,	Basis for funding	Funding - funding	Who	Service	Quality of care -	Interpretation	Limitations (identified	Methods	Relevance of
and title	Objectives	review (vears	sources/data bases	funding	service and provider focus		or service impact.	capitation, salary.	- performance measures.	amount, frequency	receives funding -	use/access - access by specific	preventative care, chronic disease quality	(authors)	by authors)	and summary of	the paper to research
		searched):	searched	primary			behaviour	blended	population		service	populations incl	care			assessment	questions
		Countries		care					coverage		(type), provider	type and frequency					
Carter 2016:	To draw	Systematic	Pubmed/	Primarv	Diabetes and	Focused on	Improved	Family			provider	nequency	In patients with	The quality of evidence	Review aimed to	GRADE	Comparison
The impact of	inferences	review	Medline;	Health	chronic	organizational	health	medicine					diabetes enrolled in a	ranged from low to	include post-2000	guidelines	of payment
primary care	whether	(2000-	Embase; Web	Care	conditions in	reforms to	system	groups,					PCN, the rate of	high. The small number	population-wide	with	models
health system	primary care	2015)	of science	Fund	Primary	Canada, namely.	based on	network					glucose monitoring was	and Ouebec suggested	longitudinal designs.	to account	
performance	reforms			(PHCTF)	care/general	the formation of	measures of						2% higher than non-	that team-based	yet diversity in	for issues of	
in Canada: a	improved			contribute	practice;	group practices,	health						enrolled patients, and	models contributed to	interventions and	selection	
systematic	health system			d \$800 million	Medical	new payment models to	service						the rate of visits to the onbthalmologist was 19	reductions in ED use.	methodologies	bias and	
Terrett.	based on			towards		support group	processes of						% higher. The rate of	care services could be	detailed meta-analyses	definitions	
	measures of			reforming		practice, or both.	care, and						cholesterol	attributed to blended	or subgroup analyses.	specific to	
	health service			primary care in		We limited the	physician						measurement was 3	capitation models and P4P in Ontario	Dependency on	our context.	
	processes of			Canadian		review to Alberta,	productivity.						15 % and 14 %	Although blended	administrative data	, Observation	
	care, and			provinces		Ontario and							decreases in the	capitation and	had limitations in	al studies	
	physician			and		Quebec where							proportion of patients	enhanced FFS practices	evaluating diversity	varied in	
	productivity.			territories		system-wide							an annual eve exam	enrolled natients and	notentially impacting	data aided	
						that meet the							after enrolling with	number of days worked	the effectiveness of	some. Team-	
						above criteria							physicians receiving	in a year.	new models in	based	
						have been							blended capitation and		differentiating	reforms in	
						provinces were							payments, respectively.		performing practices.	Quebec	
						excluded.							Patients with diabetes			showed	
													receiving		Qualitative studies	consistent	
													Family Health		from Quebec and Ontario highlighted	emergency department	
													Organizations (FHO)		varying nurse	reductions,	
													were 8 % more likely to		integration based on	rating	
													receive recommended		responsibilities,	moderately.	
													in the Family Health		for future reviews to	care studies	
													Group (FHG) enhanced		incorporate qualitative	had bias and	
													FFS model.		evidence when	indirectness,	
													in the FHG model who		addressing normative	low quality. Physician	
													were paid by enhanced		actual versus ideal	productivity	
													FFS, those in the		healthcare practices	studies had	
													blended capitation FHO		within organizations.	high-quality	
													more likely to deliver			evidence.	
													cancer screening and				
													preventive care targets.				
													traditional FFS				
													practices, statistically				
													significant increases of				
													8.5% found in delivery				
													of senior flu shots, Pap				
													smears, mammograms				
													screening, respectively				
													Significant increase of				
													4.7 % in the rate of				
													screening after the				
													introduction of the P4P				
													incentive. No changes				
													in the rates of breast				
													screening.				
													Patients treated in				
													community health				
													significant effect of				
													FHGs on patient				
													assessment of chronic				
			<u> </u>						1		1	1	illness care.	1	1		
Author, year and title	Aims and Objectives	Type of review (years searched): Countries	Information sources/data bases searched	Main funding for primary care	Health, service and provider focus	Selection criteria	Aim - health or service impact, behaviour	Type - P4P, capitation, salary, blended	Basis for funding - performance measures, population coverage	Funding - funding amount, frequency	Who receives funding - service (type), provider	Service use/access - access by specific populations incl type and frequency	Quality of care - preventative care, chronic disease quality care	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of assessment	Relevance of the paper to research questions
---	---	---	--	---	---	---	--	--	---	--	--	--	---	---	--	--	--
Emmert 2012: Economic evaluation of pay-for- performance in health care: a systematic review	identify and analyse the existing literature regarding economic evaluation of P4P	Systematic review (2000- 2010): USA, Taiwan	Pubmed/ Medline; PsycInfo; EconLit; Cochrane Library; Web of Science; Other: Business Source Complete	Fee for service	NA; Mixed; Medical, Nursing, Allied health	Experimental or observational studies including a quantitative assessment of costs and/or consequences Comparative evaluations: full economic evaluations and partial economic evaluations At least one process or outcome measure of quality	Service or health provider behaviour	Р4Р	Adherence to quality of care Referrals for lifestyle Achievement of target % of immunisation Follow up assessments Appropriate admission s, improved health status, timely discharge. Early visits for antenatal care	Variable - some lump sum (eg \$5000) some, higher fee for service if achieved target rate. Incentive size 1-2% of base reimbursements. Frequency of payments varied between immediate to over a year.	Variable - 5 service provider or 4 service	Reduce inappropriate hospitalization and admission to RAC Increased referral	Vaccination coverage According to guidelines	Efficiency of P4P is scarce and inconclusive. P4P efficiency could not be demonstrated.	Small number of included studies (9) and weak design of 5	Checklist containing 35 quality criteria - 10 categories - design, selection of alternatives, form of evaluation, benefit measuremen t and valuation, costing, modelling, adjust for timing of costs and benefits, allowance for uncertainty, presentation of results.; The methodologi cal quality of the studies was highly variable with scores ranging from 32 to 65%	Comparison of payment models
riodgren 2011: An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes	overview of systematic reviews that evaluates the impact of financial incentives on healthcare professional behaviour and patient outcomes	Review of systematic reviews (1970- 2010): USA, UK, Finland, Canada, Denmark, Germany, Ireland,	Pubmed/ Medline; Embase; EconLit; Cochrane Library; Other: DARE, TRIP, Science citation index, Social Science Citation index, NHS EED, HEED, Program and Policy Decision making	capitation or fee for service	NA; Mixed; Medical, Nursing, Allied health	SRS OF RCTs, CCTs, CBAs, and ITS reporting numerical data evaluating the effectiveness of any type of financial incentives reporting measure of health provider behaviour, health service use and healthcare costs.	health provider behaviour and patient outcomes	capitation, salary, fee for service	 Payment for working in specified time period (Salary) Payment for each service: high vs low FFS; delivery of particular care. Payment for care for a patient or specific group _mostly capitation. Payment for specified level or care or providing a change in activity or quality of care. 	of support described in one review but not in the other reviews.	service or provider	effectiveness on consultation or visit rates. Effective in improving referrals.	improved process of care including prevention. Ineffective in improving compliance with guideline outcomes	working for a specified time period, b) each service, episode or visit c) providing care for a patient or specific population, d) providing a pre- specified level or e) providing a change in activity or quality of care, was generally effective.	publication bias. 75% of evidence was at least 10 years old (now 20 years).	AIVISTAR; 2 Reviews scored highly and 2 moderately	of payment models; Behaviour change in primary care providers

Author, year and title	Aims and Objectives	Type of review (years searched): Countries	Information sources/data bases searched	Main funding for primary care	Health, service and provider focus	Selection criteria	Aim - health or service impact, behaviour	Type - P4P, capitation, salary, blended	Basis for funding - performance measures, population coverage	Funding - funding amount, frequency	Who receives funding - service (type), provider	Service use/access - access by specific populations incl type and frequency	Quality of care - preventative care, chronic disease quality care	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of assessment	Relevance of the paper to research questions
Gillam 2012: Pay-for- performance in the United Kingdom: impact of the quality and outcomes framework: a systematic review.	Examine the impact of the Quality and Outcomes Framework (QOF) on the quality of UK primary medical care, using broad inclusion criteria.	Systematic review (2004- 2011): UK	Pubmed/ Medline; Embase; PsycInfo	P4P	NA; Primary care/general practice; Medical, GP practices	All publications that sought to evaluate the QOF in the UK	To evaluate effectiveness, efficiency, equity, patient experience and provider/tea mwork of P4P QOF	Р4Р	1,000 points available to practices, and an average payment per practice in 2011-2012 of £130 (US \$204) for each point achieved	(see left)	Practice based		Significant increases in 22/23 QOF indicators for the first year but plateaued in second year. Quality of care in the third year remained higher than predicted by pre-incentive trends for 14 incentivised QOF indicators. No overall improvement effect for non-incentivized indicators in the first year of the scheme, but by the third year, improvements were significantly below those predicted by pre- incentive trends. Some evidence that disparities based on ethnicity have been reduced, with improvements in blood pressure control. Fleetcroft et al modelled a potential saving of 11 lives per 100,000 people per year aggregated across all clinical indicators and domains in the first year of the contract, with no further gain in the second year as performance for a typical practice already exceeded the target payment levels.	Large and diverse body of research, but some consistent themes have emerged. Implementation of the QOF helped consolidate evidence-based methods. QOF increased rate of improvement of quality of care during the first year of implementation, returning to preintervention rates of improvement in subsequent years. There have been modest reductions in mortality and hospital admissions in some areas, and where they have been assessed, these modest improvements appear cost-effective. The QOF has led to narrowing of differences in performance in deprived areas compared with areas not deprived. QOF strengthened team working.	Cross sectional before and after studies. The influence of many other regulatory, workforce-related, and educational changes on the quality of general practice is hard to disentangle.	Nil; NR	Comparison of payment models

