

BMJ Open How can patient experience scores be used to predict quality inspection ratings? A retrospective cross-sectional study of national primary care datasets in the UK

Amy Tallett ¹, Alan J Poots ¹, Chris Graham ¹, Michele Peters ², Rory Corbett¹, Steve Sizmur ¹, Julien Forder ³

To cite: Tallett A, Poots AJ, Graham C, *et al*. How can patient experience scores be used to predict quality inspection ratings? A retrospective cross-sectional study of national primary care datasets in the UK. *BMJ Open* 2020;**10**:e041709. doi:10.1136/bmjopen-2020-041709

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<https://www.bmjopen.com>).

Received 16 June 2020

Revised 23 September 2020

Accepted 06 November 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Pickering Institute Europe, Oxford, UK

²Nuffield Department of Population Health, University of Oxford, Oxford, UK

³PSSRU, University of Kent, Canterbury, UK

Correspondence to

Dr Amy Tallett;
amy.tallett@pickeringeurope.ac.uk

ABSTRACT

Objectives The relationship between patient feedback in the General Practice Patient Survey (GPPS) and Care Quality Commission (CQC) inspections of practices was investigated to understand whether there is an association between patient views and regulator ratings of quality. The specific aims were to understand whether patients' self-reported experiences of primary care can predict CQC inspection ratings of GP practices by: (i) Measuring the association between GPPS results and CQC inspection ratings of GP practices; (ii) Building a predictive model of GP practice quality ratings that use GPPS results; and (iii) Evaluating the predictive model for risk stratification.

Design Retrospective analysis of routinely collected data using decision tree modelling.

Setting Primary care: GP practices in England.

Primary and secondary outcome measures GPPS scores and GP practice CQC inspection ratings during 2018.

Results Most GP practices (72%, 974/1350) were rated as 'Good' overall by CQC. Simply assuming that all practices will be rated as 'Good' results in a correct prediction 72% of the time, and it was not possible to improve on this overall level of predictive accuracy using decision tree modelling (correct in 73% of cases). However, a set of GPPS questions were found to have value in identifying practices at elevated risk of a poor inspection rating.

Conclusions Although there were some associations between GPPS data and CQC inspection ratings, there were limitations to the use of GPPS data for predictive analysis. This is a likely result of the majority of CQC inspections of GPs resulting in a 'Good' or 'Outstanding' rating. However, some GPPS questions were found to have value in identifying practices at higher risk of an 'Inadequate' or 'Requires Improvement' rating, and this may be valuable for surveillance purposes. For example, the CQC could use key questions from the survey to target inspection planning.

INTRODUCTION

Globally, defects in quality create wide-spread and varied shortfalls of healthcare

Strengths and limitations of this study

- Decision tree modelling was useful for risk stratification, whereby a subset of General Practice Patient Survey (GPPS) questions could identify general practice at risk of a poorer inspection rating.
- Mapping questions in the GPPS to inspection domains, informed by an exercise undertaken by an expert group and by stakeholder consultation, showed that GPPS items map only to two of the Care Quality Commission domains.
- As GPPS asks about experience over the last 12 months, it was not possible to match patient experience to a specific inspection date.
- Due to the varying frequency of inspections, the dataset was restricted to those practices inspected during 2018, and data from earlier years could not be used due to substantial changes in the GPPS questionnaire between 2017 and 2018.
- Aggregate data may mask within-practice variation that could be reflected in inspection ratings.

for populations.¹ Assessing care quality is essential to improving care delivery.² Quality of healthcare is multifaceted and has been defined in different ways. The Organisation for Economic Cooperation and Development (OECD) defines quality of care in terms of safety, effectiveness and patient-centredness,³ which overlaps with the definition of the National Health Service (NHS) in England. The NHS Next Stage Review⁴ defined quality in terms of patient safety; clinical effectiveness; and patient experience, which serve as three overarching concepts for the five outcome domains of the NHS Outcomes Framework⁵ (table 1). Evidence suggests that although quality of care is particularly lacking in low-income and middle-income countries; high-income countries such as the UK and the USA are not exempt from poor quality

Table 1 The five outcome domains of the NHS Outcomes Framework⁵

Concepts	Domain
Effectiveness	(1) Preventing people from dying prematurely. (2) Enhancing quality of life for people with long-term illness. (3) Helping people recover from episodes of ill health or following injury.
Patient experience	(4) Ensuring people have a positive experience of care.
Safety	(5) Treating and caring for people in a safe environment and protecting them from avoidable harm.

NHS, National Health Service.

care.¹ While the National Academies of Sciences and the OECD aim for international comparison,⁶ the Next Stage Review⁴ aims to evaluate care quality specifically in the NHS.

