

BJGP OPEN

General practice characteristics associated with pay-for-performance in the UK: a systematic review

Srai, Rhatika Kaur; Cromwell, David; Mays, Nicholas; Pettigrew, Luisa M

DOI: <https://doi.org/10.3399/BJGPO.2024.0174>

To access the most recent version of this article, please click the DOI URL in the line above.

Received 24 July 2024

Revised 09 August 2024

Accepted 04 November 2024

© 2024 The Author(s). This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>). Published by BJGP Open. For editorial process and policies, see: <https://bjgpopen.org/authors/bjgp-open-editorial-process-and-policies>

When citing this article please include the DOI provided above.

Author Accepted Manuscript

This is an 'author accepted manuscript': a manuscript that has been accepted for publication in BJGP Open, but which has not yet undergone subediting, typesetting, or correction. Errors discovered and corrected during this process may materially alter the content of this manuscript, and the latest published version (the Version of Record) should be used in preference to any preceding versions

General practice characteristics associated with pay-for-performance in the UK: A Systematic Review

Authors:

Rhatica Srai, General Practitioner (ORCID ID: 0000-0001-5933-6930)

David Cromwell, Professor of Health Services Research (ORCID ID: 0000-0002-6516-8125)

Nicholas Mays, Professor of Health Policy (ORCID ID: 0000-0001-9808-8466)

Luisa M Pettigrew, NIHR Doctoral Research Fellow, General Practitioner (ORCID ID: 0000-0001-6427-6696) luisa.pettigrew@lshtm.ac.uk (corresponding author)

Department of Health Services Research & Policy, Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine

Accepted Manuscript - BJGP Open - BJGP2024.0174

General practice characteristics associated with pay-for-performance in the UK: A Systematic Review

Rhatica Srai, David Cromwell, Nicholas Mays, Luisa M Pettigrew

Department of Health Services Research and Policy, Faculty of Public Health and Policy, London School of Hygiene & Tropical Medicine

Abstract

Background

The Quality and Outcomes Framework (QOF), a pay-for-performance programme, has been the most widespread quality initiative in National Health Service (NHS) general practice since 2004. It has contributed between 25% and 8% of practices' income during this time, but concerns about its effect on equity have been raised.

Aim

Understand which practice characteristics are associated with QOF performance.

Design and setting

Systematic review, NHS general practice.

Method

MEDLINE, Embase, CINAHL+, Web-of-Science and grey literature were searched for studies examining the association between general practice characteristics and QOF performance.

Results

Twenty-two studies, published between 2006 and 2022, exploring the relationship between six population and 15 organisational characteristics and QOF measures were found. Most studies were cross-sectional, of English general practices, and used data from the early years of QOF. A negative association was frequently found between overall QOF performance and socioeconomic deprivation; proportion of registered patients >65; list size; mean general practitioner (GP) age; and Alternative Provider Medical Services contracts. Group practices (versus single-handed); more full-time-equivalent GPs; and being a training practice were frequently associated with better overall QOF performance. The associations of most other characteristics with performance were inconsistent.

Conclusion

Associations with characteristics both within and outside practices' control were identified. Pay-for-performance instruments may systematically disadvantage practices serving those at greatest risk of ill-health, such as older and more deprived populations. Given the cross-sectional design of many studies and focus on the early years of QOF, more up-to-date evidence is needed to understand if and why these relationships persist.

Key words

General Practice; Primary health care; Quality of Health Care; Incentive Reimbursement; Organisational Management

How This Fits In

The Quality and Outcome Framework (QOF) has been the most widespread service quality initiative in UK NHS general practice over the past 20 years. This systematic review examined studies of the association between QOF performance and general practices' population and practice characteristics. Associations were identified with characteristics both within and outside practices' control. Some of these may be exacerbating inequities in health and care. Up-to-date evidence is needed to understand whether and, if so, why these relationships persist as the scope of QOF is reviewed and new pay-for-performance schemes are introduced in general practice in the UK and elsewhere.

Accepted Manuscript - BJGP Open - BJGPO.2021.0174

Introduction

The Quality and Outcomes Framework (QOF) was introduced in 2004 to UK NHS general practice as a pay-for-performance incentive scheme. It was viewed as a mechanism to increase government funding into general practice while trying to ensure value for money. The QOF covers a range of clinical and organisational quality indicators which are revised every year (1-3). It has been associated with improved recording of, and reduced variation in incentivised care, but evidence is limited on its impact on health outcomes and health inequalities (4-17).

QOF scores are publicly reported and in England they inform the Care Quality Commission's (CQC) regulatory inspections and ratings. The QOF initially contributed up to 25% of practice income, however, it was removed in 2016 in Scotland and its contribution to practice income in England had declined to around 8% by 2022/23 (18-20). Despite this, it remains the most widespread quality incentive scheme used in UK general practice over the past 20 years. In 2022/23, NHS England reported spending £769 million on QOF payments (19). Pay-for-performance indicators similar to those in QOF now form part of NHS Primary Care Networks' (PCN) 'Investment and Impact Fund' (IIF) in England and other pay-for-performance schemes are widespread in the UK general practice under 'Local Enhanced Services' (LES) (21, 22).

Various population (e.g. location, patient demographics) and organisational characteristics (e.g. list size, training practice status) have been found to be associated with general practices' performance as measured by the QOF at different time points, across different QOF measures and geographical areas. However, this evidence has not been reviewed as a whole. This study therefore systematically reviews evidence, using national level data from the four countries in the UK, to understand which general practice characteristics have been studied and their associations with QOF.