Author, year	Aims and	Type of	Information	Main	Health,	Selection criteria	Aim - health	Type - P4P,	Basis for funding	Funding - funding	Who	Service	Quality of care -	Interpretation	Limitations (identified	Methods	Relevance of
and title	Objectives	review	sources/data	funding	service and		or service	capitation,	- performance	amount, frequency	receives	use/access -	preventative care,	(authors)	by authors)	and	the paper to
		(years	bases	for	provider focus		impact,	salary,	measures,		funding -	access by specific	chronic disease quality			summary of	research
		searched):	searched	primary			behaviour	blended	population		service	populations incl	care			assessment	questions
		Countries		care					coverage		(type),	type and					
				-							provider	irequency					
Hamilton	To examine	Systematic	Pubmed/		Smoking	Inclusion criteria:	Smoking	P4P QOF, FFS,	P4P QOF for	P4P QOF practices	QOF	QOF P4P studies	1) For QOF process	Almost all studies	Ten out of the 18	Modified	Complex
2013. Effectiveness	the	(1047	Embaso:		Drimony	Studies:	cessation	Donus	SHICKING	fie paid on average	EES and	in SC advice but	measures such as	in process measures	impact of the UK OOF	the Downs	diseases
of providing	offectiveness	(1947- 2011): LIK	Ellipase, Psycinfo:		care/general	and studies with	including	payments	activities are naid	2011/12	honus	other studies	status smoking	such as recording of	Findings may not be	and Black	uiseases
financial	of providing	Germany	Web of		nractice.	before and after	smoking		to general	representing a	navments	showed mixed	cessation advice and/or	smoking status	generalisable to other		
incentives to	financial	Taiwan	Science:		Medical some	design.	cessation		nractices rather	notential maximum	were mix	results Non-OOF	referral to smoking	smoking cessation	countries as the size of	studies in	
healthcare	incentives to	USA	Other:		practices were	quantitative	advice.		than individual	income for smoking	of	showed increase	cessation services. For	advice, referral to	the incentive is large	mid-range	
professionals	healthcare	03/1	GrevNet		HMOs in US	results.	referral		clinicians and	cessation work of	GPs/physic	in prescriptions.	these measures, most	smoking cessation	and is supported by	for quality	
for smoking	professionals		International,			Participants: 15	and/or		depend on the	£10 309.50 (US\$16	ians and	QOF studies	of the studies showed	services when financial	prompts from the		
cessation	on the		Open Grey			years and over,	prescription		achievement of	325). Bonus	groups of	showed mixed	an improvement after	incentives were	electronic medical		
activities:	provision and		. ,			registered with	of medication		points for quality	payments included:	physicians	results for	financial incentives	introduced. Studies	records (EMRs) and a		
systematic	impact of					any healthcare			targets in chronic	US\$5000 was		smoking rates.	were introduced, 2)	examining quit rates	comprehensive		
review.	smoking					provider.			disease	provided to			Three studies examined	had mixed results.	specialist smoking		
	cessation					Financial			management	participating clinics			the impact of financial	Those examining	cessation service. Most		
	interventions.					incentives: P4P			rather than	for achieving 50			incentives on quit rates.	system-level incentive	studies focused on		
						for individual and			improvements	referrals to a stop			There was too great a	schemes found a	process measures		
						groups of			from baseline.	smoking telephone			degree of statistical	reduction in smoking	(recording smoking		
						healthcare			There are a	advice service, then			heterogeneity for the	prevalence, but	status and advice		
						providers			maximum of	US\$25 per patient			studies to be	limitations in study	given) rather than quit		
						providing			1000 points	after the first 50;			combinable for meta-	design meant it was not	rates as outcomes, so		
						smoking			available per	US\$5000-US\$10 000			analysis even when	possible to determine	improvements may be		
						cessation advice,			practice for	if 75% of patients at			undertaken separately	whether this was due	due to improved		
						referral and/or			achieving quality	participating clinics			for QUF and non-QUF	to smokers quitting	recording rather than		
						prescription of			79 are for	status recorded and			studies ($12 > 90\%$,	management or	smoking cossistion		
						help with			smoking	if 65% of smokers			P<0.001 USINg Rev(Man)	external factors OOF	interventions		
						smoking			indicators	had been given			neviviarij.	naners were not able to	Interventions		
						cessation			navable on a	smoking cessation				look at quitting			
						Exclusion criteria:			sliding scale.	advice: US\$24 per				smoking as an outcome			
						Studies that				patient advised: and				as this is not recorded			
						examined				\$\$2432 to U\$\$152				consistently by GPs,			
						financial or other				per patient who				possibly because			
						rewards to				stopped smoking.				currently practices are			
						patients, to help								not incentivised to do			
						with smoking								so. Smoking cessation			
						cessation (unless								activity takes place			
						associated with a								outside primary care, in			
						provider financial								community pharmacies			
						incentive). Papers								and stop-smoking			
						that reported								clinics, and information			
						results as a								about quitters is often			
						composite quality								not relayed to GPs			
						score including											
						other measures											
						management if it											
						was not possible											
						to isolate impacts											
						on smoking-											
						related activities.											