In England, most patient contact with the NHS happens in primary care.⁷ Multiple approaches evaluate primary care services in England, including the Quality and Outcomes Framework; patient surveys such as the General Practice Patient Survey (GPPS); the Health and Social Care Information Centre's Indicator Portal and inspections by the Care Quality Commission (CQC).²

The CQC is the independent regulator for Health and Social Care providers in England, and has conducted inspections for health and social care services since 2009, and GP since 2013.⁸ The CQC asks all services five key questions (box 1) covering 37 indicators, which are flagged as showing 'no evidence of risk', 'risk' or 'elevated risk'.^{9 10} Additionally, CQC inspections bring together information held about GP practices and compares this to local and national data.¹¹

Studies have explored relationships between CQC ratings and a range of associated factors. Higher CQC ratings are associated with better quality, for example, higher quality ratings are associated with better quality of life in care home residents.¹² Another study indicated that acute NHS Trusts with better CQC ratings tend to

have high employee engagement, and lower ratings occur with high financial deficits.¹³

Quality of healthcare, including primary care, is multifaceted and can be complex.¹⁴ Consequently, defined indicators that evaluate quality only measure a part of what constitutes quality. Thus, a holistic approach to measuring quality is needed for that measurement to be useful. Holistic can refer to the perspectives by which quality is evaluated (eg, professionals and patients) or the methods used for measurement (eg, using indicators in conjunction with surveys; qualitative approaches and mixed methods).^{14–17} There is a wide array of initiatives to measure quality in primary care, yet these initiatives are poorly coordinated, overlap in some areas and leave other areas unmeasured.⁷ Therefore, indicators of quality of primary care should be consolidated, with a small number of 'vital signals' on what matters most, rather than all indicators being presented separately.²

If we use a holistic approach to quality evaluation, the patient perspective must be included. Experiences of healthcare services can be captured from patients through self-complete surveys.¹⁸ Such surveys complement other measures of care quality by accessing information that only patients hold.¹⁹ More positive patient experience is associated with better adherence to preventative and curative processes, better safety, better clinical outcomes, lower use of healthcare services¹⁹ and fewer complication rates.²⁰ Furthermore, patient feedback can identify hospitals likely to have poorer quality in CQC inspections.²¹

In England, the GPPS invites respondents to report their experiences of primary healthcare.¹⁵ The GPPS provides practice-level data that is comparable across organisations and time.²² The survey has been updated to account for changes in primary care provision, and the findings inform the distribution of NHS resources.¹⁵ GPPS survey results have provided important insights into primary care provision. For example, people with long-standing psychological problems or emotional conditions have similar experiences of primary care to the rest of the population.²³ Most people with multimorbidity report a positive experience of care and this experience deteriorates with a higher number of conditions.²⁴ Furthermore, practices where GPPS participants express lower levels of satisfaction, in particular for doctor patient communication, are more likely to experience higher levels of patients leaving the practice.²⁵

Box 1 Key questions Care Quality Commission asks all services, including primary care¹⁰

1. Are they safe?
Safe: you are protected from abuse and avoidable harm
2. Are they effective?
Effective: your care, treatment and support achieves good outcomes, helps you to maintain quality of life and is based on the best available evidence
3. Are they caring?
Caring: Staff involve and treat you with compassion, kindness, dignity and respect
4. Are they responsive to people's needs?
Responsive: services are organised so that they meet people's needs
5. Are they well-led?
Well-led: the leadership, management and governance of the organisation make sure it is providing high-quality care that is based around your individual needs, that it encourages learning and innovation, and that it promotes and open and fair culture

CQC inspection ratings come at a huge expense and it is not possible for all practices to be inspected annually with the available resource. Thus, alternative approaches that use existing data such as the GPPS, may be a more cost-effective way of understanding care quality. In particular, a predictive method that uses GPPS data could provide a sufficient assessment of practice ratings to complement CQC inspection activity. A 2020 study²⁶ suggests that GPPS data is not a good predictor of inspection ratings. However, in the current study, we explore a different modelling approach to examine risk stratification opportunities, to determine how far GPPS data can contribute to the assessment of practice ratings.

The GPPS was sent out by Ipsos MORI on behalf of NHS England to approximately 2.32 million adults registered at an English GP practice in 2019. The sample design involved a proportionately stratified, unclustered sample at each practice. The required number of patients per practice was selected on a '1 in n' basis. For statistical representativeness, weights are generated to correct for potential design effects and non-response bias. The 2019 survey achieved over 770 000 responses.²⁷ Data of this volume may make it possible to use patient experience to identify GP practices at greater risk of poorer care quality and to therefore identify priorities for inspection.