Methods

The review is reported in accordance with PRISMA and Synthesis Without Meta-analysis (SWiM) in Systematic Review guidelines (23, 24). The protocol was registered with PROSPERO (CRD42021225146)(25).

Search strategy

MEDLINE, Embase, CINAHL +, Web of Science databases were searched using terms related to (i) QOF and (ii) statistical measures of association (Supplementary Box 1) up to January 2022. The reference lists of selected articles were searched for additional studies and Google was used to search for grey literature using key words from the search strategy.

Inclusion and exclusion criteria

Studies were included if they examined whether QOF performance was associated with any population or organisational general practice characteristics using national datasets from England, Scotland, Wales and/or Northern Ireland. The review excluded studies that used QOF or other performance measures as explanatory variables in their models, or used QOF exception reporting rates (when patients are excluded from the eligible QOF population for a justified reason (26)) as the outcome variable.

Study screening, selection, data extraction and quality assessment

RS and LP independently screened and selected the studies. Both extracted data and quality-assessed the studies with disagreement resolved by discussion between the authors. Quality was assessed using the modified Newcastle-Ottawa Scale for cross-sectional studies and Critical Appraisal Skills Programme (CASP) tool for cohort studies (27, 28). Data were extracted on the population, study design, year(s) of data used, exclusion criteria, explanatory and outcome variable(s), adjustment for confounding, direction of associations and their statistical significance. Associations were taken from the final statistical model(s) published, where available, including supplementary material.

Categorising characteristics and synthesis

Explanatory variables

Where possible, related explanatory variables were grouped. For example, various measures of deprivation were combined into one group. Variables were left in subcategories if they could not be combined in a logical way due to a lack of overlap, for example, some patient age groupings.

Outcome variables

The protocol was revised to group QOF outcome variables into three categories, rather than one, due to the range used in studies (25). The categories were: (i) 'Overall' where the total QOF or whole domain(s) (e.g. 'clinical' domain) scores were used; (ii) 'Subdomain' where one or more disease/condition specific QOF subdomain(s) scores were used (e.g. asthma, diabetes mellitus, mental health - if only the register indicator, which captures reported prevalence, was omitted, then this was still considered as a full subdomain); and (iii)

'Subgroup' where a group of indicators had been selected by the study authors (e.g. influenza immunisation, blood pressure, cholesterol). Associations with the percentage of QOF points achieved, with ('reported achievement') and without ('population achievement') exception reporting were counted separately. Studies were grouped by the time period of the QOF figures into early-, mid- and recent periods to identify time trends. If multiple years of data were reported separately, each year was counted as a separate analysis as QOF indicators, target thresholds and associated payments changed over time.