not bit Objective bit Norme Description bit Description protect	Author, year	Aims and	Type of	Information	Main	Health,	Selection criteria	Aim - health	Type - P4P,	Basis for funding	Funding - funding	Who	Service	Quality of care -	Interpretation	Limitations (identified	Methods	Relevance of
Let Justice Justice Tot Justice Justic	and title	Objectives	review	sources/data	funding	service and		or service	capitation,	- performance	amount, frequency	receives	use/access -	preventative care,	(authors)	by authors)	and	the paper to
LetterLetterRestrict <th></th> <th></th> <th>(years</th> <th>bases</th> <th>for</th> <th>provider focus</th> <th></th> <th>impact,</th> <th>salary,</th> <th>measures,</th> <th></th> <th>funding -</th> <th>access by specific</th> <th>chronic disease quality</th> <th></th> <th></th> <th>summary of</th> <th>research</th>			(years	bases	for	provider focus		impact,	salary,	measures,		funding -	access by specific	chronic disease quality			summary of	research
Control Control <t< th=""><th></th><th></th><th>searched):</th><th>searched</th><th>primary</th><th></th><th></th><th>behaviour</th><th>blended</th><th>population</th><th></th><th>service</th><th>populations incl</th><th>care</th><th></th><th></th><th>assessment</th><th>questions</th></t<>			searched):	searched	primary			behaviour	blended	population		service	populations incl	care			assessment	questions
Table of the construction Material o			Countries		care					coverage		(type),	type and					
Higher Source Production Prod												provider	frequency					
Liket of number more day (mode) more day (Heider 2020:	the effects of	Other:	Pubmed/		NA.	Systematic	The effect of	2 Salary, 5 Fee	Not reported	Not reported; Bonus	All	FFS treated	Salaries showed lower	For salaries,	Lack of details on	AMSTAR; 4	Comparison
Note-Norma Note-No	Effects of	monetary	Review of	Medline;			literature review	monetary	for service, 2		payments were	provider	specific	referral rate and lower	heterogeneity of	design aspects (e.g.,	Low, 13	of payment
Bindball Secretard Secretard <th< td=""><td>wonetary</td><td>Incentives on</td><td>reviews</td><td>PSycinto;</td><td></td><td></td><td>with a</td><td>Incentives on</td><td>Bonus</td><td></td><td>Detween \$US24-152</td><td>based; P4P</td><td>conditions</td><td>activities compared to</td><td>indicators preciuded</td><td>size of incentives or</td><td>Woderate, 4</td><td>models;</td></th<>	wonetary	Incentives on	reviews	PSycinto;			with a	Incentives on	Bonus		Detween \$US24-152	based; P4P	conditions	activities compared to	indicators preciuded	size of incentives or	Woderate, 4	models;
Stepstige Stepstige <t< td=""><td>Physician</td><td>convicos and</td><td>(NOL</td><td>EconLit:</td><td></td><td></td><td>description of the</td><td>sonvicos</td><td>Pupdlod</td><td></td><td>or referred bonuses</td><td>was providor or</td><td>montal health</td><td>higher foo led to</td><td>boing drawn about the</td><td>No cloar generally</td><td>півії</td><td>chango in</td></t<>	Physician	convicos and	(NOL	EconLit:			description of the	sonvicos	Pupdlod		or referred bonuses	was providor or	montal health	higher foo led to	boing drawn about the	No cloar generally	півії	chango in
System Other level at Tavare, but on the form, but on the dense the form of the level at the server of the level at the server of the level at the server of the level at the	Groups: A	the influence		Cochrane			review process	Services	navments 10		varied from a	medical	PAP had a	increase of number of	results of the two	accentable conclusion		nrimary care
Pierwork Incertives Incertives Netherin Netherin Nethology Netherin Netherin Netherin Netherin Netherin	Systematic	of the level at	Taiwan.	Library: Web			Examined the		P4P. 5		practice bonus paid	group	positive impact	visits made by	reviews. FES had no	can be drawn from the		providers
Revenue monetary applied France Description monetary applied increments of present increments of	Review of	which	Germany.	of Science:			effects of		capitation. 3		per quarter of	based	on service	physicians vs deputies.	clear impact on	analysis of these		promacio
Incertivies (production, segments) Networks, (production, segments) Networks, (production, segments) Networks, (production, segments) Construction, (production, segments) Construction, (production, segments) <td>Reviews</td> <td>monetary</td> <td>France,</td> <td>Other: Grey</td> <td></td> <td></td> <td>monetary</td> <td></td> <td>accountable</td> <td></td> <td>approximately 5% of</td> <td>(e.g., in the</td> <td>utilisation,</td> <td>more elective</td> <td>structure or outcome of</td> <td>reviews in terms of the</td> <td></td> <td></td>	Reviews	monetary	France,	Other: Grey			monetary		accountable		approximately 5% of	(e.g., in the	utilisation,	more elective	structure or outcome of	reviews in terms of the		
applied group) dx, lenderd, b, lenderd, group) uropean b, lenderd, Suff		incentives are	Netherlan	literature:			incentives,		care		capitation through	case of		procedures,	care, process quality	effects of monetary		
Individual N Agentina, group) Obcervatory of systems and Locus, Norma, Sociand, Astron, Sociand		applied	ds, Ireland,	European			explicitly included		organisations		to year-end	"tourname		comparable health	may be negatively	incentives on quality of		
group South on Health (psy, manged (ps, manged) (psy, manged		(individual vs.	Argentina,	Observatory			the setting of				physician bonuses.	nt-based		outcomes to managed	affected by FFS	care.		
Kore, Systems and (e.g., manged prodes: sceles payments dial obsov oreall, stempate 5 Land, Health organizations organizations payments based on significant fields Sociand, Systems and (MCOL), Health organizations operations significant fields Sociand, Notice organizations operations operations significant fields Solin, Monter organizations operations operations operations operations Solin, Monter organizations operations		group)	South	on Health			physician groups				Bundled care	pay")		care. Bundled	compared with other			
Italy Policies, The care predetermined squifty of effects of Overall, the impact of Socialidad, So			Korea,	Systems and			(e.g., managed				providers receive			payments did not show	reimbursement types.			
Canada, Health organizations payments based on quality of cars; Pay boous payments (anot) Australia, Yealtry maintenance conclusions there de paide of a strained boous payments (anot) Spain, Montor, organizations conclusions there de paide of a strained boous payments (anot) Spain, Montor, organizations conclusions there de paide of a strained boous payments (anot) Spain, Gohan there de paide of a strained conclusions there de paide of a strained payments (anot) Spain, Gohan there de paide of a strained there de paide of a strained payments (anot) payments (anot) Spain, Gohan there de paide of a strained there de paide of a strained payments (anot) payments (anot) Beigin It fund, (PPool, trained a strained trained a strained on proces on payment (anot) payment (anot) Cantado, Connonvex organizations to tay de payment (anot) to tay de payment (anot) payment (anot) payment (anot) Cantado, Connonvex organizations to tay de payment (anot) to tay de payment (anot) payment (anot) payment (anot) Cantado, connonvex organizatio			Italy,	Policies, The			care				predetermined			significant effects on	Overall, the impact of			
Sociation, System sand (MCDS), Health expected costs for a showed partial quality bb cloaded cassified. Australia, Policy maintername defined episode of improvement overali. bb cloaded cassified. Swedey, Robert, Monitor, organizations care. in one SR PAP improvement overali. improvement overali. mixed season Rowed, Robert, Graham Preferrad strongly from episod episod Rowed, Robert, Graham Preferrad episod episod episod episod Rowed, Robert, Graham Preferrad episod episod episod episod episod episod Rowed, Robert, Graham Preferrad strongly from episod ep			Canada,	Health			organizations				payments based on			quality of care; P4P	bonus payment cannot			
Australia, Sweden, Sweden, Wew Policy mintechance organizations Cate: notes XP4Pa incentive size varied mindee cate: showed mindee cate: showed incentive size varied mindee cate: showed mindee cate: showed incentive size varied Japan, Wew Center, The Organizations preferred product product approximately 0.5 overall. windee cate: showed product Beigum Center, The Organizations organizations to 22.56 d approximately 0.5 overall. overall. overall. Reviews and Dissemination n [University] organizations to 22.56 d approximately 0.5 another SR PAP approximately 0.5 overall. mindee cate: showed organizations Vew Center for Reviews and Dissemination n [University] organizations categorised as one (POPS). of three shomes another SR PAP was another SR PAP was demonstrations internations was demonstrations Social Science Network, We also screamente reviews. (POPS). group practice (POPS). of three shomes par (rewarding organizations par (rewarding organizations We also screamente reviews. Social Science lists of the reviews. Social Science lists of the			Scotland,	Systems and			(MCOs), Health				expected costs for a			showed partial quality	be clearly classified.			
Spain, Note: Note: Optimizations Preference Callet: In other Sh P4P Interdiver Surviver Interdiver Surviver Structure Interdiver Surviver Preference Interdiver Preference			Australia,	Policy			maintenance				defined episode of			improvement overall.	Bundled care showed			
Japan, Graham Pretered approvide provide Zarland, Log Center, The provide approvide registrons overall provide Zarland, Zarland, Zarland, Zarland, Zarland, Zarland, Zarland, Zarland, Log opprize registrons to 12% of all overall provide It is fund, Secondaria opprize registrons to 12% of all overall process quality but impact on outcome Belgium It is fund, Contractions opprize registrons physician's total overall provide overall provide V PPOS, Inter for Accountable care comparisations another SR P4P was impact on outcome Reviews and in Opprize registrons of three schemes dimension was overall provide V PVS/Stard group practice of three schemes dimension was possible. V PVS/Stard group practice opprize registrons opprize registrons performance, based V PVS/Stard group practice opprize registrons opprize registrons performance, based V PVS/Stard group practice opprize registrons opprize registrons performance, based V PVS/Stard group practice opprize registrons opprize registrons performance, based V PVS/Stard group practice reficiency, precisere			Spain,	Nionitor,			organizations				care. In one SR P4P				mixed results for			
Wear Center, The Zealand, Beigium Center, The organizations provide organizations approximately 0.5 approximately 0.5 bit 52% of a to 12% of a physical stotal overall positive impact or process quality but impact on outcome quality was Beigium If Fund, Center for Accountable care or granizations approximately 0.5 commonwea opproximately 0.5 commonwea opproximately 0.5 composes quality but impact on outcome quality was Disseminatio Disseminatio of Vork), and Social Science (ACOS), Physician dimension was categorised as one composed as one dimension was quality was V ork), and Social Science (Coordinately 0.5) dimension was demonstrations pay (rewrding medical groups on combined performance, based paysible. Network), We also screened the reviews. Her was infinition of the reviews. infinition (Goordinately 0.5% infinition of combined performance, based paysible. I ists of the reviews. I ists of the reviews. I ists of the reviews. I ist of the reviews.			Janan	Graham			(HIVIUS), Proforrod				strongly from				quality PAP had an			
Zealand, Belgium Commonwea (Ith Fund, Centre for Reviews and Disseminatio organizations (IPPOs), Contable care organizations bb 12% of a (IPPOs), Contable care organizations on process quality but impact no utocome quality was inconclusive. No clear In University of York), and Social Science Social Science Social Science Research Network, (Economics Research Network), We also screened the reviews. (ICOS), Physician (IPOS), (IPOS), International Control of the schemes: (IPOPDS)). on process quality but impact no utocome quality was inconclusive. No clear In University of York), and Social Science Social Science Research Network, We also screened the reviews. Internations (IPOPDS)). of three schemes: tournamet-based performance, based on on clinical quality, patient anticitation, and process quality but international fee, threshold target payments (conditional on reviews. Internations (International fee), threshold target payments (conditional on reviews. Internations (International fee), threshold target payments			New	Center The			nrovider				approximately 0.5				overall nositive impact			
Beigium th Fund. (PPO). physician's total compensations, in another SR PAP was inconclusive. No clear organizations organizations organizations of three schemes: unity was inconclusive. No clear evidence on any quality was inconclusive. No clear evidence on any quality of three schemes: of York), and Social Science group practice demonstrations (PPO). of three schemes: unity was inconclusive. No clear evidence on any quality of three schemes: of York), and Social Science (PPO). group practice demonstrations (PPO). physician's total compensation. physician's total compensation. Network (PPO). group practice demonstrations (PPO). of three schemes: unity was inconclusive. No clear evidence on any quality dimension was inconclusive. No clear evidence on any quality of three schemes: unity was inconclusive. No clear evidence on any quality dimension was inconclusive. No clear evidence on any quality of three schemes: unity was inconclusive. No clear evidence on any quality of three schemes: Network (Economics Research Network). Network). Network). Network). Network inconclusive. No clear evidence on on clinical quality. patient statisfiction, and providence efficiency. practice efficiency. praproments in the school t			Zealand.	Commonwea			organizations				to 12% of a				on process quality but			
Centre for Reviews and Disseminatio Accountable care organizations (ACOS), Physician group practice demonstrations (ACOS), Physician group practice demonstrations (ACOS), Physician group practice demonstrations (ACOS), Physician group practice demonstrations (ACOS), Physician group practice demonstrations (GEOPOS). compensation, in another SPAP was of three schemes: do three schemes: dumension was possible. quality was inconclusiene. No cear evidence on any quality dimension was possible. Network/ (Economics Research Network), We also screened the reviews. GPOPS). Pay (rewarding medical groups on on on clinical on on clinical satisfaction, and particle efficiency, at approximately 3% of each physician's annual fee), threshold target poyments. Image: pay (rewarding medical groups on on on clinical on on clinical on on clinical on on clinical on on clinical on an clinical of each physician's annual fee), threshold target poyments. Image: pay (rewarding medical groups on on on clinical on on clinical on on clinical on on clinical on on clinical on on clinical on on clinical of each physician's annual fee), Image: pay (rewarding medical groups on clinical on on clinical on clinical on on clinical on on clinical on on clinical on on clinical on on clinical on c			Belgium	lth Fund,			(PPOs),				physician's total				impact on outcome			
Reviews and organizations another SR P4P was inconclusive. No clear Disseminatio (ACOS), Physician categorised as one evidence on any quality of York), and demonstrations of three schemes: dimension was of York), and demonstrations tournament-based pay (rewarding medical groups on combined tournament-based Social Science (PGPDs)). evidence on any quality dimension was Research research pay (rewarding medical groups on combined tournament-based on on clinical quality, patient satisfaction, and satisfaction, and satisfaction, and satisfaction, and satisfaction, and satisfaction, and server hereicher, at approximately 5% of each physician's annual fee), threshold target payments (conditional on in carcinal in the schemes in threshold target payments (conditional on in threshold target payments (conditarget payments (Centre for			Accountable care				compensation, in				quality was			
Disseminatio (ACOS), Physician categorised as one evidence on any quality n (University group practice of three schemes: evidence on any quality of York), and demonstrations of three schemes: bit schemes: pay(rewarding Research Research medical groups on commande, based possible. possible. Network (Economics Research group practice quality, patient satification, and Network, Ve also screened the satification, and partice efficiency, at aproximate/ 5% of each physician's satification annual fee), threshold target paymanuality I ists of the reviews. I ists of the reviews. reviews. reviews. reviews. reviews.				Reviews and			organizations				another SR P4P was				inconclusive. No clear			
Image: Second				Disseminatio			(ACOs), Physician				categorised as one				evidence on any quality			
of Vork), and demonstrations tournamet-based possible. Social Science (PGPDs)). pay (rewarding Research medical groups on combined Network performance, based on on clinical Research quality, patient Network). screened the screened the practice efficiency, reference annual fee), reviews. is of the reviews. (Conditional on				n (University			group practice				of three schemes:				dimension was			
Social Science (PGPDs)). pay (rewarding medical groups on combined Network (Economics performance, based (Economics) on on clinical Network). quality, patient We also screened the screened the reference lists of the of each physician's annual fee), threshold target reviews. (Conditional on combined				of York), and			demonstrations				tournament-based				possible.			
Network (Economics Research Network). Network). We also screened the reference lists of the reviews. reviews. annual fee), threshold target payments (conditional on				Social Science			(PGPDs)).				pay (rewarding							
Network (Economics Research Network). Ve also screened the reference lists of the reviews. reviews. annual fee), threshold target payments (conditional on conditional on				Research							medical groups on							
Research Network). We also screened the reference lists of the reviews. Here is a set of the set o				Network (Feenemies							combined							
Network). We also satisfaction, and Vereau screened the practice efficiency, reference at approximately 5% lists of the of each physician's reviews. annual fee), threshold target payments (conditional on (conditional on				Research							on on clinical							
We also screened the reference lists of the reviews. we also screened the reference lists of the reviews. satisfactice afficiency, at approximately 5% of each physician's annual fee), threshold target payments (conditional on reaching certain				Network)							guality, patient							
screened the reference lists of the reviews.				We also							satisfaction, and							
reference lists of the reviews.				screened the							practice efficiency.							
lists of the reviews. reviews. of each physician's annual fee), threshold target payments (conditional on reaching certain)				reference							at approximately 5%							
reviews.				lists of the							of each physician's							
threshold target payments (conditional on reaching certain				reviews.							annual fee),							
payments (conditional on reaching certain											threshold target							
(conditional on reaching certain											payments							
reaching certain											(conditional on							
											reaching certain							
targets) and a fixed											targets) and a fixed							
Tee for a patient											achieving a cortain							
											outcome.							

