How do financial incentives impact on the distribution of the quality of care? Evidence from the Advancing Quality (AQ) initiative

Thomas Mason, Yiu-Shing Lau, Matthew Sutton

University of Manchester

Email for correspondence: thomas.mason@manchester.ac.uk

Acknowledgements

This project was funded by the National Institute for Health Research Health Services and Delivery Research (NIHR HS&DR) programme (project number 08/1809/250). Visit the HS&DR website for more information. The views and opinions expressed are those of the authors and do not necessarily reflect those of the HS&DR programme, NIHR, NHS or the Department of Health. We are grateful to Linda Flavell of the NHS Greater Manchester Business Intelligence Service and Lesley Kitchen of the Advancing Quality Alliance (http://www.advancingqualityalliance.nhs.uk/) for previous discussions of this work and providing the data. We would also like to thank our colleague Eleonora Fichera for kindly providing us with data on trust characteristics.

Abstract

Background

Evidence suggests that financial incentives can increase the quality of care patients receive but may also lead to more exclusion of patients from the metrics used to measure performance. However, less is understood regarding the consequences for the distribution of quality. Providers may select easier patients or be induced to treat the harder patients that they would have neglected without the financial incentive.

Aims

In this study we consider: (i) whether patients' probabilities of being excluded from, or achieving, performance metrics depend on their characteristics and the characteristics of the hospitals that treat them; (ii) we also examine whether these probabilities have changed over time and whether quality differences have narrowed or widened over time.

Data

We investigate the specific case of the Advancing Quality (AQ) initiative incepted into hospitals in the North West of England. We linked two datasets, firstly we have data on characteristics of all patients across four health conditions (n=165,000); linked with records of whether quality was received. These data cover eleven quarters from October 2008 until June 2011 and comprise a unique dataset regarding who receives quality and who is declared ineligible.

Methods

Multinomial logistic regression

Results

The distribution of quality varies by quality indicator. There is a substantial age gradient in both rates of achievement and ineligibility which persists over time. There is also evidence of significant differences across deprivation and gender. Our estimates suggest a change over time in the distribution of quality from being pro-poor to pro-rich. We find evidence of correlations between hospital characteristics and exclusion rates.

Implications

Financial incentive initiatives impact on equity via the distribution of quality. Hospitals can use ineligibility reporting to maximise revenue.

Introduction

A wide variety of 'pay-for-performance' (P4P) initiatives have been adopted in hospitals across the world. Typically, the primary if not sole objective of these initiatives is to improve the quality of care provided to patients and thus improve health outcomes. Increased attention is being given to evaluating the impact of P4P schemes on the quality of care provided in addition to a limited number of analyses of the impact on health outcomes. However, little evidence currently exists in relation to the impact of P4P on the distribution of quality, a potential unintended consequence which may arise from imposing P4P.

In this paper, we analyse the impact of the 'Advancing Quality' (AQ) initiative on the distribution of quality and how the distribution of quality changed over time. We ask two key questions: (i) whether patients' probabilities of being excluded from, or achieving, performance metrics depend on their characteristics and the characteristics of the hospitals that treat them; (ii) we also examine whether these probabilities have changed over time and whether quality differences have narrowed or widened over time.

In answering these questions, we provide background on the use of P4P internationally and within the NHS; summarising both the rationale for its adoption, and the current evidence base relating to its impact on equity. We then describe the AQ initiative, followed by our methodological approach and the data set used. Results are then presented and the implications discussed. Finally, we draw conclusions based on our analysis.

Background to P4P

Research regarding the health care 'quality' has produced evidence that quality varies substantially between patients, including in countries with well-resourced health systems (van Herck et al. 2010). P4P has been adopted by health policymakers as a mechanism by which 'performance' and thus 'quality' can be improved: by linking the remuneration of providers with reported achievement on quality indicators. Linking payments to performance has, both internationally and in the UK, been one stage in broader gradual reforms that have promoted the use of 'quasi-markets' and payment mechanisms to address concerns regarding incentives for efficiency and quality improvement. The World Health Organization (WHO) noted that greater 'strategic purchasing' ought to be embraced to affect quality improvements (WHO 2000); and in 2001 the US Institute of Medicare recommended that purchasing ought to be aligned to achieve quality improvements (IoM 2001). Both recommendations essentially translate as calling for the reform of payment mechanisms to increase the incentives for providers to improve quality.