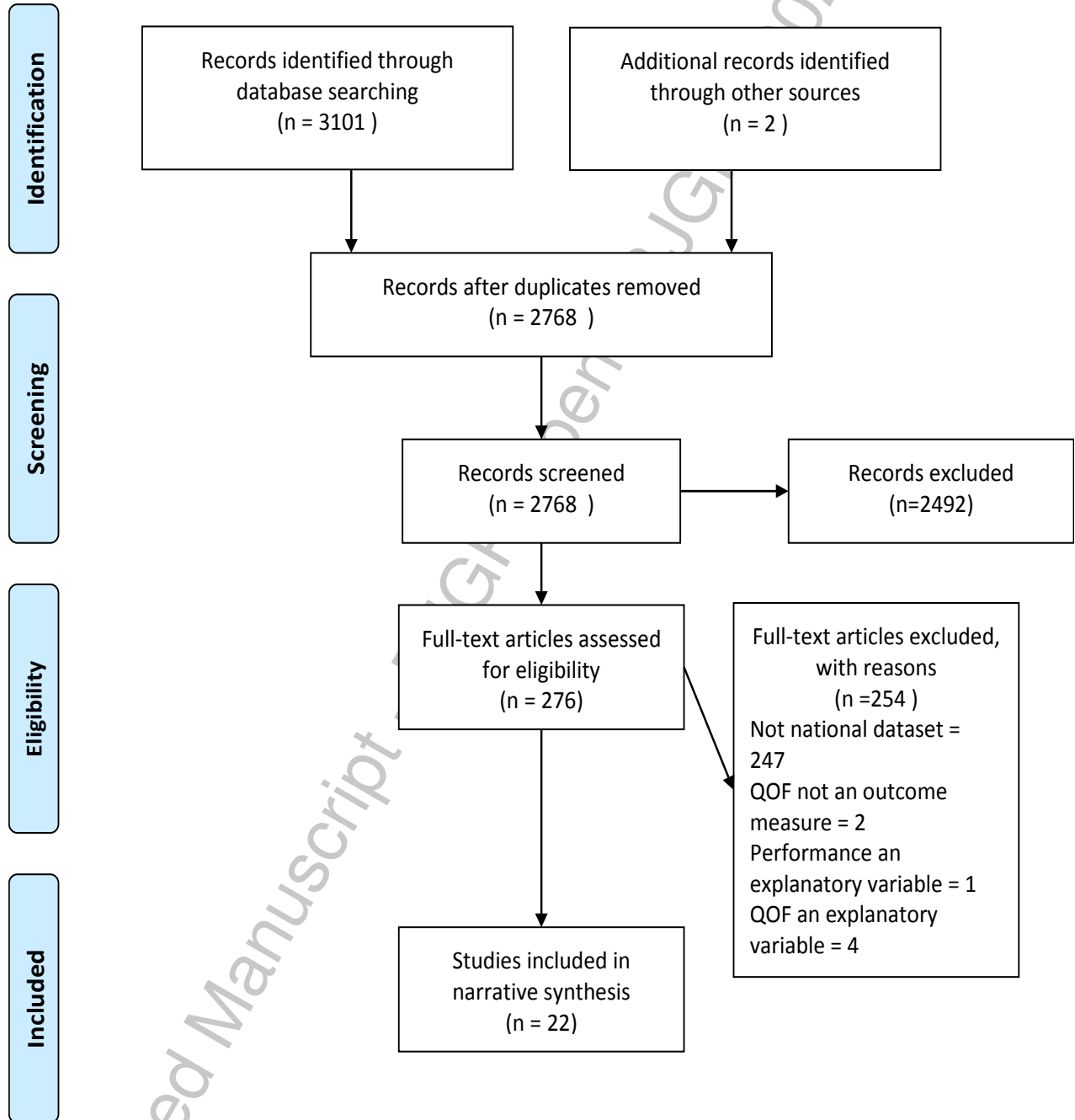
Synthesis

Heterogeneity between studies precluded synthesis beyond capturing the direction and strength of association. Associations were classed as 'positive', 'negative' or 'no association'. Associations which were not statistically significant ($P > 0.05$) were classified as 'no association'. Associations reported after adjusted for confounding factors scored one point; unadjusted associations, where they were the final results, were awarded half a point. Points per association were added within and across all studies to give the direction and consistency of association. If all associations were in the same direction, we considered this a 'consistent' association. If $\geq 60\%$ of associations were the same direction, this was considered to be a 'relatively consistent' association. If fewer than 60% of the associations were in the same direction, this was considered to be an 'inconsistent association'.

Results

Search strategy results are summarised in Figure 1. Twenty-two studies published between 2006 and 2022 were included: 15 cross-sectional and 7 cohort studies. Study characteristics are detailed in Supplementary Tables S1 and S2.

Figure 1: PRISMA Flowchart of study selection



Study locations and time periods

Fifteen studies were based in England (14, 29-43), three in England and Scotland (one of which examined the countries separately) (44-46), and three in Scotland (47-49). We found no studies from Northern Ireland or Wales. Most studies excluded practices (i) with less than 1000 registered patients due to their atypical nature, (ii) with missing data, or (iii) that were not consistently within merged datasets. Most studies reported including over 90% of practices in the UK country studied (Supplementary Tables S1 and S2). A description of the typical characteristics of excluded smaller practices can be found elsewhere (7).

Studies used QOF data from 2004/05 to 2016/17, but 14 studies examined the first two years of QOF implementation. Longitudinal studies varied from two to five years, two presented the years separately (14, 38) and the remainder reported average values over the study period (31, 34, 37, 41, 49) (Supplementary Tables S1 and S2).

Methodological quality of studies

Fourteen studies were rated as being of high methodological rigour (14, 29-31, 33, 35-42, 49), four were rated as good (34, 43-45) and four as satisfactory (32, 46-48) (Supplementary Tables S3 and S4).

Figure 2a: Associations between General Practice Population Characteristics and Quality and Outcomes Framework (QOF) performance

Population Characteristics	'Overall': Total QOF or Whole QOF Domains	'Subdomain': Condition/Disease Specific QOF Subdomains	'Subgroup': Selected QOF Indicators Subgroup
1. Increasing Deprivation	Neg-r (14, 29-31, 35, 37, 41, 43)	Inc (33, 39, 40, 44, 45)	Inc (30, 34, 40, 46, 47, 49)
2. Increasing Rurality	Inc (14, 30, 35, 37, 41, 43)	Inc (39, 40)	No-r (40, 48, 49)
3. Patient Age: % pts 0 to 44	No (14, 35, 37)	-	-
% female pts 45 to 64	No-r (37)	-	-
% male pts 45 to 64	No (37)	-	-
Increasing proportion of pts >65	Neg-r (14, 29, 35, 41)	Inc (33, 39, 40)	Inc (40, 49)
Increasing proportion of male pts >65	Inc (37)	-	-
Increasing proportion of female pts >65	No (37)	-	-
4. Patient Gender: Increasing proportion of female patients	No-r (14, 35, 41)	-	-
5. Patient Ethnicity: Increasing proportion of non-white patients / ethnic minorities / born in developing country	Inc (14, 29, 31, 35, 37)	No-r (39, 40)	Neg-r (40)
6. Disease Prevalence: Increased overall disease burden (including disabilities allowance/ nursing home pts)	Inc (29, 37)	-	Inc (46)
Asthma	Inc (37)	-	-
Cancer	Inc (37)	-	-
Coronary Heart Disease	Neg (37)	-	-
Chronic Obstructive Pulmonary Disease	No-r (37)	-	-
Diabetes Mellitus	No (37)	-	-
Epilepsy	No-r (37)	-	-
Hypertension	Pos-r (37)	-	-
Mental Health	Inc (37)	-	-
Stroke	No-r (37)	-	-
Individual indicators for various conditions	-	Inc (39, 40, 45, 46)	Inc (40)

Figure 2b: Associations between General Practice Organisational Characteristics and Quality and Outcomes Framework (QOF) performance