Author, year and title	Aims and Objectives	Type of review (years searched): Countries	Information sources/data bases searched	Main funding for primary care	Health, service and provider focus	Selection criteria	Aim - health or service impact, behaviour	Type - P4P, capitation, salary, blended	Basis for funding - performance measures, population coverage	Funding - funding amount, frequency	Who receives funding - service (type), provider	Service use/access - access by specific populations incl type and frequency	Quality of care - preventative care, chronic disease quality care	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of assessment	Relevance of the paper to research questions
Houle 2012: Does performance- based remuneration for individual health care practitioners affect patient care?: a systematic review.	To evaluate the effect of P4P remuneration targeting individual health care providers.	Systematic review (Database inception until June 2012): United States, Germany, UK and Canada (detail from Appendix Table 2)	Pubmed/ Medline; Embase; Cochrane Library; Other: OpenSIGLE, Canadian Evaluation Society Unpublished Literature Bank, and New York Academy of Medicine Library Grey Literature Collection		NA; Primary care/general practice; Medical			Р4Р	Performance		Individual provider		Evidence indicates that P4P has modest effectiveness, showing some improvement in preventive measures like immunisation rates, but little impact on other outcomes currently. Uncontrolled before- after studies indicated that P4P enhances adherence to chronic illness care quality indicators, but higher- quality studies with control groups or analyses accounting for secular trends did not confirm these benefits. Four large UK studies found that quality care for conditions like asthma and diabetes was already improving before the 2004 P4P scheme, with no post- implementation acceleration and no improvements in non- incentivized indicators.	While P4P may be effective in business and signal valued care aspects in healthcare organisations, current evidence does not support its widespread adoption for individual practitioners in healthcare systems. Further high-quality research is needed to assess P4P's impact on patient care, outcomes, and healthcare costs. "We believe the enthusiasm for P4P as a driver of quality improvement is disproportionate to the amount and quality of the current evidence."		Cochrane Collaboratio n tool for assessing ROB; Varied quality, but largely low to moderate. Potential for positive bias due to participant awareness of compensatio ns, inconsistent comparison groups, and lack of concealment of allocation.	Complex chronic diseases
Jackson 2017: The impact of financial incentives on the implementati on of asthma or diabetes self- management: A systematic review	To investigate the impact of financial incentives that promote implementatio n of supported self- management on quality of care including: organisational process outcomes, individual behavioural outcomes, and health outcomes for individuals with asthma or diabetes.	Systematic review (2004- 2017): US (n = 7), UK (n = 4) and Canada (n = 1)	Pubmed/ Medline; PsycInfo; CINAHL; Cochrane Library; Web of Science; Other: ScienceDirect		Diabetes and asthma; Any healthcare setting; Healthcare professionals incentivised (or whose organisation is incentivised) to provide self- management.		Implementati on of supported self- management . All schemes focused on clinician actions that were likely to improve health outcomes.	P4P - cash incentives were paid to the clinician or practice for achieving targets (n=11), one scheme used avoidance of penalty.		Incentives paid annually (n=8); quarterly (n=1);	The clinician or practice	Programme participation (n=1). Participation in a case management scheme was associated with fewer hosptialisations	Asthma study (n=1): written action plans, provision of controller medication, patients' control recorded based on national guideline. The proportion of receiving all three components of care increased from 4% to 88% after the intervention. Diabetes: proportion of patients who received HbA1c tests (n=9). The incentive had no effect (n=6); a positive effect (n=1). Diabetes: effect on HbA1c levels (n=6). A positive effect (n=2); no effect (n=2)	The impact of financial incentives for implementing self- management to patients with asthma or diabetes is inconsistent. The potential for unintended consequences should be considered.	The patient population was not considered. the heterogeneity of the studies made them difficult to compare. Studies were conducted on the UK, Canada and the US which limits generalisability. No able to complete funnel plots to measure the extent of publication bias. The initial screening of title and abstracts was conducted by a single reviewer.	NR; NR	Behaviour change in primary care providers

Author, year and title	Aims and Objectives	Type of review (years searched): Countries	Information sources/data bases searched	Main funding for primary care	Health, service and provider focus	Selection criteria	Aim - health or service impact, behaviour	Type - P4P, capitation, salary, blended	Basis for funding - performance measures, population coverage	Funding - funding amount, frequency	Who receives funding - service (type), provider	Service use/access - access by specific populations incl type and frequency	Quality of care - preventative care, chronic disease quality care	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of assessment	Relevance of the paper to research questions
Jia 2021: Payment methods for healthcare providers working in outpatient healthcare settings.	To assess the impact of different payment methods for healthcare providers working in outpatient healthcare settings on the quanity and quality of health service provision, patient outcomes, healthcare provider outcomes, cost of service provision, and adverse effects.	Systematic review; Meta analysis (Inception to present (2017, 2018 or 2019). Cochrane, Medline, Embase, WoS searched 2019. Searches rerun in 2020.): Australia, Taiwan, Canada, Germany, Denmark, USA, India.	Pubmed/ Medline; Embase; EconLit; Cochrane Library; Web of Science; Other: Dissertations and Theses Database, China National Knowledge Infrastructur e (CHKD- CNKI), Chinese Medicine Premier (Wanfang Data), IDEAS and POPLINE. Grey literature sources: OpenGrey, WHO, World Bank Trial Registries: US National Institutes of Health Ongoing Trials Register, World Health Organization International Clinical Trials Registry Platform (ICTRP)		NA; Primary care/general practice; Medical, Nursing, Allied health	- Study type: RCTs, non- randomised trials, ITS, controlled before-after studies - Participants: Healthcare providers working in outpatient care facilities (including, primary care physicians and other non- surgical specialists, dentists, midwives, nurses or allied health). - Interventions: Salary, FFS, capitation, P4P and blended payments. - Studies evaluated changes from one type of payment method to another, changes to design of a payment.	The impact of different payment methods for healthcare providers working in outpatient healthcare settings on the quantity and quality of health service provision, patient outcomes, healthcare provider outcomes, cost of service provision, and adverse effects.	P4P; fee-for- service; salary; capitation; or mix of these approaches.				Results from interventions and comparisons grouped into 4 categories: 2. FFS compared with existing payment models - The measurements of quantity of services and patient health outcomes for the studies in this comparison were heterogeneous, and it is uncertain if FFS results in an increase of outpatients and inpatients provision. 4.Enhanced FFS compared with FFS - Two randomised trials evaluated the effect of increasing FFS rate levels on the immunisation status of children, finding that compared with the low-rate group, the high- rate FFS group probably had higher immunisation rates.	Results from interventions and comparisons grouped into 4 categories: 1. P4P plus existing payment methods compared with existing payment methods - Extra funding paid by P4P probably increases up-to-date immunisation coverage for children between 3 and 35 months on Medicaid. - Extra P4P incentives may result in a slight increase in pharmacists asking more detailed questions on patients' diseases. - Uncertain of the effects of adding P4P to existing payment methods on influenza vaccination rates amongst the outpatient elderly, as well as blood pressure control, compared with existing payment methods. 2. FFS compared with existing payment models - Uncertain of the effect of FFS on patient health outcomes compared with capitation or salary. 3.FFS mixed with existing payment methods compared with existing payment methods on patient ewith existing payment methods - Data from only 1 RCT: uncertain of the effect of FFS mixed payment methods on patient outcomes, measured as smoking abstinence behaviour.	"For healthcare providers working in outpatient healthcare settings, P4P or an increase in FFS payment level probably increases the quantity of health service provision (moderate- certainty evidence), and P4P may slightly improve the quality of service provision for targeted conditions (low-certainty evidence)."		EPOC 'risk of bias' criteria; Largely unclear ROB; 10/12 RCTs judged as unclear ROB, other 2 high ROB. 8/13 controlled before and after were unclear ROB and 5 studies judged as high ROB. Other 2 studies unclear ROB.	Comparison of payment models
pay for performance scheme in primary care: Meta- synthesis of qualitative studies on the provider experiences of the quality and outcomes framework in the UK.	meta-synthesis of qualitative studies on the Quality and Outcomes Framework (QOF) to identify themes on the impact of the QOF on individual practitioners and other professionals in primary care; to	synthesis (2004 - 2018): UK	Medline; Embase; CINAHL; Web of Science; Other: Healthstar		care/general practice; Medical, Nursing, Other: Groups of professionals in primary care.	qualitative research methods; focused on the QOF; focused on the UK.								incongruent with values such as self- direction and benevolence values that are pivotal to professionalism; but compatible with the pursuit of wealth, authority, success and ambition, linked to personal success. Creativity, social justice, equality, benevolence were deemed restricted because of QOF targets.	have access to any raw data.		change in primary care providers