Reforming NHS Payment Mechanisms

Since the purchaser-provider split was established in the NHS, the relationship between the parties has been governed by contracts. These payment mechanisms have been reformed several times: first, block contracts

were replaced with activity based financing since block contracts had failed to link volume or quality to revenue. Activity based financing pays hospitals on a case-mix adjusted average cost per case basis, establishing a link between revenue and volume. Whilst this payment model rewards activity indiscriminately of quality, it was hoped that this model would create an indirect link between quality and revenue: as providers compete for patients based on quality, they would improve their quality to attract more patients and make higher revenues.

NHS funding mechanisms have been further reformed since the introduction of activity based financing aiming to improve quality of care (for both during the admission and post discharge). For example, from April 2011 until April 2012, hospitals in England were not reimbursed for emergency readmissions within thirty days of an elective readmission. This is also being implemented for hospitals in the US reimbursed under Medicare from October 2012 onwards. Payment mechanisms continue to be altered as policymakers seek to make a more explicit link between quality and revenue than the link created by activity based financing.

In addition to modification of the activity based payment mechanism to incorporate incentives to boost quality, P4P mechanisms have been adopted in some clinical areas to compliment activity based funding. These mechanisms have typically related payment to process measures. These are usually binary indicators relating to discrete steps on a clinically established 'best practice pathway' which demonstrate whether a patient has received that particular dimension of 'quality'.

It is clear that, both in the context of the NHS and further afield, that P4P has been embraced (as with previous reforms to payment mechanisms) as a means to improving quality given financial constraints. The evidence regarding the impact on quality is mixed. However, the impact of P4P on health inequalities has been far less extensively assessed. One of the unintended consequences of adopting P4P might be to increase health inequalities if the schemes impact differentially of societal groups. For example, the 'inverse equity hypothesis' proposes that affluent segments of society benefit disproportionately from such interventions (Victora et al. 2000). A socioeconomic gradient in use could be caused by the demand side: the result of disparities in the behaviour of different groups in society (for example differences in uptake rates between socioeconomic groups). However, it could also be the consequence of provider behaviour (supply side): for example, providers could exclude *eligible* individuals from their achievement score calculations in order to attain a higher performance score; or providers serving deprived populations may behave differently relative to those serving less deprived populations as their rational expectations of a bonus are much lower.

These examples are hypothetical reasons to expect that P4P initiatives would impact on health inequalities, the reduction of which is a formalized NHS goal (DH 2003). The WHO has previously asserted that all government policies and programmes be analysed for their impact on health and health equity (WHO 2008).

On this basis, the impact of the introduction of P4P schemes on equity is a key research question relevant for policymakers. The following section reviews the current evidence base evaluating this important question.

Evidence on the impact of P4P upon equity

A systematic review undertaken by Alshamsan et al. (2010) found 22 relevant articles, the majority of which were observational studies conducted in the UK evaluating the impact of the Quality and Outcomes Framework (QoF). The QoF was introduced in 2004 in the NHS and rewarded providers of primary care services for achievement on a large number of evidence-based quality indicators, particularly emphasising the management of common cardiovascular conditions (over half of the indicators relate to CHD, diabetes, stroke and hypertension). The introduction of this P4P scheme was intended to address the objectives stated in the NHS Plan (DH 2000): (i) Greater plurality of provision of primary care services; (ii) greater access to care and patient choice; (iii) more flexible contractual arrangements; and (iv) increased focus on paying for 'performance'.

The QoF defined performance as being comprised of four dimensions: (i) clinical care; (ii) organizational characteristics; (iii) patient experience; and (iv) provision of 'additional' services. Practices were awarded points based on their performance as defined above and their payments were linked to their points total. Relatively higher performing practices earned higher incomes, with payment varying according to population size.

In their review, Alshaman et al. (2010) identify that studies have considered the impact of P4P upon three dimensions of inequality: socioeconomic inequalities; age and gender inequalities; and ethnic inequalities. This classification most probably stems from the patient characteristics available in relevant data sets, but it is useful and covers primary dimensions of population heterogeneity.