Organisational Characteristics	'Overall': Total QOF or Whole QOF Domains	Subdomain': Condition/Disease Specific QOF Subdomains	'Subgroup': Selected QOF Indicators Subgroup
1. Group partnership Vs single-handed practice	Pos-r (29-31, 49)	No-r (39, 40)	No-r (30, 34, 40)(6)(10)(21)
2. Number of FTE GPs	Pos (29, 31)	Pos-r (33)(8)	-
3. Proportion of GPs Salaried	Inc (32, 37, 41)	-	-
4. List size/Nurse	-	Inc (39, 40)	Inc (39, 40)
5. List size/GP	Inc (14, 29, 31, 35, 37, 49)	No-r (33, 39, 40)	No-r (34, 40)
6. List Size	Neg-r (14, 35, 37, 41, 49)	Inc (39, 40, 45)	No-r (40, 44)
7. List turnover	No (29)	-	-
8: Average GP Age: >45yrs	-	No-r (39, 40)	Neg-r (34, 40)
Increasing mean age	Neg-r (14, 31, 35, 37, 41, 49)	-	-
9. GP Gender: Proportion of Female GPs	Inc (14, 31, 35, 37, 41)	Inc (39, 40)	Pos-r (40)
10. Proportion of GPs qualified in UK/Europe Vs elsewhere	Inc (14, 31, 35, 37, 41)	Pos-r (33, 39, 40)	Pos-r (34, 40)
11. Contract: GMS (General Medical Services)	No-r (37)	-	-
PMS (Personal Medical Services)	Inc (14, 29, 35, 41)	No-r (33, 34, 39, 40)	No-r (40)
APMS (Alternative Provider Medical Services)	Neg (38, 41)	Neg (38)	Neg (38)
PCTMS (Primary Care Trust Medical Services)	Neg (41)	-	-
12. Capitation Payment Supplement	No (42)	-	-
13. Training Practice	Pos-r (29-31)	-	No (30)(6)
14. Dispensing Practice	-	-	No-r (36)
15. Computer System (EMIS LV): PCS	Neg (41)	-	Inc (41)
Practice manager	Inc (41)	-	Inc (41)
Premiere	Inc (41)	-	Inc (41)
SystemOne	Inc (41)	-	No-r (41)
Synergy	Pos (29, 31, 41)	-	Neg-r (41)
Vision 3	Pos-r (41)	-	Inc (41)

Neg	Consistent Negative	Pos	Consistent Positive	No	Consistent No Association	Inc	Inconsistent Association
Neg-r	Relatively Consistent Negative	Pos-r	Relatively Consistent Positive	No-r	Relatively Consistent No Association	-	Not studied

'Consistent': all associations the same	'Relatively consistent': ≥60% of associations the same	'Inconsistent': <60% of associations the same
---	--	---

(Numbers in brackets represent the study citations)

Type and frequency of explanatory characteristics studied

Twenty-one explanatory variables - six population and 15 organisational characteristics - were included in studies. Which, how often and for what purpose these were explored varied. For example, computer system was only studied once, while deprivation was used in 17 studies in various formats. Fifteen studies included multiple explanatory characteristics in their regression model(s) and adjusted for confounding. Two adjusted for confounding in a very limited way (i.e. only included two explanatory variables) (43, 45), five did not adjust at all (e.g. reported univariate analysis) (32, 44, 46-48) (Supplementary material Tables S3 and S4).

Type and frequency of QOF outcomes studied

Most studies examined associations with more than one QOF measure. Overall measures of QOF achievement were used in 12 studies, with the 'clinical domain' being the most studied (14, 29-32, 35, 37, 38, 41-43, 49); subdomains were used in nine studies (30, 33, 38-40, 44, 45, 47, 48); and, subgroups in seven (30, 34, 36, 44-46, 49).

Consistency of associations

Associations which were examined in two or more studies and that showed consistent or relatively consistent positive or negative associations for the 'overall' QOF performance category, as well as the most frequently studied explanatory variables are discussed below. All explanatory variables with their direction and consistency of association with QOF performance are presented in Figures 2a&b with the relevant citations. Full data extraction tables are available from the authors.

Population characteristics

Deprivation was the most studied characteristic, showing a relatively consistent negative association with the overall QOF performance (14, 29-31, 35, 37, 41, 43), but an inconsistent relationship with specific clinical subdomains (33, 39, 40, 44, 45) and indicator subgroups (30, 34, 40, 46, 47, 49). Rurality was the second most frequently studied explanatory variable. This showed an inconsistent association with overall QOF performance (14, 30, 35, 37, 41, 43) and subdomains (39, 40), and relatively consistently no association with indicator subgroups (40, 48, 49).

Having a higher proportion of patients over 65 was frequently studied and showed a relatively consistent negative association with overall QOF achievement (14, 29, 35, 41), but an inconsistent relationship with subdomains and indicator subgroups (33, 39, 40, 49). Patient ethnicity was also frequently studied and showed an inconsistent association with overall achievements (14, 29, 31, 35, 37) and relatively consistently no associations with subdomain achievements (39, 40). In one study examining subgroup indicators, there was a relatively consistent negative association between the percentage of patients from an ethnic minority and QOF performance, driven by indicators related to diabetes and epilepsy (40).

Organisational characteristics

Group practices (versus single-handed practices) were relatively consistently associated with better overall achievement (29-31, 49); however, this association did not hold when examining clinical subdomains and indicator subgroups (30, 34, 39, 40). Similarly, higher numbers of full time equivalent (FTE) GPs were consistently associated with better overall performance (29, 31). This association was also present in clinical subdomains, although to a lesser degree (33). In contrast, there was a relatively consistent negative association with list size (14, 35, 37, 41, 49) and an inconsistent relationship between list size per GP and overall performance (14, 29, 31, 35, 37, 49).

Increasing mean GP age was relatively consistently associated with poorer overall performance (14, 31, 35, 37, 41, 49), notably for over 45s in indicator subgroups related to diabetes and stroke (40). Seven studies examined the association with GP gender (14, 31, 35, 37, 39-41) and reported inconsistent associations with achievement, except in one study of subgroup indicators where female GPs showed a relatively consistent positive association with QOF achievement, driven by diabetes and epilepsy indicators (40). The proportion of GPs qualified in the UK or the rest of Europe (versus elsewhere) was frequently studied. While there was an inconsistent association with overall QOF measures (14, 31, 35, 37, 41), a higher proportion of UK or rest of Europe qualified GPs was relatively consistently associated with higher achievement in clinical subdomains and indicator subgroups driven by COPD, coronary heart disease, diabetes, epilepsy, hypertension, hypothyroid and stroke indicators (33, 34, 39, 40).