Author, year and title	Aims and Objectives	Type of review (years searched): Countries	Information sources/data bases searched	Main funding for primary care	Health, service and provider focus	Selection criteria	Aim - health or service impact, behaviour	Type - P4P, capitation, salary, blended	Basis for funding - performance measures, population coverage	Funding - funding amount, frequency	Who receives funding - service (type), provider	Service use/access - access by specific populations incl type and frequency	Quality of care - preventative care, chronic disease quality care	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of assessment	Relevance of the paper to research questions
	that will be useful for decision- makers in designing and implementing new incentive schemes.																
Langdown 2014: The use of financial incentives to help improve health outcomes: is the quality and outcomes framework fit for purpose? A systematic review.	Examines the evidence on the efficacy of the quality and outcomes framework (QOF) for improving health outcomes, its impact on non- incentivized activities and the robustness of the clinical targets adopted in the scheme.	Systematic review (2004 - 2012): UK	Pubmed/ Medline; Embase; EconLit; Cochrane Library; Web of Science; Other: Health Management Information Consortium	P4P - quality and outcomes framework (QOF)	Asthma, CHD, diabetes; Primary care/general practice; Medical		health outcomes	quality and outcomes framework (QOF)			GPs		Significant increase in the achievement of intermediate health outcome targets (blood pressure ≤145/85 mmHg; cholesterol <5 mmol/l; HbA1c ≤ 7.4%)	The evidence from the results suggests that the QOF has led to an improvement in health outcomes for some conditions including Diabetes, although the results are mixed for others such as CHD.	The review's methods and selection criteria could have been too specific, excluding studies which may have contributed to the findings. Another limitation of the study is that the review was conducted by one individual, which may also have impacted on the selection of abstracts and the quality assessment conducted.	CASP tool; No summary provided (detail in suppl)	Complex chronic diseases; Behaviour change in primary care providers
Lin 2016: Impact of pay for performance on behavior of primary care physicians and patient outcomes	To assess the impact of pay- for- performance on behavior of primary care physicians and patient outcomes aiming to provide a comprehensive and objective evaluation of pay-for- performance	Systematic review (1980- 2013): UK, USA, France, Taiwan, Ireland, Netherlan ds, Argentina,	Pubmed/ Medline; Embase; Cochrane Library	Variable - capitation and fee for service	Various: Multiple, Hypertension, Diabetes, Preventive care, Cancer Screening, CVD, stroke, COPD, Influenza vaccination, Asthma, Hypothyroidis m, severe mental illness, Smoking, Epilepsy, ; Primary care/general practice; Medical, Nursing, Allied health	The inclusion criteria: Study design: randomized controlled trials (RCTs), controlled before-after studies (CBAs), interrupted time series (ITSs), or cross-sectional studies (CSs). (ii) Participants: primary care physicians or physician groups, or patients being taken care of by primary care physicians. (iii) Intervention: P4P programs in which all or part of physician income depended on the quality of medical services they provided. (iv) Control: studies reported the effect of P4P compared to a control group without P4P program, or baseline data before implementation of P4P. (v) outcomes: change in behaviour of primary care physicians like prescription, order for tests,	Primary care physicians behaviour and patient outcomes	Р4Р	Variable: when performance reaches certain level, for each individual reached (eg fully immunised)	Variable 1.5 to 10% of practice income; ,	Mainly practice	Service use - eg immunisations.	Preventive 12+ and 4 - Hypertension 13+ and 2- CHD 16+ Diabetes 28+ COPD 5+ Asthma 7+ Stroke 6+	An overall positive effect was found on the management of disease, which varied in accordance with the baseline medical quality and the practice size.		Quality Index (I); Range of scores 14-25 <20: 11 20+: 33	Complex chronic diseases

Author, year and title	Aims and Objectives	Type of review (years searched): Countries	Information sources/data bases searched	Main funding for primary care	Health, service and provider focus	Selection criteria	Aim - health or service impact, behaviour	Type - P4P, capitation, salary, blended	Basis for funding - performance measures, population coverage	Funding - funding amount, frequency	Who receives funding - service (type), provider	Service use/access - access by specific populations incl type and frequency	Quality of care - preventative care, chronic disease quality care	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of assessment	Relevance of the paper to research questions
						admission, referral, and outcome measures like blood pressure or serum lipid level) concerning the specific disease, including management of diseases, equity, satisfaction of patients, cost of health care, baseline with indicators, and practice size. Exclusion criteria (i) Review articles or perspective articles with no detailed data; (ii) Studies about specific P4P settings and design, without reporting effect of P4P programs; (iii) Studies which were updated and replaced by later studies.											
Mandavia 2017: Effectiveness of UK provider financial incentives on quality of care: a systematic review.	To review the UK evidence on whether provider financial incentives are an effective way of improving the quality of health care.	Systematic review (1980- 2016): UK	Pubmed/ Medline; Embase		CVD -2; Diabetes-7; Smoking 1, Cholecystecto my, ; Mixed; Medical	All included articles assessed financial incentives as the independent variable, and quality of health care as the dependent variable. Articles were excluded if there was no comparison group or baseline analysis before the intervention.	Improved quality of care against QOF indicators	P4P	Performance measures		Service		Of 28 studies, nine reported positive effects of incentives on quality of care, 16 reported intermediate effects, two reported no effect, and one reported a negative effect.	Of 28 studies included, 9 reported positive effects on quality of care, 16 reported intermediate effects, two reported no effect, and one reported a negative effect.	The generalisability of findings is limited, with the majority of studies focusing on the QOF incentive.	The quality of included papers was assessed using the quality assessment checklist published by Downs and Black; Quality scores ranged from 15 to 19, out of a maximum of 22 points	Comparison of payment models; Behaviour change in primary care providers
Markovitz 2017: Pay-for- performance: Disappointing results or masked heterogeneity ?	To assess whether hospital and physician performance in P4P vary by patient and catchment area factors, organizational and structural capabilities, and P4P program characteristics.	Systematic review (2012- 2017): UK 13, Canada 2, USA 42	Pubmed/ Medline; Google scholar; Other: RAND review of P4P publications (Damberg et al., 2014) Reports by Medicare Payment Advisory Commission (MedPAC), Kaiser Permanente, and the Commonwea Ith Fund.	UK Capitation US Fee for service	NA; Mixed; Medical	Evaluated programs that provided financial bonuses or penalties for performance on prespecified measures; program was implemented in the United States, the United Kingdom, or Canada; provided qualitative or quantitative empirical analysis; was conducted among physician practices and hospitals: and (p)	Mixed results	P4P	Performance clinical recording and outcomes	Inconsistent and contradictory effects of bonus likelihood, bonus size, and marginal costs on performance	Mixed		Improvement variable - least improvement for low SES groups	P4P has largely failed to realize substantial quality gains. Organizations with the most to gain financially did not necessarily improve; those that improved the most did not necessarily stand to receive large financial awards; the tasks on which institutions improved were not necessarily the easiest; nor were the institutions that did improve those with the lowest marginal costs.	Difficult to disentangle the effects of baseline quality and true P4P responsiveness.	Informal; Maybe publication bias	Comparison of payment models

Author, year	Aims and	Type of	Information	Main	Health,	Selection criteria	Aim - health	Type - P4P,	Basis for funding	Funding - funding	Who	Service	Quality of care -	Interpretation	Limitations (identified	Methods	Relevance of
and title	Objectives	review (years searched): Countries	sources/data bases searched	funding for primary care	service and provider focus		or service impact, behaviour	capitation, salary, blended	- performance measures, population coverage	amount, frequency	receives funding - service (type),	use/access - access by specific populations incl type and	preventative care, chronic disease quality care	(authors)	by authors)	and summary of assessment	the paper to research questions
Author, year and title Mauro 2019: Effect of financial incentives on breast, cervical and colorectal cancer screening delivery rates: Results from a systematic literature review.	Aims and Objectives	Type of review (years searched): Countries Systematic review (Inception to Feb or March 2018 [included studies 1998 to 2017]): United States, Taiwan, The Netherlan ds, Australia, Canada, France	Pubmed/ Medline; Google scholar; Cochrane Library; Web of Science	Main funding for primary care	Health, service and provider focus Screening for breast, cervical and colorectal cancer.; Primary care/general practice; Medical	Selection criteria reported sub analyses by patient and catchment factors, organizational structure and capabilities, or P4P program characteristics. Inclusion criteria: Participants: Primary care healthcare providers; medical practice/facility teams. Intervention: Explicit financial incentives tied to measured quality-of-care targets, impacting income or quality improvement investment; performance based on achievement/imp rovement. Comparator group or period, including physicians or groups in/out of incentive programs, cancer patients in/out of P4P programs, and pre/post	Aim - health or service impact, behaviour Test whether P4P improves delivery rates of preventative care, namely breast, cervical and colorectal cancer screening.	P4P	Basis for funding - performance measures, population coverage	Funding - funding amount, frequency	Who receives funding- service (type), provider Provider	Service use/access - access by specific populations incl type and frequency For breast cancer screening, most of the studies (9[/13 studies reporting on BCS]) showed partial or no effects; one explanation could be that women may take a proactive role in breast cancer screening, making physician incentives less important. For cervical cancer screening, 6 studies showed positive effects, 5 no effect, 3 partial effects, and 1 negative effects [of 15 studies reporting on CerCS]. Few positive or irrelevant effects were found regarding	Quality of care - preventative care, chronic disease quality care	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of assessment Not formally assessed.	Relevance of the paper to research questions
						financial incentive screening rate trends. Outcome: Quantitative measures of clinical effectiveness using validated statistical methods (e.g., randomized trials, longitudinal analysis). Study Design: Experimental, quasi- experimental, or observational studies focusing on incentives for breast, cervical, and colorectal cancer screening in primary care, with quantitative assessment of P4P effects on screening rates.						colorectal cancer screening [of 7 studies reporting on CoICS)					

Author, year and title	Aims and Objectives	Type of review (years searched): Countries	Information sources/data bases searched	Main funding for primary care	Health, service and provider focus	Selection criteria	Aim - health or service impact, behaviour	Type - P4P, capitation, salary, blended	Basis for funding - performance measures, population coverage	Funding - funding amount, frequency	Who receives funding - service (type), provider	Service use/access - access by specific populations incl type and frequency	Quality of care - preventative care, chronic disease quality care	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of assessment	Relevance of the paper to research questions
Mendelson 2017: The Effects of Pay- for- Performance Programs on Health, Health Care Use, and Processes of Care: A Systematic Review.	The purpose of the current review is to update and expand the prior systematic review in order to summarize current understanding of the effects of P4P programs targeted at physicians, groups, and institutions on process-of-care and patient outcomes in ambulatory and outpatient settings in and outside the United States.	Systematic review (June 2007 to October 2016): United Kingdom (27 studies), the United States (17 studies), Taiwan (13 studies), France (3 studies), the Netherlan ds (3 studies), Canada (3 studies), Australia (1 study), South Korea (1 study), and Italy (1 study)	Pubmed/ Medline; PsycInfo; CINAHL; Scopus; Other: Business Economics, Theory, Business Source Elite, Faulty of 1000, Gartner Research		NA; Primary care/general practice; Medical	Adult patients that evaluated ambulatory care or hospital-based P4P programs targeting health care providers at the individual, group, managerial, or institutional level; Reported any process-of-care, utilisation, health, or intermediate health (clinical measures, such as a laboratory value or blood pressure) outcome. We included studies from other countries that have health systems similar to portions of the U.S. health care system; Clinical or cluster randomised, controlled trials (RCTs) of any size, plus observational studies limited to those with a comparison group, interrupted time series (ITS) studies, or large (n > 10 000) cross-sectional or uncontrolled before-after studies.		Р4Р					Low-strength evidence that P4P programs may improve process-of- care outcomes over the short term (2-3 yrs) in ambulatory settings. Evidence on long-term effects was limited. Many of the studies reporting positive findings were conducted in the United Kingdom, where incentives were much larger than any P4P programs in the United States. The largest improvements were seen in areas where baseline performance was poor. Low-strength evidence that P4P had little to no effect on intermediate health outcomes (changes in laboratory measures), though there were inconsistencies among study results. Evidence examining patient health outcomes was insufficient because few methodologically rigorous studies reported these outcomes. In the hospital setting, low- strength evidence showed that P4P had a neutral effect on patient health outcomes and a positive effect on reducing hospital readmissions.	"We found low- strength, contradictory evidence that P4P programs could improve processes of care, but we found no clear evidence to suggest that they improve patient outcomes."		Newcastle- Ottawa Scale for observationa I studies, the Cochrane Risk-of-Bias tool for RCTs; 2 RCTs were found to have low ROB; observationa I studies largely demonstrate d high ROB (Checked using Supplementa I material)	Comparison of payment models
Ogundeji 2016: The effectiveness of payment for performance in health care: A meta- analysis and exploration of variation in outcomes.	This paper systematically explores the extent and sources of heterogeneity in the results of evaluations of P4P schemes to identify features associated with success in P4P schemes.	Meta analysis (Up to 2016 (no date restrictions)): Majority of studies based in UK and USA (approx 26 and 25 but total numbers can't be calculated from the s6) - at least one study from the following countries: China,	Pubmed/ Medline; Cochrane Library; Other: Database of Abstracts and Reviews of Effect(DARE), National Health Service Economic Evaluation Database (NHS EED), Health Technology Assessment (HTA)	P4P	NA; Primary care/general practice; Other: Hospital settings; Medical			P4P, which was stratified by authors using self- made typology tool (see S7)		Incentives were small (<5% of the individuals salary, or payments such as total budget for the healthcare team/facility/hospit al for an intervention, fee for service and capitation) or large (>or equal to 5% of the individuals salary, or payments such as total budget for the healthcare team/facility/hospit al for an intervention, fee for service and capitation)	Individuals and groups (MDT, charities, community groups, practices)		Multi variate regression (all 96 studies) found no effect in improving health outcomes (e.g., hospital mortality rates, smoking cessation). However larger effects were found for intermediate (e.g., blood pressure or cholesterol reduction) and process outcomes (e.g., cancer screening, smoking cessation advice).		Estimates of the effectiveness of incentive schemes on health outcomes are probably inflated due to poorly designed evaluations and a focus on process measures rather than health outcomes.	AMSTAR; NR	