Alshamsan et al. (2010) found that socioeconomic status is the most frequently examined dimension of equity and that studies have commonly adopted a cross sectional design, scrutinizing the associations between area deprivation score and quality after the implementation of the QoF. The majority of these studies found a significant relationship between relatively higher deprivation and lower achievement scores on the QoF. The magnitude of these differences, however, was small and decreased after the first year of the QoF. Only one study (Sutton & McLean 2006) found a positive association between deprivation and quality; however this particular study was based on a small sample; the study was conducted a largely rural and relatively deprived area in Scotland, which may explain its divergence from the majority of studies. A further study found that an initial socioeconomic gradient favouring relatively less deprived populations at the inception of the QoF in 2004-5 was substantially reduced over a three-year period (Doran et al. 2008).

Most studies assessing the impact on socioeconomic equity have tended to focus on the associations between socioeconomic status and 'rates of achievement', where achievement is defined as the proportion of

the eligible population in receipt of quality. However, Sigfrid et al. (2006) looked at associations between deprivation and rates of exception reporting for sixteen diabetes indicators, noting that providers could exclude patients from the achievement calculation and, ceteris paribus, increase their achievement score. Thus, providers may be induced to exclude eligible patients and if their exclusions are not uniform across population groups, then such exclusion will impact on equity.

Sigfrid et al. (2006) established that providers operating in more deprived areas were more likely to report exceptions and raised concerns that reported achievement in the QoF could disguise broader effects on equity. McLean, Sutton & Guthrie (2006) also incorporate exclusion rates into their analysis: they assess both reported performance (achievement in the eligible population) and an additional measure 'delivered quality' (achievement in the overall registered population). Their study found that delivered quality rates were significantly lower in more deprived populations.

Gravelle et al. (2010) looked for differences in reported disease rates between provider and differences in exception rates both within and between providers using QoF data. They use two tests to do so: first, they consider whether practice characteristics are correlated with exclusion rates; and in addition they note that the bonus structure of QoF would marginally benefit some practices more than others - and exploit this to examine differences in exception reporting between practices whose scope to benefit from improved performance differed. The results of both tests imply that practices gamed their performance measures: practices characteristics were correlated with exception rates; and practices whose expected marginal benefit from an improved performance was greater had higher exception rates. Gravelle et al. (2010) also suggest that the effects of the incentive scheme on achievement (in the eligible population) are weaker than its effects on exclusion reporting behaviour.

Crawley et al. (2009) examined the impact of the introduction of the QoF on social class inequalities for Chronic Heart Disease (CHD) and hypertension patients using achievement rates as their dependent variable using individual-level data from the Health Survey for England. Overall, they found little evidence of significant differences between social groups suggesting that the QoF gains were distributed equitably.

Some studies have adopted a 'before and after' methodological approach. One study established for individuals with coronary heart disease, being from a relatively deprived area reduced the probability that their practice had recorded their smoking status (McGovern et al. 2008). However, their overall results are ambiguous as the opposite result was found on a different quality indicator and, for many indicators, there were no significant differences. Another evaluation using similar methods found that Stroke patients from more deprived backgrounds were less likely to receive quality (Simpson et al. 2006).

Regarding inequalities in between age groups and genders, four studies have found mixed results. For some particular QoF indicators, equity before and after the introduction of the QoF was unchanged, for other

indicators inequalities either increased or decreased. Overall, on the basis of the current evidence base, the global effect of the QoF on age and gender inequalities is ambiguous.

This is also the case with the studies that have examined the impact of the QoF on ethnic inequalities: the impact varies by indicator and in there are findings of no change, reduced and increased inequality. One study from the US spanning 3449 hospitals assessed the impact of P4P on the distribution clinical quality for three health conditions (Karve et al. 2008). For two of these conditions, findings indicated that having a patient population with a higher proportion of ethnic minorities was associated with a lower probability of improvement (Karve et al. 2008). The majority of studies in the US do not examine the impact of P4P on equity.

The majority of the evidence base we have considered analyses the impact of the QoF on the distribution of quality. However, it is unclear from the published evidence as to whether changes to inequalities are attributable to the introduction of the QoF, or some other predicating effect (for example, other quality improvement initiatives such as the National Service Frameworks). The QoF was introduced across the board, making a quasi-experimental approach less straightforward as there were no (English) counterfactual cases (where the QoF was not introduced). Studies have been unable to examine underlying trends in quality causing an identification problem (Alshamsan et al. 2010). Furthermore, most studies used provider level data and not individual level data with patient characteristics such as age, gender, etc.