Aim

In this study, we examine the relationship between two indicators of quality in primary care in England: GPPS scores and CQC inspection ratings. Specifically, the research explores whether patients' self-reported

experiences of care can predict CQC inspection ratings of GP practices by:

- Measuring the association between GPPS results and CQC inspection ratings of GP practices.
- Building a predictive model of GP practice quality ratings that use GPPS results.
- Evaluating the potential of the predictive model for risk stratification.

METHODS

Design

Retrospective analysis of routinely collected data using decision tree modelling.

Variable identification

Six members of the team (including authors of this paper along with wider members of the Quality, Safety and Outcomes Policy Research Unit) mapped CQC inspection rating domains to questions in the 2018 GPPS. Additionally, stakeholder consultation resulted in CQC providing their own document showing GPPS items they consider to map to their inspection domains. This expert group of measurement specialists and academic professionals independently selected GPPS items that related to the CQC inspection domains (Caring, Responsive, Well-led, Safe, Effective). Results were reviewed and decisions to include items were based on where there was consensus. Where there was not unanimous agreement, these were further discussed among the expert group until consensus was reached. After synthesis, 10 questions were mapped to the Responsive domain, and six to the

Table 2 Mapping of GPPS items to CQC inspection rating domains

GPPS items mapping to 'Responsive' inspection domain (Accessibility of services and how responsive services are to people's needs)	GPPS items mapping to 'Caring' inspection domain (Relational aspects of care including communication with staff, confidence and trust and involvement in decisions)
<ol style="list-style-type: none"> 1. Generally, how easy is it to get through to someone at your GP practice on the phone? 2. How easy is it to use your GP practice's website to look for information or access services? 3. How satisfied are you with the general practice appointment times that are available to you? 4. How often do you see or speak to your preferred GP when you would like to? 5. Were you satisfied with the type of appointment (or appointments) you were offered? 6. Overall, how would you describe your experience of making an appointment? 7. Last time you had a general practice appointment, how good was the healthcare professional at giving you enough time? 8. Thinking about the reason for your last general practice appointment, were your needs met? 9. Have you had a conversation with a healthcare professional from your GP practice to discuss what is important to you when managing your condition (or conditions)? 10. Have you agreed a plan with a healthcare professional from your GP practice to manage your condition (or conditions)? 	<ol style="list-style-type: none"> 1. How helpful do you find the receptionists at your GP practice? 2. Last time you had a general practice appointment, how good was the healthcare professional at listening to you? 3. Last time you had a general practice appointment, how good was the healthcare professional at treating you with care and concern? 4. During your last general practice appointment, were you involved as much as you wanted to be in decisions about your care and treatment? 5. During your last general practice appointment, did you have confidence and trust in the healthcare professional you saw or spoke to? 6. Overall, how would you describe your experience of your GP practice?

CQC, Care Quality Commission; GP, General Practitioner; GPPS, General Practice Patient Survey.

Caring domain (table 2). No GPPS items were mapped to the Safe, Effective or Well-led domains.

Data sources

General Practice Patient Survey

The GPPS data set for 2018 was downloaded from the GPPS portal.²⁷ The survey had been sent to a sample of 2,221,068 individuals, and 758,165 questionnaires were completed (response rate 34%). The midpoint of the fieldwork period assigned as the fieldwork date, as the exact date could not be determined from the available documentation.

Care Quality Commission

The CQC Ratings data for inspection dates during 2018 were obtained by extract request from the CQC. GP surgeries, identified by Organisation code, could be present in the data set several times, with observations displayed at the level of the key question on a given inspection date. Duplicate instances were removed by retaining those records that represented a full review, identified as containing assessment for all domains.

Data merging

The CQC Ratings and GPPS scores were matched via the Organisation Code for the GP practice and a derived date that assigned the most recent past GPPS fieldwork date to the CQC inspection date. The merge was conducted by programmatically identifying the most recent fieldwork midpoint in the past of an inspection date.

Study variables

Predictors

The question fields that were mapped to responsive and caring (Table 2) were used as predictor variables, using the percentage of responses in most positive and most negative categories for each question as separate variables (ie, percentage of people selecting *Yes, definitely*; and percentage of people selecting *Not at all*). GPPS questions that were not mapped to any of the CQC inspection domains were not included in the analysis.

Outcomes

The outcome variables assessed were CQC inspection ratings for 'Overall' (a combined rating of all the key questions), which uses an ordinal scale (Outstanding, Good, Requires Improvement and Inadequate).

Analytical approach

Polyserial correlation was used to provide a measure of association between the predictor variables (percentages), as identified by the expert group and the outcome variable (four category ordinal).