The relationship with different NHS contract types was also frequently analysed. Two studies found Alternative Provider Medical Services (APMS) service contracts, showed consistently negative associations across all QOF groupings (38, 41). Training practice status was relatively consistently associated with better overall performance measures in three studies (29-31).

Trends in associations

We did not identify any differences in patterns of association comparing QOF measures with or without exception reporting, nor comparing specific clinical subdomain or indicator subgroup explanatory variables. We also did not find any trends in associations over time or in the three Scottish studies compared to those in England, although ability to do so was limited due to the lack of studies in more recent years and from outside England.

Discussion

Summary

Twenty-two studies were found exploring the relationship between 21 general practice characteristics and QOF performance. Most studies used data from the early years of QOF in England. An association was frequently identified between poorer overall QOF performance and higher deprivation; proportion of patients over 65; list size; mean GP age; and having an APMS contract. A positive association with overall QOF performance was frequently seen with group practices (versus single-handed); total FTE GPs; and training practices. The proportion of GPs whose primary medical qualification was from the UK or Europe (vs elsewhere) showed a relatively consistent positive association with better performance across QOF sub-domains and sub-group indicators, but an inconsistent association with overall performance. Inconsistent associations were found with most other characteristics.

Strengths and limitations

We only included studies using national datasets as QOF was a national policy, however an opportunity exists to examine sub-national studies. Variables which had multiple definitions were grouped to make sense of the findings but in the process lost granularity. Most studies used cross-sectional data from the first few years of QOF, however, practice characteristics, contextual factors and QOF itself have changed over time (50). The use of vote-counting across heterogeneous studies has limitations, including not being able to comment on the magnitude of associations and the risk of subjective interpretation (51). Counting only statistically significant associations, giving less weighting to studies which did not adjust for confounding, and setting a 60% rather than a 50% cut-off to define the consistency of associations reduced the risk of overestimating the presence of these, although it may have resulted in an underestimation (24, 51). Importantly, association does not mean causality, however it does signal areas that merit further attention, in particular, where associations appear more consistently and are plausible causally (52).

Comparison with existing literature

Studies have shown that inequalities in performance related to levels of deprivation diminished during the early years of QOF performance (14, 15, 53). However the association persisted in studies using later data and is seen in sub-national level studies (53, 54). The association between poorer performance and proportion of patients over 65 contrasts with findings of a recent longitudinal study suggesting practices with a higher proportion of over 65s perform better on QOF, this difference may be due to methodological differences adjusting for confounding variables (53). Associations between poorer performance, indices of deprivation and older populations suggest that the socioeconomic determinants of health and the 'inverse care law' may be at play in determining practices' ability to perform on QOF measures (55-59).

The mixed picture of associations with the number of FTE GPs, list size and patients/FTE GP may reflect differences in methods and data used. However, it mirrors mixed findings in the wider literature regarding the relationship between list size and clinical quality of care,

suggesting variables other than organisational size drive performance and that the relationship may not be linear (37, 60-62).

The association between older GPs and poorer performance on QOF may seem counterintuitive and could indicate intentional disengagement from QOF, for example, to prioritise other dimensions of quality, rather than an inability to deliver. However, a 2005 systematic review also identified a negative relationship between clinical experience and measured quality of healthcare (63). New evidence in this area would be helpful as the volume of information and mechanisms for clinicians to keep up-to-date have changed significantly.

Around 25% of GPs working in England qualified outside the UK (50, 64). Their contributions in, typically, more challenging and socioeconomically deprived areas has often been under recognised, and while hypothesised, evidence that international medical graduates deliver poorer quality care has been limited to date (65-67). However an association remained after controlling for socioeconomic deprivation between better performance in some aspects of QOF and practices with a higher proportion of UK-qualified GPs.

APMS service contracts are time-limited, they account for a small (4% in England) and diminishing proportion of general practice contracts and are often used to enable the contracting of incorporated limited companies (19, 68, 69). Greaves et al. identified that practices with APMS contracts are usually smaller, serving younger, more diverse and deprived populations, but despite adjustment still found an association with poor performance (38). Commercial interests and the use of APMS contracts when failing practices have been put out to tender have been hypothesised as reasons for poorer performance (38). In contrast, the proportion of training practices is increasing (50), and have been associated with better clinical quality as well as patient satisfaction due to their greater focus on education and clinical governance (64, 70-73).

Implications for research and practice

This review identifies commonly used measurable population and organisational general practice characteristics. It identifies that up-to-date research into characteristics associated with QOF is needed. Synthesising existing evidence at sub-national level and comparing associations with other quality measures, notably the General Practice Patient Satisfaction Survey (GPPS), CQC ratings, and other pay-for-performance schemes, such as LESs and the IIF, would inform policy regarding general practice funding and its organisational structure. Given the evolving structure of general practice other explanatory characteristics could be explored including the ratio of non-GP:GP FTE roles per 1000 patients; or the use of different digital solutions to drive quality, noting that Kontopantelis et al. found that the choice of clinical computing system was the strongest predictor of QOF performance in their model (41, 50).

Pay-for-performance schemes, while they have the potential to help address inequities in health and care, if not carefully designed they may exacerbate these (16, 74, 75). Practices in socioeconomically deprived neighbourhoods need adequately adjusted capitated funding

and support to address the social determinants of health, as well as to care for older populations who have greater multimorbidity (76-79).