Having undertaken a systematic review of the current evidence base, Alshamsan et al. (2010) made some recommendations for future studies. They note that future studies should: (i) include analyses of patient level data; (ii) consider various dimensions of inequality such as age, gender, ethnicity, and socioeconomic status; (iii) take into account current trends in underlying inequality; and (iv) examine the impact on inequalities over a wider time horizon. They also note that piloting P4P schemes would allow for non-intervention comparator groups and the use of quasi experimental methods.

Advancing Quality (AQ)

The AQ initiative was introduced in 2008 in all 24 NHS hospitals in the North West of England that provide emergency care. Hospitals are required to collect and submit data on 28 conditions relating to five health conditions: Acute Myocardial Infarction (AMI); Pneumonia (PN); Hip and Knee Replacement (HK); Heart Failure (HF); and Coronary Artery Bypass Grafting (CABG). Patients attending a participating hospital with one of the five included conditions can receive one of three outcomes for each quality indicator. These mutually exclusive / exhaustive outcomes are:

1. **Excluded** on an AQ indicator: this signifies that the hospital did not consider the patient eligible for this quality measure under agreed rules

2. **Achievement** on an AQ indicator: this signifies that the patient was eligible for the quality measure and this was subsequently provided to the patient
3. **Failure** on an AQ indicator: this signifies that the patient was eligible for the quality measures but that this was not subsequently provided by the hospitals

Each indicator of clinical quality is measured as the ratio of achieved patients divided by the total population of patients reported as eligible by the hospital: that is, not including exclusions. Whilst it would be hoped that hospitals would seek to increase their revenue by increasing the numerator (the number of patients in receipt of quality), they can also increase their score by reducing the denominator (increasing the number of exclusions). The reasons for which a hospital can legitimately exclude a patient are discussed below.

Exclusion Rules

AQ developed detailed rules for excluding patients. There are general exclusion reasons for all indicators within a health condition and specific exclusion reasons for most indicators. Patients under 18 years of age should be excluded, as should those involved in clinical trials. Patients who die during the hospital stay are admitted for comfort measures only or who are discharged to hospices can also be excluded. Patients who self-discharge can generally be excluded from indicators that require completion of an activity before discharge.

Some indicators allow exclusion when patients are transferred in from or transferred out to another hospital. Many indicators allow providers to exclude patients because of contraindications, which may include competing diagnoses, or prior or planned surgical procedures. For pneumonia, patients can be excluded if the hospital had no evidence from initial diagnostic procedures that the patient had pneumonia on admission. Patients can also be marked as excluded for several indicators if the period of hospital stay or surgery is very short or too short for sufficient time to have elapsed to measure achievement.

Patients can be excluded from some indicators if another indicator requires an action to be performed and the patient has failed that earlier indicator (e.g. it cannot be judged whether the antibiotic selected was appropriate if no antibiotic was provided).

Financial Rules

Financial rewards were based on the relative performance of the participants. At the end of the first year, participants whose performance was in the top quartile received a bonus payment equal to four percent of the revenue they receive for the patient activity for this condition. The next-best performing quartile received bonus payment equal to two percent of revenue. The remaining participants did not receive bonus payments and there were no penalties for relatively poor performance. The total value of bonuses paid to participating hospitals was £3.2 million.

In the following six months, all providers received a bonus if they exceeded the median performance level reported in the first year. Providers that received this bonus could get additional bonuses for being a top improver or a top performer. Total payments in these six months were £1.6 million. Beyond this, AQ merged with 'CQUIN' for which there are no financial bonuses. Instead, providers have income withheld if they do not achieve performance levels that they negotiate with the local commissioner.

When AQ was first introduced, Chief Executive Officers (CEOs) of hospitals collectively decided to assign bonuses to clinical teams responsible for earning. These payments could not be taken as personal income; instead they would be reinvested into improving clinical quality.

Methods

Our outcome variable is categorical and our choice of methods reflects this. Multinomial methods model a single outcome from categories that are mutually exclusive. The multinomial logit model is our specific choice based on considerations relating to simplicity of computation.