The predictive modelling approach adopted was decision tree analysis, using conditional inference trees (CTree). CTree is a non-parametric class of regression trees that embed tree-structured models into a conditional inference procedure. For this study, CTree was implemented using the R function `party::ctree`.²⁸ The

algorithm applies recursive partitioning, at each step selecting the independent variable that provides the most information about the dependent target and dividing the data into subsets of cases (nodes) based on the values of the selected independent variable and according to statistical significance rules. The segregation occurs such that the data in each descendent node are more homogeneous than in the parental nodes.^{29 30} Partitioning continues at each node until it is no longer possible to create statistically distinct groups. CTree overcomes the disadvantages of other implementations of decision trees as it incorporates statistical significance testing and because the selection of split variables is not biased towards those with more categories.³¹ The algorithm selects the most appropriate form of model (regression or classification) at each node based on the type of variable (continuous or categorical); in this study, regression models were appropriate. Missing data for continuous predictors are handled by the algorithm through a system of surrogate splits: splits that preserve the distribution of the original split.³¹

The outcome of the decision tree was a set of terminal nodes with a predicted inspection rating and the error rate: the mismatch between the predicted rating and the actual inspection result. By identifying terminal nodes containing a high proportion of ratings in the 'Inadequate' or 'Requires Improvement' categories, the results provide both a predicted rating and a means of identifying practices at risk of a poor rating.

Patient and public involvement

Patient and public involvement representatives were involved in designing the study, by selecting the setting (primary care). The authors consulted stakeholders including CQC, NHS England and academics with a relevant research interest, about the research, and discussions shaped the focus of the study.

RESULTS

Restricting the data set to cases with an inspection date in 2018 and fieldwork year in 2018 left 1269 records, an additional 81 records had an inspection date in 2019 and fieldwork year in 2018 (total 1350). The ratings for these were predominantly 'Good' (72%, 974/1350). With 1350 rows and 64 candidate variables, there was missing data in 4.3% (3697/86,400), the algorithm automatically used surrogate splits, described previously.

Polyserial correlations

None of the polyserial correlations between predictors and outcome variable coefficients has an absolute value of greater than 0.354 (online supplementary material S1), the largest was for a negative rating of the overall experience of the service on the GPPS question. All predictor variables were included as candidates in the decision tree analysis.

Decision tree analysis

A three level tree provided an optimised classification of the data set provided. This tree assigned 97%