Caution is needed with current policy driving the formation of larger general practice organisations and task shifting to allied healthcare professionals, as its relationship with quality and cost-effectiveness unclear (37, 50, 60, 61, 80, 81). The associations between increasing mean GP age and being qualified outside the UK/Europe requires further investigation, but may suggest the need for careful workforce planning and additional support for certain GPs cohorts to engage with continuous professional development.

Inconsistencies in associations with other practice characteristics and QOF may reflect methodological differences, such as, the QOF performance measure used, year of study and degree of adjustment for confounding. They may also be due to non-linear relationships. Importantly, inconsistencies in associations highlight the complexity of quality as a concept, and the limitations of quantifiable characteristics being able to explain variation (9, 82, 83).

Conclusion

Relatively consistent associations with QOF performance and both characteristics which are within practices' control and those which are not were found. Up-to-date evidence is needed to understand if and why these relationships persist as they may be exacerbating inequities in health and care which need to be addressed.

Additional information

Funding: LP is funded by a NIHR Doctoral Research Fellowship (DRF-2017-10-088). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Ethical approval: not required

Competing interests: nil

References

1. NHS England. Quality and Outcome Framework 2022. Available from: <https://qof.digital.nhs.uk/>.
2. Reeves D, Doran T, Valderas JM, Kontopantelis E, Trueman P, Sutton M, et al. How to identify when a performance indicator has run its course. *BMJ*. 2010;340.
3. Doran T, Kontopantelis E, Reeves D, Sutton M, Ryan AM. Setting performance targets in pay for performance programmes: what can we learn from QOF? *BMJ*. 2014;348:g1595.
4. Forbes LJ, Marchand C, Doran T, Peckham S. The role of the Quality and Outcomes Framework in the care of long-term conditions: a systematic review. *Br J Gen Pract*. 2017;67(664):e775-e84.
5. 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*. 2020;21(1):142.
6. Roland M, Guthrie B. Quality and Outcomes Framework: what have we learnt? *BMJ*. 2016;354:i4060.
7. Doran T, Campbell S, Fullwood C, Kontopantelis E, Roland M. Performance of small general practices under the UK's Quality and Outcomes Framework. *Br J Gen Pract*. 2010;60(578):e335-44.
8. 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 (Oxf)*. 2014;36(2):251-8.
9. Heath I, Rubinstein A, Stange KC, van Driel ML. Quality in primary health care: a multidimensional approach to complexity. *BMJ*. 2009;338.
10. Doran T, Kontopantelis E, Valderas JM, Campbell S, Roland M, Salisbury C, Reeves D. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. *BMJ*. 2011;342:d3590.
11. Zaresani A, Scott A. Is the evidence on the effectiveness of pay for performance schemes in healthcare changing? Evidence from a meta-regression analysis. *BMC Health Serv Res*. 2021;21(1):175.
12. Harrison MJ, Dusheiko M, Sutton M, Gravelle H, Doran T, Roland M. Effect of a national primary care pay for performance scheme on emergency hospital admissions for ambulatory care sensitive conditions: controlled longitudinal study. *BMJ*. 2014;349:g6423.
13. Ryan AM, Krinsky S, Kontopantelis E, Doran T. Long-term evidence for the effect of pay-for-performance in primary care on mortality in the UK: a population study. *Lancet*. 2016;388(10041):268-74.
14. Doran T, Fullwood C, Kontopantelis E, Reeves D. Effect of financial incentives on inequalities in the delivery of primary clinical care in England: analysis of clinical activity indicators for the quality and outcomes framework. *Lancet*. 2008;372(9640):728-36.
15. 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*. 2012;10(5):461-8.
16. Tao W, Agerholm J, Burstrom B. The impact of reimbursement systems on equity in access and quality of primary care: A systematic literature review. *BMC Health Serv Res*. 2016;16(1):542.
17. Alshamsan R, Majeed A, Ashworth M, Car J, Millett C. Impact of pay for performance on inequalities in health care: systematic review. *Journal of Health Services Research & Policy*. 2010;15(3):178-84.
18. Guthrie B, McLean G, Sutton M. Workload and reward in the Quality and Outcomes Framework of the 2004 general practice contract. *Br J Gen Pract*. 2006;56(532):836-41.
19. NHS England. NHS payments to general practice 2023. Available from: <https://digital.nhs.uk/data-and-information/publications/statistical/nhs-payments-to-general-practice>.
20. Moberly T, Stahl-Timmins W. QOF now accounts for less than 10% of GP practice income. *BMJ*. 2019(365):11489).