Author, year and title	Aims and Objectives	Type of review (years searched): Countries	Information sources/data bases searched	Main funding for primary care	Health, service and provider focus	Selection criteria	Aim - health or service impact, behaviour	Type - P4P, capitation, salary, blended	Basis for funding - performance measures, population coverage	Funding - funding amount, frequency	Who receives funding - service (type), provider	Service use/access - access by specific populations incl type and frequency	Quality of care - preventative care, chronic disease quality care	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of assessment	Relevance of the paper to research questions
		Rwanda, Tanzania, Turkey, Zambia, Israel, Taiwan, Uganda, Australia, Netherlan ds, Canada, Italy, Haiti, Malawi															
Rashidian 2015: Pharmaceutic al policies: effects of financial incentives for prescribers.	To determine the effects of pharmaceutical policies using financial incentives to influence prescribers' practices on drug use, healthcare utilisation, health outcomes and costs (expenditures).	Systematic review; Meta analysis (Variable till about 2 Feb 2015): 9 UK, 1 Ireland, 2 Germany, 3 UK and Netherlan ds, 1 Sweden, 2 Taiwan	Pubmed/ Medline; Embase; EconLit; PAIS; WoS; Other: Effective Practice and Organisation of Care Group Register, EBM Reviews, the Cochrane Central Register of Controlled Trials, CSA Worldwide Political Science Abstracts, SIGLE, System for Information on Grey Literature in Europe, INRUD, International Network for Rational Use of Drugs, International Network for Rational Use of Drugs, International Political Science Abstracts, NHS EED, National Health Services Economic Evaluation Database, CRD, NTIS, National Technical Information, IPA, International Pharmaceutic al Abstract Global Jolis,JOLIS, World Bank Library, WHO OECD.		NA; Community programs; Medical	Participants: Healthcare consumers and providers within a large jurisdiction / system of care (regional, national or international). Studies within organisations, such as health maintenance organisations, were included. Interventions: Prescribing policies that intend to affect prescribing by means of financial incentives, and other policies specifically targeted at prescribing or drug utilisation.	Service impact	Drug Budget policies, P4P, Reimburseme nt Rate Reduction Policy	Change in prescribing patterns/drug use - especially to prescribe fewer or cheaper medicines P4P Targets for these policies include administrative goals, waiting time, patient satisfaction and diagnostic and treatment goals.	P4P Interventions vary greatly in terms of implementation approaches, magnitude of the 'incentives' (e.g. from 2% to 25% of physician total earnings) and whether accompanying interventions are included	Providers	Drug Budget - Drug use (item per patient or prescription) ۥ2.8% (ۥ28.9 to 1.5); Drug use (generic percentage) 15% (ۥ43.7 to 190.5). This policy may lead to a modest reduction in overall drug use per patient (low certainty evidence) P4P - Drug use Range 2.5 to 2.6	P4P - Mean ― 1.49% (95% Cl ― 6.32 to 3.34) [1 comparison (percentage of patients with controlled blood pressure) from 1 setting]	Limited studies on budgetary policies, P4P and reimbursement rate reduction policies from six countries met the inclusion criteria. The certainty of the evidence is low, to lacking for all financial incentives. There is low to very low certainty evidence that drug budgets decreasing drug use and costs. Effects of P4P on improving quality of care and health outcomes are uncertain.	Investigators in the included studies did not pay enough attention to potential side effects of the policies. Also we were not able to identify any evidence to assess the applicability of the review findings to disadvantaged groups. None of the included studies were randomised trials. Hence for all included policies, selection bias may occur, especially when policies are implemented on a voluntary basis or on the basis of presumed 'readiness' for policy implementation. Risk of selection bias for all included CBA study results might lead to overestimation of the effects.	Risk of Bias based on EPOC group (multiple tools); *Note - Interpretatio n of their reporting as the authors did not summarize their RoB assessment 16 High Risk of Bias 2 Medium Risk	Comparison of payment models; Behaviour change in primary care providers

Author, year and title	Aims and Objectives	Type of review (years searched): Countries	Information sources/data bases searched	Main funding for primary care	Health, service and provider focus	Selection criteria	Aim - health or service impact, behaviour	Type - P4P, capitation, salary, blended	Basis for funding - performance measures, population coverage	Funding - funding amount, frequency	Who receives funding - service (type), provider	Service use/access - access by specific populations incl type and frequency	Quality of care - preventative care, chronic disease quality care	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of assessment	Relevance of the paper to research questions
Scott 2018: Financial incentives to encourage value-based health care.	recent (since 2010) empirical literature that evaluates the effect of value- based purchasing in health care.	Review (): United States, United Kingdom, Taiwan, China, Canada, Italy, Australia, France, the Philippines , Rwanda	Pubmed/ Medline; EconLit; Other: screening the citations of included articles and previous review s, forward citation tracking of previous reviews and included studies, searching websites of authors frequently publishing in the field, and grey literature searches of key websites		Value-based health care; Primary care/general practice; Mixed; Medical	studies were included if: they examined the impact of schemes on any type of outcome (e.g., costs, utilization, expenditures, quality of care, health outcomes). If incentives were targeted at individual or groups of medical practitioners or hospitals. Study designs: Before and after studies and controlled before and after studies that used regression analysis to attempt to control for confounders were included. Studies that used interrupted time series designs, difference-in- difference designs, and randomized controlled trials were included. Exclusions: observational and modelling studies with no control group, as were before and after studies and controlled before and after studies that only compared the means of primary outcomes with no attempt to adjust for confounders.	value-based purchasing in health care	P4P, incentives for performance combined with incentives for reducing costs	measures	19 primary care/medical group schemes reported size of payments. 4 were below 5% and 8 were 5-10%, and 7 were 10-30% of payments.	ndividual medical profession al, hospital		Improved quality of care in about half the studies	differences between schemes combining pay for performance with rewards for reducing costs, relative to pay for performance schemes alone. Paying for performance improvement is less likely to be effective. Allowing payments to be used for specific purposes, such as quality improvement, had a higher likelihood of a positive effect, compared with using funding for physician income. Finally, the size of incentive payments relative to revenue was not associated with the proportion of positive outcomes.	 The search terms used were limited and a full critical appraisal of the quality of the studies and their risk of bias was not conducted. Vote counting was used rather than sum- arising effect sizes. Unobserved factors may have contributed to success of programs. 	Not assessed; Not assessed	of payment models

Author, year and title	Aims and Objectives	Type of review (years	Information sources/data bases	Main funding for	Health, service and provider focus	Selection criteria	Aim - health or service impact,	Type - P4P, capitation, salary,	Basis for funding - performance measures,	Funding - funding amount, frequency	Who receives funding -	Service use/access - access by specific	Quality of care - preventative care, chronic disease quality	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of	Relevance of the paper to research
		searched): Countries	searched	care			behaviour	blended	coverage		service (type), provider	populations incl type and frequency	care			assessment	questions
Tao 2016: The	Impact of	Systematic	Pubmed/		Diabetes, CVD,	Inclusion: Studies		4 Capitation				Capitation vs FFS	P4P -Diabetes - mixed	Few studies addressed	The heterogeneity of	Instruments	Complex
impact of	reimbursemen	review (1st	Medline;		respiratory,	with		and FFS, 1				- Increase primary	results. Levels of mean	this. No evidence	the outcomes	developed	chronic
reimburseme	t system on	of January	Web of Scionco		multiple diagnosos:	experimental or		Capitation and				care access	HDA1C, total	ravouring one	complicates synthesis	by Zaza et al.	diseases;
equity in	and racial	30th of	Science		Primary	designs		Canitation 1					systolic blood pressure	over another as regards	reimbursement system	University of	of navment
access and	inequalities in	September			care/general	conducted in		Capitation,					were decreasing in all	impact on	might have differential	Manchester	models
quality of	access,	2013): 15			practice;	primary care		FFS and					ethnic groups prior to	socioeconomic or racial	impact on equity	Center for	
primary care:	utilization and	UK, 6 US, 1			Medical,	settings. Only		Salaried, 16					QOF. Existing ethnic	inequalities in access,	depending on the	Occupational	
A systematic	quality of	Canada			Other: Not	studies from		P4P					disparities remained	utilization and quality	outcome under study.	and	
literature	primary care.				explicitly	high-income							unchanged after the	of primary care.	Patient characteristics	Environment	
review.					noted	countries were							CVD - mixed results for	neterogeneity	other than	al Health;	
						Reimbursement							whites and blacks. In	comparisons.	socioeconomic status	Medium - 18	
						systems assessed							contrast, a higher	Outcomes measured	might also influence	Low	
						were capitation,							proportion of south	(e.g. changes in	the effect of		
						fee-for-service							Asians achieved target	laboratory values or	reimbursement		
						and pay-for-							levels of blood pressure	blood pressure) may	systems and confound		
						performance. All							in 2003, and the	not be sensitive to	the results in these		
						outcomes were of							2005 Respiratory - no	capturing inequalities.	the effect of a given		
						interest.							change in inequity. In a		reimbursement system		
													serial cross-sectional		might be context-		
													study from the UK, the		specific and vary		
													introduction of QOF		between different		
													and new clinical		health care systems,		
													guidelines for COPD		funded systems versus		
													registration of		private health		
													spirometry data and		insurance systems. Our		
													use of combination		review was dominated		
													inhalers for all patients,		by studies from the US		
													with no differences		and the UK, which		
													between deprivation		limits generalizability		
													diagnosis - no change in		hetween studies since		
													inequity; Preventive		the structure of the		
													care - no change in		health systems differ		
													inequity		widely between these		
															countries. Additionally,		
															studies from the US		
															vulnerable to selection		
															bias since patients		
															choose their insurance		
															plan. None of the		
															studies distinguished		
															petween horizontal		
															caveat of studying		
															health care reforms		
															are the natural		
															changes that would		
															have occurred over		
															time regardless of the		
															reform, which could be		
															time trends into		
															account, but few		
															studies applied this		
															method.		