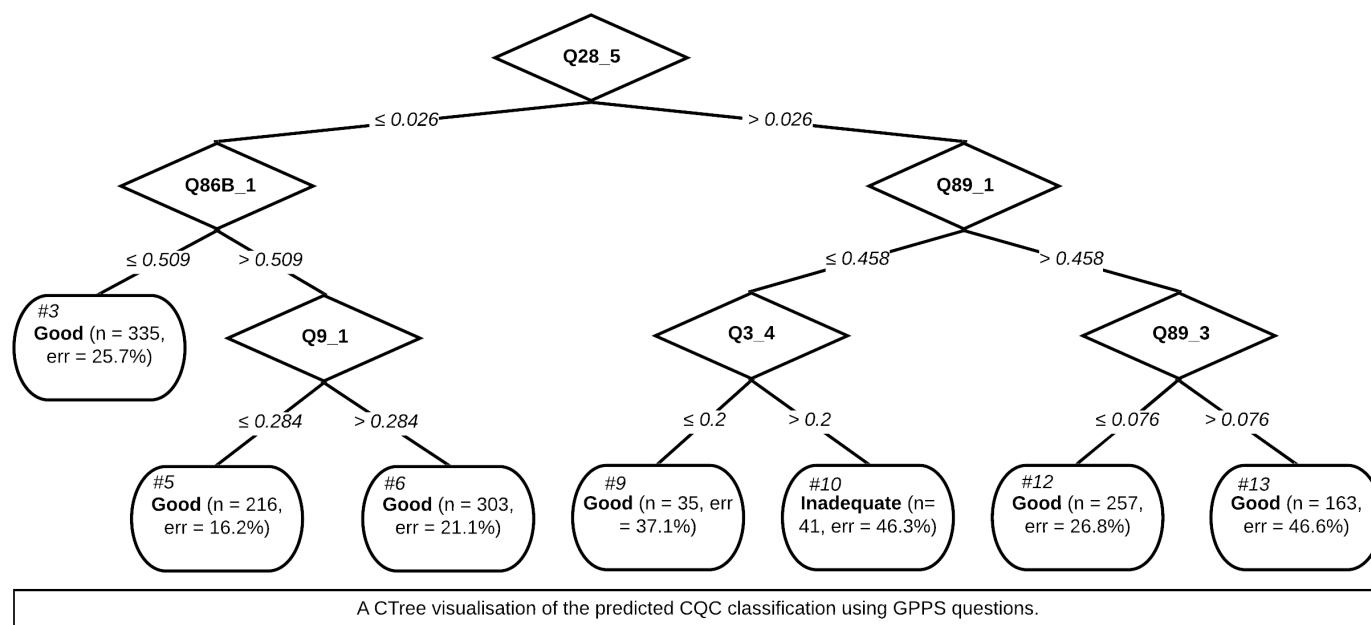


Figure 1 Decision tree analysis model results. GPPS, General Practice Patient Survey; CTree, conditional inference trees; CQC, Care Quality Commission.

(1309/1350) cases a rating of 'Good' and 3% (41/1350) 'Inadequate', an overall correct classification of 73% (988/1350; [figure 1](#)).

The indicators used in the tree model ([box 2](#)), were not necessarily those with the highest bivariate correlations.

To evaluate the performance of the tree in classifying practices, the tree classification rules were applied to the dataset and practices allocated to one of the terminal nodes (or to 'Unclassified' if any required data were missing). Both the tree-based classification and actual inspection rating were then collapsed into two categories: 'Good/Outstanding' and 'Inadequate/Requires Improvement'. The level of agreement corrected for chance (Cohen's kappa) between these classifications was 0.12, implying poor agreement. This is primarily a result of the large bulk of both classifications being in the 'Good/Outstanding' category (so that there is a high probability of chance agreement).

Box 2 Questions in the model

Q28_5: % answering *Very poor* to 'Overall experience of GP'

Q86B_1: % answering *Very good* to 'Last time you had a general practice appointment, how good was the healthcare professional at listening to you?'

Q9_1: % answering *Always* or *Almost always* to 'Frequency of seeing preferred GP'.

Q89_1: % answering *Yes, definitely* to 'During your last general practice appointment, were you involved as much as you wanted to be in decisions about your care and treatment?'

Q3_4: % answering *Not at all easy* to 'Ease of getting through to someone at GP practice on the phone'.

Q89_3: % answering *No, not at all* to 'During your last general practice appointment, were you involved as much as you wanted to be in decisions about your care and treatment?'

Examining the performance of the decision tree within each terminal node, [table 3](#) shows the distribution of inspection ratings for each node.

Practices classified in nodes 9–13 have a relatively high probability of being rated less than 'Good', followed by node 3. Of these, the highest risk of a poor classification was in nodes 10 and 13.

DISCUSSION

Quality of healthcare is an important topic worldwide, with evidence of poor quality services in all countries.¹ Primary care is critical in the provision of coordinated care.³ Exploring the association between GPPS data and CQC inspections allows us to understand whether routinely collected patient experience data can be used to identify poor quality healthcare within the NHS in England. The GPPS and CQC inspection outcomes both assess the quality of GPs, although from different perspectives and with a different focus. High quality primary care can lead to better outcomes while being cost effective.¹ Some data are collected in England, such as GP patient experience survey data, that is not available for other countries. Exploring how these data can be used for understanding care quality is useful both within and beyond the context of the NHS in England.

In our study, there was limited overlap in the themes covered in the two data sources, with narrow alignment of the mapping of GPPS questions to only two of the five key areas assessed by CQC. There were no strong bivariate associations between the GPPS variables and CQC outcome. We found that GPPS data did not provide for more accurate predictions of CQC ratings compared with simply assuming that all practices would be rated as 'Good' overall. Yet, by using a decision tree approach,

Table 3 Distribution of inspection ratings within tree classifications

Decision tree terminal	Inspection rating			
	Good/Outstanding		Inadequate/Requires Improvement	
Node and classification	N	%	N	%
0 Unclassified	42	77.8	12	22.2
3 Good	244	74.2	85	25.8
5 Good	196	92.9	15	7.1
6 Good	235	83.9	45	16.1
9 Good	21	63.6	12	36.4
10 Inadequate	9	22.0	32	78.0
12 Good	177	72.8	66	27.2
13 Good	85	53.5	74	46.5

we identified groups of GP practices at elevated risk of an 'Inadequate' or 'Requires Improvement' rating. The decision tree analysis does not include solely the most correlated variables, potentially due to non-linear effects or redundancies in the predictors.