21. NHS England. Network Contract Directed Enhanced Service. Investment and Impact Fund 2022/23: Updated Guidance 2022. Available from: <https://www.england.nhs.uk/wp-content/uploads/2022/03/B1963-iii-Network-contract-IIF-Implementation-Guidance-September-2022.pdf>.
22. NHS England. Local enhanced service commissioning through GP contracts. 2023.
23. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
24. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ*. 2020;368.
25. Srari R, Pettigew L. A systematic review of associations between characteristics of UK general practices and quality of care as measured by QOF. PROSPERO 2021 CRD42021225146 2021. Available from: https://www.crd.york.ac.uk/prospéro/display_record.php?ID=CRD42021225146.
26. Doran T, Fullwood C, Reeves D, Gravelle H, Roland M. Exclusion of patients from pay-for-performance targets by English physicians. *N Engl J Med*. 2008;359(3):274-84.
27. Critical Appraisals Skills Programme. CASP Checklist: 12 questions to help you make sense of a Cohort Study Oxford2022. Available from: <https://casp-uk.net/checklists/casp-cohort-studies-checklist.pdf>.
28. Wells G, Shea, B., O'Connell, D., Peterson, J., Welch, Losos, M., Tugwell, P., Ga, S.W., Zello, G.A., & Petersen, J.A. . The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-Analyses. .2014. Available from: https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.
29. Ashworth M, Armstrong D. The relationship between general practice characteristics and quality of care: a national survey of quality indicators used in the UK Quality and Outcomes Framework, 2004-5. *BMC Fam Pract*. 2006;7:68.
30. Ashworth M, Seed P, Armstrong D, Durbaba S, Jones R. The relationship between social deprivation and the quality of primary care: a national survey using indicators from the UK Quality and Outcomes Framework. *Br J Gen Pract*. 2007;57(539):441-8.
31. Ashworth M, Schofield P, Seed P, Durbaba S, Kordowicz M, Jones R. Identifying poorly performing general practices in England: a longitudinal study using data from the Quality and Outcomes Framework. *J Health Serv Res Policy*. 2011;16(1):21-7.
32. Ding A, Hann M, Sibbald B. Profile of English salaried GPs: labour mobility and practice performance. *Br J Gen Pract*. 2008;58(546):20-5.
33. Dixon A, Khachatryan A, Tian Y. Socioeconomic differences in case finding among general practices in England: analysis of secondary data. *J Health Serv Res Policy*. 2012;17:18-22.
34. Dixon A, Khachatryan A, Gilmour S. Does general practice reduce health inequalities? Analysis of quality and outcomes framework data. *Eur J Public Health*. 2012;22(1):9-13.
35. Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, Roland M. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med*. 2006;355(4):375-84.
36. Gomez-Cano M, Wiering B, Abel G, Campbell JL, Clark CE. Medication adherence and clinical outcomes in dispensing and non-dispensing practices: a cross-sectional analysis. *Br J Gen Pract*. 2021;71(702):E55-E61.
37. Gravelle H, Liu D, Santos R. How do clinical quality and patient satisfaction vary with provider size in primary care? Evidence from English general practice panel data. *Soc Sci Med*. 2022;301.
38. Greaves F, Lavery AA, Pape U, Ratneswaren A, Majeed A, Millett C. Performance of new alternative providers of primary care services in England: an observational study. *J R Soc Med*. 2015;108(5):171-83.
39. Griffiths P, Maben J, Murrells T. Organisational quality, nurse staffing and the quality of chronic disease management in primary care: Observational study using routinely collected data. *Int J Nurs Stud*. 2011;48(10):1199-210.
40. Griffiths P, Murrells T, Maben J, Jones S, Ashworth M. Nurse staffing and quality of care in UK general practice: cross-sectional study using routinely collected data. *Br J Gen Pract*. 2010;60(570).

41. Kontopantelis E, Buchan I, Reeves D, Checkland K, Doran T. Relationship between quality of care and choice of clinical computing system: retrospective analysis of family practice performance under the UK's quality and outcomes framework (vol 3, e003190, 2013). *BMJ Open*. 2013;3(8).
42. L'Esperance V, Sutton M, Schofield P, Round T, Malik U, White P, Ashworth M. Impact of primary care funding on secondary care utilisation and patient outcomes: a retrospective cross-sectional study of English general practice. *Br J Gen Pract*. 2017;67(664):E792-E9.
43. Wright J, Martin D, Cockings S, Polack C. Overall Quality of Outcomes Framework scores lower in practices in deprived areas. *Br J Gen Pract*. 2006;56(525):277-9.
44. Millett C, Car J, Eldred D, Khunti K, Mainous AG, Majeed A. Diabetes prevalence, process of care and outcomes in relation to practice size, caseload and deprivation: national cross-sectional study in primary care. *J R Soc Med*. 2007;100(6):275-83.
45. Saxena S, Car J, Eldred D, Soljak M, Majeed A. Practice size, caseload, deprivation and quality of care of patients with coronary heart disease, hypertension and stroke in primary care: national cross-sectional study. *BMC Health Serv Res*. 2007;7.
46. McLean G, Guthrie B, Watt G, Gabbay M, O'Donnell CA. Practice postcode versus patient population: a comparison of data sources in England and Scotland. *Int J Health Geogr*. 2008;7.
47. McLean G, Sutton M, Guthrie B. Deprivation and quality of primary care services: evidence for persistence of the inverse care law from the UK Quality and Outcomes Framework. *J Epidemiol Community Health*. 2006;60(11):917-22.
48. McLean G, Guthrie B, Sutton M. Differences in the quality of primary medical care services by remoteness from urban settlements. *Qual Saf Health Care*. 2007;16(6):446-9.
49. Lowrie R, McConnachie A, Williamson AE, Kontopantelis E, Forrest M, Lannigan N, et al. Incentivised chronic disease management and the inverse equity hypothesis: findings from a longitudinal analysis of Scottish primary care practice-level data. *BMC Med*. 2017;15.
50. Pettigrew L, Petersen I, Mays N, Cromwell D. The changing shape of English general practice: a retrospective longitudinal study using national datasets describing trends in organisational structure, workforce and recorded appointments. *BMJ Open*. 2024 14:e081535.
51. McKenzie JE, Brennan SE. Chapter 12: Synthesizing and presenting findings using other methods *Cochrane Training*2023.
52. Shimonovich M, Pearce A, Thomson H, Keyes K, Katikireddi SV. Assessing causality in epidemiology: revisiting Bradford Hill to incorporate developments in causal thinking. *Eur J Epidemiol*. 2021;36(9):873-87.
53. Mann O, Bracegirdle T, Shantikumar S. The relationship between Quality and Outcomes Framework scores and socioeconomic deprivation: a longitudinal study. *BJGP Open*. 2023;7(4).
54. Glidewell L, West R, Hackett JEC, Carder P, Doran T, Foy R. Does a local financial incentive scheme reduce inequalities in the delivery of clinical care in a socially deprived community? A longitudinal data analysis. *BMC Fam Pract*. 2015;16.
55. Hart JT. The inverse care law. *Lancet*. 1971;1(7696):405-12.
56. The Lancet. 50 years of the inverse care law. *Lancet*. 2021;397(10276):767.
57. Cookson R, Doran T, Asaria M, Gupta I, Mujica FP. The inverse care law re-examined: a global perspective. *Lancet*. 2021;397(10276):828-38.
58. Whitty CJM, MacEwen C, Goddard A, Alderson D, Marshall M, Calderwood C, et al. Rising to the challenge of multimorbidity. *BMJ*. 2020;368:l6964.
59. Pearson-Stuttard J, Ezzati M, Gregg EW. Multimorbidity-a defining challenge for health systems. *Lancet Public Health*. 2019;4(12):e599-e600.
60. Pettigrew LM, Kumpunen S, Mays N, Rosen R, Posaner R. The impact of new forms of large-scale general practice provider collaborations on England's NHS: a systematic review. *Br J Gen Pract*. 2018;68(668):e168-e77.
61. Pettigrew LM, Kumpunen S, Rosen R, Posaner R, Mays N. Lessons for 'large-scale' general practice provider organisations in England from other inter-organisational healthcare collaborations. *Health Policy*. 2019;123(1):51-61.