Author, year	Aims and	Type of	Information	Main	Health,	Selection criteria	Aim - health	Type - P4P,	Basis for funding	Funding - funding	Who	Service	Quality of care -	Interpretation	Limitations (identified	Methods	Relevance of
and title	Objectives	review	sources/data	funding	service and		or service	capitation,	- performance	amount, frequency	receives	use/access -	preventative care,	(authors)	by authors)	and	the paper to
		(years	bases	for	provider focus		impact,	salary,	measures,		funding -	access by specific	chronic disease quality			summary of	research
		searched):	searched	primary			behaviour	blended	population		service	populations incl	care			assessment	questions
		countries		care					coverage		nrovider	frequency					
Tildy 2022	Idontify	Systematic			Smoking	Focus on	Sonvice and		Poodinoss to quit'		provider		Three studies indicated	1/24 studios woro of	roviow ovaluated		Comparison
Implementati	implementatio	systematic	Medline:		cessation.	implementation	Health		and 'counselling			not change	an increase in cessation	4/54 studies were of	observational studies	(Risk Of Bias	of navment
on strategies	n strategies	Qualitative	Embase:		Primary	strategies which	lication		given' fields to			overall	while one showed no	Evidence that	which, whilst at risk of	In Non-	models:
to increase	aiming to	synthesis	PsycInfo;		care/general	were nation- or			the vital sign			prescribing of	effect.	incentives increase	bias and unable to	randomized	Behaviour
smoking	increase	(inception	CINAHL;		practice;	state-wide to test			section of the			smoking	FORMULARY In the	smoking cessation is	demonstrate causality,	Studies of	change in
cessation	smoking	to April	Other: Global		Medical,	scalability.			medical record,			cessation	USA, one study	unclear. A national	can provide evidence	Intervention	primary care
treatment	cessation	2021):	Health, Social		Nursing, Allied	Inclusions: Any			incentivise			medications. For	(Massachusetts) found	financial incentive for	of real-world	s); 24 Serious	providers
provision in	treatment	UK23, USA	Policy &		health	type of			recording of			practitioner-level	no difference in quit	GPs increased	implementation. A	20 Moderate	
primary care:	provision in	(13),	Practice,			healthcare			smoking status			outcomes, in the	attempts, but two	recording of smoking	large number of	4 Low	
review of	their	the	Annlied			within family						access to health	multi-state) found a	advice and (one study)	in the evidence		
observational	effectiveness.	Netherlan	Social			medicine or						insurance	positive effect on	referral to cessation	synthesis, however.		
studies.	cost-	ds (3) ,	Sciences			general medical						coverage which	smoking cessation	services. Practices	only half were at		
	effectiveness	Australia	Index and			practice, that						included smoking	following the increases	receiving funding to	moderate or low risk of		
	and any	(2) Turkey	Abstracts			included the						cessation	in medication	deliver national CVD	bias. Despite an		
	perceived	(1), Poland				whole practice						treatment,	prescribing. The Dutch	prevention increased	international scope,		
	facilitators and	(1),				patient						increased	study indicated	smoking status	most studies were set		
	barriers for	Finland (1),				population.						smoking status	Increased cessation,	recording, cessation	In the UK and the USA.		
	enectiveness.	compared				health						state Oregon)	was less robust	medication and	intervention involved		
		different				interventions						cessation advice	Patient-level outcomes	cessation.	multiple		
		policies in				delivered outside						provision	were not measured in		implementation		
		the				primary care,						(Colorado) and	the studies assessing		strategy categories and		
		Germany				dental practices						cessation	the introduction of new		it was challenging to		
		and the UK				or pharmacies. If						medication	medications.		disentangle their		
						outcome data						prescribing	INCENTIVE one study		individual effects.		
						extracted						(Oregon, muiti- state) In the	cessation too. In				
						exclusively for the						Netherlands.	contrast, the only study				
						primary care						increasing health	assessing a patient-				
						setting.						insurance	level outcome of the				
												coverage for	QOF 2004 found no				
												smoking	effect on cessation.				
												cessation also					
												Increased					
												medication					
												prescribing.					
												INCENTIVE Two					
												studies found					
												that there was no					
												OOF on cessation					
												medication					
												prescribing while					
												one found an					
												increase in					
												cessation					
												prescribing. The					
												one study					
												investigating the					
												2012 QOF					
												amendment					
												in the provision					
												of cessation					
												advice and					
												referrals to NHS					
												SSS, but no					
												increase in					
												medication					
												prescribing.					
												CAPITATION					
												found no effect					
												on cessation					
												advice provision,					
												medication					
												prescribing.					

Author, year and title	Aims and Objectives	Type of review (years searched): Countries	Information sources/data bases searched	Main funding for primary care	Health, service and provider focus	Selection criteria	Aim - health or service impact, behaviour	Type - P4P, capitation, salary, blended	Basis for funding - performance measures, population coverage	Funding - funding amount, frequency	Who receives funding - service (type), provider	Service use/access - access by specific populations incl type and frequency	Quality of care - preventative care, chronic disease quality care	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of assessment	Relevance of the paper to research questions
Vahidi 2013: A systematic review of the effect of payment mechanisms on family physicians service provision and referral rate behaviour	to discuss the impact of different methods of payment to family physicians and general practitioners, quantity of service provision and referral rate behaviour.	Systematic review (1985- 2011): 3 were from Norway, 3 from Canada and rest were from the United States, England, South Africa, Denmark and Uruguay	Pubmed/ Medline; Other: Science Direct, Emerald, Wiley Inter Science, Springer Link and ANNFAMME D published	Fee for service	Variable, one focused on caesarean section; Primary care/general practice; Mixed; Medical	Assessed effect of three basic types of payment mechanisms (salary, FFS, capitation) on physician behaviour; Addressed confounding factors (by adjusting), had >60% response rate; Used valid data sources; Reported quantitative results, effects, or impacts of payment mechanism on family physician program and general practitioners' behaviour		Capitation Salary FFS	FFS system, healthcare providers are reimbursed for each service provided. Capitation system, the payment for all services is bundled depending on diagnosis/proced ure. Capitation system also includes payments for providing comprehensive care to a patient throughout a defined period of time irrespective of the amount and intensity of services rendered. Salary - based on work hours.				Compared with FFS, capitation payment decreased the number of provided services (14% lower visits in the outpatient settings and 50%-60% lower visits in the inpatient settings) due to budget limitations.	Salary payments were associated with low service provision and higher referral rate compared with FFS and capitation methods; FFS cause higher service production and induced more services to the patient. FFS leads to lower referral rate. Capitation decreased service production due to budget limitations.	Future studies should evaluate programs from a more diverse set of countries, in particular from developing countries. In these studies, researchers should attempt to control selection biases as rigorously as possible, using selection models in observational studies and randomized controlled trials where fund providers and policy makers are willing to support such experiments	None done;	Comparison of payment models; Behaviour change in primary care providers
VanHerck 2010: Systematic review: Effects, design choices, and context of pay-for- performance in health care.	summarizes evidence, obtained from studies published between January 1990 and July 2009, concerning P4P effects, as well as evidence on the impact of design choices and contextual mediators on these effects. Effect domains include clinical effectiveness, access and equity, coordination and continuity, patient- centeredness, and cost- effectiveness.	Systematic review (January 1990 and July 2009): 63 USA, 57 UK. 2 Australia, 2 Germany, 2 Spain. 1 Argentina 1 Italy	Pubmed/ Medline; Embase; PsycInfo; EconLit; Cochrane Library; Web of Science. Published in a peer- reviewed journal or published by the Agency for Healthcare Research and Quality (AHRQ), the Institute of Medicine (IOM), the National Health Service Department of Health or a non-profit independent academic institution		NA; Primary care/general practice; Other: Acute hospital care provider, team of providers or an individual physician; Medical	Healthcare providers in primary and/or acute hospital care; being a provider organization (hospital, practice, medical group, etc.); team of providers or an individual physician The use of an explicit financial positive or negative incentive directly related to providers' performance with regard to specifically measured quality-of-care targets and directed at a person's income or at further investment in quality improvement; performance measured as achievement and/or improvement At least one structural, process, or (intermediate) outcome measure on		P4P (The use of an explicit financial positive or negative incentive directly related to)	Providers' performance with regard to specifically measured quality-of-care targets and directed at a person's income or at further investment in quality improvement; performance measured as achievement and/or improvement Various measures include those focused on efficiency, productivity and cost reduction			P4P did not have negative effects on patients of certain age groups, ethnicity, or socio- economic status, or patients with different comorbid conditions (mainly UK data) Small difference implicating less P4P achievement for female as compared to male patients was found Note: No data from RCTs	Negative clinical effectiveness results were found only in a minority of cases: in three studies on one target, each of which also reported positive results on other targets ('negative' in this context means less quality improvement compared to non P4P use and not a quality decline) In general there was about 5% improvement due to P4P use, but with a lot of variation, depending on the measure and program. P4P most frequently failed to affect acute care. In chronic care, diabetes was the condition with the highest rates of quality improvement due to P4P implementation. Positive results were also reported for asthma and smoking cessation. This contrasts with finding no effect with regard to coronary heart disease (CHD) care. The effect of P4P on non-incentivized quality measures varied from none to positive. However, one study	Six recommendations are supported by evidence 1) Define P4P targets 2) Use process and (intermediary) outcome indicators as target measures. 3) Involve stakeholders and communicate 4) Implement a uniform P4P design 5) Focus on quality improvement and achievement, 6) Distribute incentives at the individual and/or team level.	First, there were restrictions in the search strategy used (e.g. number of databases consulted). Secondly, although quality appraisal was performed, and most studies controlled for potential confounders, selection bias cannot be ruled out with regard to observational findings. However, an analysis of randomized studies identified the same effect findings, when assessed on a RCT only inclusion basis. Thirdly, publication bias is likely to impact the evidence-base of P4P effectiveness and data quality bias may make the comparison of results across P4P programs problematic. Fourthly, behavioural health care and nursing home care were excluded from the study as potential settings for P4P application. Fifthly, the large degree of voluntary participation in P4P programs might lead to a self-selection of higher performing providers with less room for improvement (ceiling effect). This	Mixed of 9 appraisal tools - used dependent on study design; Score 3 - 2 5 - 1 6 - 1 7 - 2 8 - 11 9 - 77 10 - 19 11 - 9 12 - 3 Over-rule (Not enough data) - 43	Complex chronic diseases; Comparison of payment models; Behaviour change in primary care providers