The study has a number of important strengths. By accessing large volumes of publically available data, we developed a substantial dataset that takes account of the timing of GP practice surveys and inspections to take an ecological view of the potential utility of the GPPS for regulation. We used a thorough approach to map questions in the GPPS to inspection domains. However, a potential weakness is the lack of evidence from the GPPS on two of the five domains used by CQC. Due to the frequency of inspections (not all GP practices are inspected annually), the dataset was restricted to those practices inspected during 2018. We were unable to use data from multiple years due to significant changes in the GPPS questions and format between 2017 and 2018. Furthermore, the study used data aggregated to practice level, and this prevents consideration of variation in experience between subgroups of patients, which might be relevant to ratings of practice performance. The model represented only a small improvement in prediction compared with assuming all practices are 'Good', and should only therefore be used for prioritising instead of selecting where to target CQC inspections. Finally, as GPPS asks about experience over the previous 12 months, it was not possible to match patient experience to a specific inspection date and instead the midpoint period of the survey fieldwork date was used to match to inspection date.

In recognising the limitations of cross-sectional studies, although they are relatively quick to conduct, they provide only a snapshot of data, and we can therefore not guarantee representativeness. This means that the model may not apply to future data. This could be explored with further research to test the model prospectively on future datasets.

A 2020 study²⁶ found that using GPPS data to predict CQC inspection ratings gave results little better than

chance. That study used ordinal logistic regression models to develop predictions; by contrast, the decision tree approach that we used is more suitable for risk stratification rather than direct prediction. Our study thus adds to the evidence by showing that GPPS data could be used for risk stratification by identifying practices at risk of poor inspection ratings, even without being able to predict exact ratings for all practices. Our study also uses data from a more recent iteration of the GPPS, which was substantially redeveloped for 2018; thus we can conclude that Allen *et al's*²⁶ findings stand despite changes to the questionnaire, while identifying specific items in the updated survey that should be of regulatory interest.

Our research has important implications for policy-makers and regulators both within England and internationally. Inspection-led assessments of provider quality are onerous and expensive to conduct. In 2018/2019, CQC spent approximately £93million on inspections, representing 41% of the organisation's overall expenditure.³² The inspection model has been criticised for having mixed effects and some unintended and negative consequences, including increasing pressure on staff, and decreasing staff and patient morale.^{33 34} Even small improvements in the efficiency with which inspections are allocated have the potential to save significant sums of money for the system, and to reduce the unmeasured, but likely substantial, costs to provider staff preparing for and conducting inspection visits. Making further use of existing data as part of a surveillance model could support this goal, and the CQC should consider the potential to incorporate the findings of this study as part of the 'CQC insight' system. Practice managers and general practitioners should note the findings as evidence of the importance of patient feedback from the GPPS.

There remain a number of unanswered questions that would benefit from future research. First, we noted a narrow association between the content in the GPPS and the themes rated by CQC in its inspections. This raises questions about whether the content of the GPPS is sufficiently broad to provide a thorough understanding of quality from the perspective of patients. Further research

could engage patients in defining quality, and to consult them on the CQC inspection approach. However, an alternative interpretation may be that patients may not be expected to have direct experience of some areas that should naturally be covered as part of regulatory oversight,³⁵ that is, those relating to the safety, effectiveness and well-led domains. This argument is also predicated on the presumption that CQC ratings are themselves appropriate measures of quality, an issue beyond the scope of this paper. A second question is whether the utility of GPPS data for surveillance could be further improved with the use of additional data. Our analysis included responses from the GP patient survey, but it is possible that other public data about practices might be added to improve the decision tree model. Further research should explore the potential value of adding geographic, demographic or organisational data to improve risk stratification.

CONCLUSION

Questions in the GPPS were narrowly aligned to two of the areas evaluated by CQC in their GP practice inspections. GPPS data were not able to predict CQC ratings of GP practices with more accuracy than a model that simply assumes all practices will be rated as 'Good'. This reflects the preponderance of 'Good' ratings in the inspection dataset. However, the decision tree analysis provides an opportunity to target resources, using the terminal classification node 'misclassification' rate (the number incorrectly assigned when compared with the known rating) to indicate those targets. For instance, if a large proportion of GP practices would be incorrectly assigned as 'Good' in a node that is defined in terms of the responses to GPPS survey questions, constituent practices for that node should be targeted for CQC inspection. Whereas, if a node correctly classified most of the GP practices as 'Good', then constituent practices of that node should be placed at a lower priority for inspection. In addition, where the classification is not 'Good' those practices should be targeted for inspection to confirm or reject that prediction, regardless of the misclassification rate on this test data. This suggests that routinely collected GPPS data may, along with other data sources, be useful in identifying practices at greater risk of a lower rating, thereby allowing CQC to prioritise inspections for this risk group. This potentially allows for more efficient resource allocation.