62. Ng CW, Ng KP. Does practice size matter? Review of effects on quality of care in primary care. *Br J Gen Pract.* 2013;63(614):e604-10.
63. Choudhry NK, Fletcher RH, Soumerai SB. Systematic review: The relationship between clinical experience and quality of health care. *Ann Intern Med.* 2005;142(4):260-73.
64. NHS England. General Practice Workforce 2024. Available from: <https://digital.nhs.uk/data-and-information/publications/statistical/general-and-personal-medical-services>.
65. Esmail A, Simpson J. International medical graduates and quality of care. *BMJ.* 2017;356:j574.
66. Pettigrew LM. The NHS and international medical graduates. *Educ Prim Care.* 2014;25(2):71-5.
67. Tsugawa Y, Jena AB, Orav EJ, Jha AK. Quality of care delivered by general internists in US hospitals who graduated from foreign versus US medical schools: observational study. *BMJ.* 2017;356:j273.
68. Heins E, Pollock AM, Price D. The commercialisation of GP services: a survey of APMS contracts and new GP ownership. *Br J Gen Pract.* 2009;59(567):e339-43.
69. Beech J, Baird B. GP funding and contracts explained. The King's Fund, 2020.
70. Ahluwalia S, Spicer J, Patel A, Cunningham B, Gill D. Understanding the relationship between GP training and improved patient care - a qualitative study of GP educators. *Educ Prim Care.* 2020;31(3):145-52.
71. Weston C, Ahluwalia S, Bassett P, Lock J, Durbaba S, Ashworth M. GP Training practices in England: a description of their unique features based on national data. *Educ Prim Care.* 2017;28(6):313-8.
72. Ahluwalia S, Sadak M, Ashworth M. Antimicrobial prescribing in post-graduate training practices: a cross-sectional study of prescribing data in general practices in England. *Educ Prim Care.* 2018;29(3):139-43.
73. Ashworth M, Schofield P, Durbaba S, Ahluwalia S. Patient experience and the role of postgraduate GP training: a cross-sectional analysis of national Patient Survey data in England. *Br J Gen Pract.* 2014;64(620):e168-77.
74. Lin Y, Yin S, Huang J, Du L. Impact of pay for performance on behavior of primary care physicians and patient outcomes. *Journal of Evidence-Based Medicine.* 2015.
75. Casalino LP, Elster A. Will pay-for-performance and quality reporting affect health care disparities? *Health Affairs.* 2007;26(3):W405-W14.
76. Fisher R. 'Levelling up' general practice in England: What should government prioritise? . Health Foundation: 2021.
77. Hiam L, Klaber B, Sowemimo A, Marmot M. NHS and the whole of society must act on social determinants of health for a healthier future. *BMJ.* 2024;385:e079389.
78. Kontopantelis E, Mamas MA, van Marwijk H, Ryan AM, Bower P, Guthrie B, Doran T. Chronic morbidity, deprivation and primary medical care spending in England in 2015-16: a cross-sectional spatial analysis. *BMC Med.* 2018;16(1):19.
79. Rhys G, Beerstecher HJ, Morgan CL. Primary care capitation payments in the UK. An observational study. *BMC Health Serv Res.* 2010;10:156.
80. Forbes LJJ, Forbes H, Sutton M, Checkland K, Peckham S. How widespread is working at scale in English general practice? An observational study. *Br J Gen Pract.* 2019;69(687):E682-E8.
81. McDermott I, Spooner S, Goff M, Gibson J, Dalgarno E, Francetic I, et al. Scale, scope and impact of skill mix change in primary care in England: a mixed-methods study. *Health and Social Care Delivery Research.* Southampton (UK)2022.
82. Hanefeld J, Powell-Jackson T, Balabanova D. Understanding and measuring quality of care: dealing with complexity. *Bull World Health Organ.* 2017;95(5):368-74.
83. Braithwaite J, Herkes J, Ludlow K, Testa L, Lamprell G. Association between organisational and workplace cultures, and patient outcomes: systematic review. *BMJ Open.* 2017;7(11).