Author, year and title	Aims and Objectives	Type of review (years searched): Countries	Information sources/data bases searched	Main funding for primary care	Health, service and provider focus	Selection criteria	Aim - health or service impact, behaviour	Type - P4P, capitation, salary, blended	Basis for funding - performance measures, population coverage	Funding - funding amount, frequency	Who receives funding - service (type), provider	Service use/access - access by specific populations incl type and frequency	Quality of care - preventative care, chronic disease quality care	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of assessment	Relevance of the paper to research questions
						clinical effectiveness of care, access and/or equity of care, coordination and/or continuity of care, patient- centeredness, and/or cost- effectiveness of care							reported a declining trend in improvement rate for non- incentivized measures of asthma and CHD after a performance plateau was reached. One study found positive effects on P4P targets concerning coronary heart disease, COPD, hypertension and stroke when applied to non- incentivized medical conditions (10.9% effect size), suggesting a spillover effect.		could induce an underestimation of P4P effectiveness in the general population of providers. Finally, P4P introduces one type of financial incentive, but does not act in isolation. Other interventions are often simultaneously introduced alongside a P4P program, which might lead to an overestimation of effects. Supportive evidence concerning interactions between funding models e.g. FFS + Capitation is currently largely absent, due to the frequent non reporting of dominant payment system characteristics and the lack of overuse focused studies as a comparison point.		

Author, year and title	Aims and Objectives	Type of review (years	Information sources/data bases	Main funding for	Health, service and provider focus	Selection criteria	Aim - health or service impact,	Type - P4P, capitation, salary,	Basis for funding - performance measures,	Funding - funding amount, frequency	Who receives funding -	Service use/access - access by specific	Quality of care - preventative care, chronic disease quality	Interpretation (authors)	Limitations (identified by authors)	Methods and summary of	Relevance of the paper to research
		searched): Countries	searched	primary care			behaviour	blended	population coverage		service (type), provider	populations incl type and frequency	care			assessment	questions
Wranik 2019: Implications of interprofessio nal primary care team characteristic s for health services and patient health outcomes: A systematic review with narrative synthesis.	synthesize the available empirical evidence of the implications of team characteristics on team processes, care outputs, and health / cost outcomes. The purpose is to support policy makers in designing IPPC teams that improve PC delivery and positively contribute to population health.	Systematic review (2003 to Feb 2016): limit to four countries: Australia, Canada, United Kingdom and New Zealand	Pubmed/ Medline; Embase; CINAHL; PAIS		NA; Primary care/general practice; Medical			P4P, Salary, FFS, Revenue Sharing	No detail into provided			Strong qualitative evidence that pay-for- performance arrangements are perceived to increase provision of billable services (at expense of other services).	Moderate quantitative evidence of remuneration methods having no impact on treatment targets.	The addition of new providers and new models of care is associated with increases in the provision of recommended tests and preventive services, and a decrease in the use of hospitals. Change in team characteristics and cost implications specifically on costs-as- outcomes, are unclear.	First, the analytical framework relies on the quantitative conceptualization of reality into causes and effects. The assumption is made that characteristics of teams contribute to the outputs and outcomes of care. The language (influences and consequences) is used to avoid terminology typically associated with quantitative studies (e.g. cause and effect, or dependent variable). The use of the framework may have resulted in the assignment of some qualitative studies to research questions that were not the primary questions intended by the study authors. Second, the MMAT rating tool proved relatively challenging to use, and we observed that it may favour qualitative studies, in that criteria for the assignment of the rating are more interpretive. Third, the need to formulate a search syntax required that our study question be narrowed in scope beyond what may have been justified substantively (e.g. by specific health conditions). Lastly, two amendments to the study protocol were made during the course of the review: (i) limitation of scope to the four countries and selected years; and (ii) decision about further conditions, beyond the initial four chronic conditions specified in the protocol.	Mixed Methods Appraisal Tool (MMAT, out of 5) ; 1 - 3 2 - 7 3 - 32 4 - 33 5 - 1 Not reported - 1	Outcomes in multidisciplin ary settings; Complex chronic diseases; Comparison of payment models; Behaviour change in primary care providers

Author, year and title	Methods Relevance of and the paper t	Limitations (identified by authors)
	summary of research assessment questions	
and title Yuan 2017: Payment methods for outpatient care facilities.	and summary of assessment EPOC suggested Risk of bias' criteria; 3 LOW Risk of Bias 12 Unclear Risk of Bias 5 High Risk of Bias	by authors) No evidence on dental clinics. information on how the incentive payments were used inside the health facilities was lacking in some of the included studies. Since P4P is intended to improve targeted behaviours through financial incentives, it is uncertain to what extent the way in which incentive payments were used influenced the effects of the P4P programs that were evaluated, and it would be difficult to replicate (or know how to improve) this component of the programs.

Table A4.3: List of included other literature.

Full citation

Bitton A, Schwartz GR, Stewart EE, Henderson DE, Keohane CA, Bates DW, Schiff GD. Off the hamster wheel? Qualitative evaluation of a payment-linked patient-centered medical home (PCMH) pilot. Milbank Q Sep 2012;90(3):484-515 United States 2012 Sep.

Chami, N and Sweetman, A. Payment models in primary health care: A driver of the quantity and quality of medical laboratory utilization. Health Economics Volume 28, Issue 10, October 2019, Pages 1166-1178.

Ding, Yu and Liu, Chenyuan. Alternative payment models and physician treatment decisions: Evidence from lower back pain. Journal of Health Economics Volume 80, December 2021, 102548.

Hollander MJ, Kadlec H. Incentive-based primary care: cost and utilization analysis. The Permanente Journal. 2015;19(4):46.

Hsieh H-M, Gu S-M, Shin S-J, Kao H-Y, Lin Y-C, Chiu H-C (2015) Cost-Effectiveness of a Diabetes Pay-For-Performance Program in Diabetes Patients with Multiple Chronic Conditions. PLoS ONE 10(7): e0133163.

Pandya, A., Doran, T., Zhu, J., Walker, S., Arntson, E. and Ryan, A.M., 2018. Modelling the cost-effectiveness of pay-for-performance in primary care in the UK. BMC medicine, 16, pp.1-13.

Qureshi, N., Weng, S. and Hex, N., 2016. The role of cost-effectiveness analysis in the development of indicators to support incentive-based behaviour in primary care in England. Journal of health services research & policy, 21(4), pp.263-271.

Rudoler, D., Deber, R., Barnsley, J., Glazier, R.H., Dass, A.R. and Laporte, A., 2015. Paying for primary care: the factors associated with physician self-selection into payment models. Health economics, 24(9), pp.1229-1242.

Skovsgaard, C.V., Kristensen, T., Pulleyblank, R. and Olsen, K.R., 2023. Increasing capitation in mixed remuneration schemes: Effects on service provision and process quality of care. Health Economics, 32(11), pp.2477-2498.

Somé, N.H., Devlin, R.A., Mehta, N., Zaric, G., Li, L., Shariff, S., Belhadji, B., Thind, A., Garg, A. and Sarma, S., 2019. Production of physician services under fee-for-service and blended fee-for-service: Evidence from Ontario, Canada. Health economics, 28(12), pp.1418-1434.

Somé, N.H., Devlin, R.A., Mehta, N., Zaric, G.S. and Sarma, S., 2020. Stirring the pot: Switching from blended feefor-service to blended capitation models of physician remuneration. Health economics, 29(11), pp.1435-1455.

Zhang, X and Sweetman, A. Blended capitation and incentives: Fee codes inside and outside the capitated basket. Journal of Health Economics 07/2018 2018;60():16-29. 2018 07/2018.

Appendix 5: Included reviews – Quality assessment

Study ID	1. Study eligibility criteria	2. Identification and selection of studies	3. Data collection and study appraisal	4. Synthesis and findings	Overall Risk of Bias
Ahmed 2021 ¹⁹	Low	Low	Low	Unclear	Unclear
Bes 2023 ²⁰	Low	High	Low	Low	High
Boeckxstaens 2011 ²¹	Low	Low	Unclear	Low	Low
Bowling 2018 ⁷	High	Unclear	Unclear	Unclear	Unclear
Brocklehurst 2013 ²²	Low	Low	Low	Unclear	Low
Carter 2016 ²⁶	Low	High	Low	Low	Unclear
Emmert 2012 ⁸	Low	Low	Low	Low	Low
Flodgren 2011 ³³	Low	Low	Low	Low	Low
Gillam 2012 ²⁵	Low	Low	Low	Low	Low
Hamilton 2013 ²⁷	Low	High	Low	High	High
Heider 2020 ³⁴	Low	Low	High	Low	High
Houle 2012 ²⁴	Low	Low	Low	Low	Low
Jackson 2017 ⁹	Low	Low	Low	Low	Low
Jia 2021 ⁵⁰	Low	Low	Low	Low	Low
Khan 2020 ³⁶	High	Low	High	High	High
Langdown 2014 ¹⁶	Low	Unclear	Unclear	High	Unclear
Lin 2016 ¹²	Low	High	Low	High	High
Mandavia 2017 ²⁸	High	Low	Low	High	Unclear
Markovitz 2017 ¹⁰	Low	High	High	High	High
Mauro 2019 ¹⁵	Low	Low	High	Unclear	Low
Mendelson 2017 ¹⁸	Low	Low	Low	Low	Low
Ogundeji 2016 ²⁹	Low	Unclear	Low	Low	Unclear
Rashidian 2015 ³¹	Low	Low	Low	Low	Low
Scott 2018 ³⁵	Low	High	High	High	High
Tao 2016 ¹⁷	Low	High	Unclear	Low	Unclear
Tildy 2023 ¹³	Low	Low	Low	Low	Low
Vahidi 2013 ¹¹	Low	Low	High	High	High
Van Herck 2010 ²³	Low	Low	Low	Low	Low
Wranik 2019 ¹⁴	Low	Low	Low	Low	Low
Yuan 2017 ³⁰	Low	Low	Low	Low	Low

Table A5.1: Risk of bias by phase 2 concerns and overall, for each review article.

LEGEND: Low=Low Risk of Bias; High=High Risk of Bias; Unclear=Unclear Risk of Bias