Twitter Amy Tallett @amytallett, Alan J Poots @DrAlanJPoots and Chris Graham @ChrisGrahamUK

Acknowledgements The authors thank the wider members of the Quality, Safety and Outcomes Policy Research Unit for their input and advice. The authors are grateful to CQC for their support in providing data and information to assist with question mapping.

Contributors All authors contributed to the writing of this article and interpretation of findings. CG and JF were responsible for study inception and providing expert advice throughout the project. AT had oversight of the project including coordination and stakeholder consultation. RC acquired and cleaned and matched the data sets. AJP matched the datasets and conducted the analyses. SS developed the analysis

approach and contributed to the analyses. MP conducted the literature review and drafted the introduction.

Funding This research is funded by the National Institute for Health Research (NIHR) Policy Research Programme, conducted through the Quality, Safety and Outcomes Policy Research Unit, PR-PRU-1217–20702.

Disclaimer The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study was assessed using the UK Health Regulatory Authority online tools 'Is my study research?' and 'Does my study require ethics?'. Whilst intending to produce generalisable knowledge and thus considered research, NHS REC approvals were not needed, as the study is secondary use of aggregated data, with no randomisation, and no intervention. The study was reviewed by Picker Institute Europe's organisational Ethics Governance Board (reference 089) with the same outcome.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated and/or analysed for this study. The study is secondary use of publicly available, aggregated data.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Amy Tallett <http://orcid.org/0000-0002-3386-0731>
 Alan J Poots <http://orcid.org/0000-0002-2331-0904>
 Chris Graham <http://orcid.org/0000-0002-6410-8171>
 Michele Peters <http://orcid.org/0000-0002-0076-5981>
 Steve Sizmur <http://orcid.org/0000-0001-9776-8531>
 Julien Forde <http://orcid.org/0000-0001-7793-4328>

REFERENCES

- 1 National Academies of Sciences, Engineering, and Medicine. *Crossing the global quality chasm: improving health care worldwide*. Washington, DC: The National Academies Press, 2018.
- 2 Dixon J, Spenceley E, Howells A, *et al*. Indicators of quality of care in general practices in England. an independent review for the Secretary of state for health. health Foundation, London, 2015. Available: <https://www.health.org.uk/publications/indicators-of-quality-of-care-in-general-practices-in-england> [Accessed 25 Jul 2019].
- 3 The Organisation for Economic Cooperation and Development. *Caring for quality in health: lessons learnt from 15 reviews of health care quality*, 2017. Available: <https://www.oecd.org/els/health-systems/Caring-for-Quality-in-Health-Final-report.pdf> [Accessed 15 Sep 2020].
- 4 Department of Health. *High quality care for all; NHS next stage review final report*. London: department of health, 2008. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/228836/7432.pdf [Accessed 11 Sep 2019].
- 5 NHS digital: about the NHS outcomes framework, 2019. Available: <https://digital.nhs.uk/data-and-information/publications/ci-hub/nhs-outcomes-framework> [Accessed 16 Oct 2019].
- 6 Carinci F, Van Gool K, Mainz J, *et al*. Towards actionable international comparisons of health system performance: expert revision of the

- OECD framework and quality indicators. *Int J Qual Health Care* 2015;27:137–46.
- 7 Goodwin N, Dixon A, Poole T, *et al.* Improving the quality of care in general practice. Report of an independent inquiry commissioned by the Kings' Fund. The Kings' Fund, London, 2011. Available: <https://www.kingsfund.org.uk/publications/improving-quality-care-general-practice> [Accessed 26 Jul 2019].
 - 8 Blake A, Sparrow N, Field S. The CQC inspections: what they mean for general practice. *Br J Gen Pract* 2015;65:112–3.
 - 9 Care Quality Commission (CQC). Nhs GP practices. indicators and methodology, 2014. Available: <https://www.cqc.org.uk/sites/default/files/20141205%20GP%20IM%20Indicators%20and%20Methodology%20Guidance.pdf> [Accessed 25 Jul 2019].
 - 10 Care Quality Commission (CQC). The five key questions we ask, 2018. Available: <https://www.cqc.org.uk/what-we-do/how-we-do-our-job/five-key-questions-we-ask> [Accessed 25 Jul 2019].
 - 11 Care Quality Commission (CQC). How CQC monitors, inspects and regulates NHS GP practices, 2019. Available: <https://www.cqc.org.uk/sites/default/files/20190402%20How%20CQC%20regulates%20primary%20medical%20services%20GP%20PRACTICES.pdf> [Accessed 25 Jul 2019].
 - 12 Towers A-M, Palmer S, Smith N, *et al.* A cross-sectional study exploring the relationship between regulator quality ratings and care home residents' quality of life in England. *Health Qual Life Outcomes* 2019;17:22.
 - 13 Wake M, Green W. Relationship between employee engagement scores and service quality ratings: analysis of the National health service staff survey across 97 acute NHS trusts in England and concurrent care quality Commission outcomes (2012–2016). *BMJ Open* 2019;9:e026472.
 - 14 Coulter A, Locock L, Ziebland S, *et al.* Collecting data on patient experience is not enough: they must be used to improve care. *BMJ* 2014;348:g2225.
 - 15 Campbell J, Smith P, Nissen S, *et al.* The GP Patient Survey for use in primary care in the National Health Service in the UK—development and psychometric characteristics. *BMC Fam Pract* 2009;10:57.
 - 16 Burt J, Newbould J, Abel G, *et al.* Investigating the meaning of 'good' or 'very good' patient evaluations of care in English general practice: a mixed methods study. *BMJ Open* 2017;7:e014718.
 - 17 Burt J, Campbell J, Abel G, *et al.* Improving patient experience in primary care: a multi-method programme of research on the measurement and improvement of patient experience. *Programme Grants for Applied Research 5 National Institute for Health Research* 2017.
 - 18 Gibbons EJ, Graham C, King J, *et al.* Developing approaches to the collection and use of evidence of patient experience below the level of national surveys. *Patient Exp J* 2016;3:92–100.
 - 19 Anhang Price R, Elliott MN, Zaslavsky AM, *et al.* Examining the role of patient experience surveys in measuring health care quality. *Med Care Res Rev* 2014;71:522–54.
 - 20 Stein SM, Day M, Karia R, *et al.* Patients' perceptions of care are associated with quality of hospital care: a survey of 4605 hospitals. *Am J Med Qual* 2015;30:382–8.
 - 21 Griffiths A, Leaver MP. Wisdom of patients: predicting the quality of care using aggregated patient feedback. *BMJ Qual Saf* 2018;27:110–8.
 - 22 NHS England and IPSOS MORI. Gp patient survey. questionnaire redevelopment. ipsos mori London, 2018. Available: <https://gp-patient.co.uk/Files/GPPS%20Y12%20Questionnaire%20redevelopment%20report%20v1%20PUBLIC.pdf> [Accessed 25 Jul 2019].
 - 23 Abel G, Mavaddat N, Elliott M, *et al.* Primary care experience of people with long-standing psychological problems: evidence from a national survey in England. *Int Rev Psychiatry* 2011;23:2–9.
 - 24 Paddison CAM, Saunders CL, Abel GA, *et al.* Why do patients with multimorbidity in England report worse experiences in primary care? Evidence from the general practice patient survey. *BMJ Open* 2015;5:e006172.
 - 25 Nagraj S, Abel G, Paddison C, *et al.* Changing practice as a quality indicator for primary care: analysis of data on voluntary disenrollment from the English GP patient survey. *BMC Fam Pract* 2013;14:89.
 - 26 Allen T, Walshe K, Proudlove N, *et al.* Using quality indicators to predict inspection ratings: cross-sectional study of general practices in England. *Br J Gen Pract* 2020;70:e55–63.
 - 27 Ipsos Mori. Gp patient survey national report. Available: <https://www.gp-patient.co.uk/SurveysAndReports> [Accessed 08 Jan 2020].
 - 28 Hothorn T, Hornik K, Zeileis A. (n.d.). ctree: conditional inference trees. Available: <https://cran.r-project.org/web/packages/partykit/vignettes/ctree.pdf> [Accessed 05 Feb 2020].
 - 29 De'ath G, Fabricius KE. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 2000;81:3178–92.
 - 30 Witten IH, Frank E. *Data mining: practical machine learning tools and techniques*. Elsevier, 2005.
 - 31 Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. *Journal of Computational and Graphical Statistics* 2006;15:651–74.
 - 32 Care Quality Commission (CQC). Annual report and accounts 2018/2019. Available: https://www.cqc.org.uk/sites/default/files/20190812_annualreport201819.pdf [Accessed 05 Feb 2020].
 - 33 Smithson R, Richardson E, Robertson R, *et al.* Impact of the Care Quality Commission on provider performance. Room for improvement? The Kings' Fund, London, 2018. Available: <https://www.kingsfund.org.uk/publications/impact-cqc-provider-performance> [Accessed 22 Jul 2019].
 - 34 Rendel S, Crawley H, Ballard T. CQC inspections: unintended consequences of being placed in special measures. *Br J Gen Pract* 2015;65:e640–1.
 - 35 Rao MClarke A, Clarke A, Sanderson C, *et al.* Patients' own assessments of quality of primary care compared with objective records based measures of technical quality of care: cross sectional study. *BMJ* 2006;333:19.