We estimate individual *j*'s probability of receiving outcome *m* for indicator *i* from provider *k* at time *t*:

$$Prob(y_{ijkt} = m) = \frac{exp(\mathbf{\Phi}'_{\mathbf{m}}\mathbf{\Upsilon}_{\mathbf{i}} + \mathbf{B}'_{m}\mathbf{X}_{j} + \mathbf{\Lambda}'_{m}\mathbf{\Pi}_{k} + \mathbf{K}'_{m}\mathbf{T}_{t})}{\sum_{l=1}^{M} exp(\mathbf{\Phi}'_{\mathbf{l}}\mathbf{\Upsilon}_{\mathbf{i}} + \mathbf{B}'_{l}\mathbf{X}_{j} + \mathbf{\Lambda}'_{l}\mathbf{\Pi}_{k} + \mathbf{K}'_{l}\mathbf{T}_{t})} \qquad (\mathbf{1})$$

where $\boldsymbol{i}$**=1, 2, ...,** $\boldsymbol{I}$**;** $\boldsymbol{j}$**=1, 2, ...,** $\boldsymbol{J}$**;** $\boldsymbol{k}$**=1, 2, ...,** $\boldsymbol{K}$**;** $\boldsymbol{t}$**=1, 2, ...,** $\boldsymbol{T}$**;** $\mathbf{\Upsilon}_{\mathbf{i}}$ is a vector of indictor-specific dummy variables; $\mathbf{X}_{\boldsymbol{j}}$ is a vector of patient characteristics (age; gender; ethnicity; income deprivation); $\mathbf{\Pi}_{\boldsymbol{k}}$ is a vector of providers charactersitcs (foundation trust status; hospital budget position; hospital quality rating); and $\mathbf{T}_{\boldsymbol{t}}$ is a vector of quarter-specific dummy variables**.**

We estimate model (1) across all indicators, individuals, time periods and providers. In addition, we estimate three models for three separate time periods. This reflects the fact that the bonus structure for these three periods changed for each period. We therefore estimate an additional set of model for three time periods $\boldsymbol{t}$***=1, 2, 3*****:**

$$Prob(y_{ijkt} = m) = \frac{exp(\mathbf{\Phi}'_{\boldsymbol{mt*}}\mathbf{\Upsilon}_{\mathbf{i}} + \mathbf{B}'_{\boldsymbol{mt*}}\mathbf{X}_{\boldsymbol{j}} + \mathbf{\Lambda}'_{\boldsymbol{mt*}}\mathbf{\Pi}_{\boldsymbol{k}})}{\sum_{l=1}^{M} exp(\mathbf{\Phi}'_{\boldsymbol{lt*}}\mathbf{\Upsilon}_{\mathbf{i}} + \mathbf{B}'_{\boldsymbol{lt*}}\mathbf{X}_{\boldsymbol{j}} + \mathbf{\Lambda}'_{\boldsymbol{lt*}}\mathbf{\Pi}_{\boldsymbol{k}})} \qquad (\mathbf{2})$$

In this paper, we are asking (i) whether patients' probabilities of being excluded from, or achieving, performance metrics depend on their characteristics and the characteristics of the hospitals that treat them; and (ii) whether these probabilities have changed over time and whether quality differences have narrowed or widened over time.

The estimates on the patient characteristics will provide evidence as to whether hospitals select patients based on their characteristics. Our second set of models for which we estimate separate models for the three different bonus structure periods will provide evidence as regard the impact of AQ on equity *over time*. The estimates on our provider characteristics will be informative as regards potential gaming, as true exception and achievement rates are determined by the characteristics of the population served by the provider but should *not* be correlated with the characteristics of providers (such as whether the providers has relatively more financial autonomy). If an individual's probability of being exception reported is correlated with hospital characteristics this would suggest that hospitals are gaming[1] exceptions provided we have not omitted patient characteristics that are correlated with incorporated hospital characteristics.

Data

We obtained a unique patient level dataset from the North West Advancing Quality team in the NHS. This dataset is an extract from Secondary Users Service (SUS) extracting AQ qualifying patients using ICD10 codes (International Classification of Diseases version 10). The patient list is presented to NHS trusts in the North West to remove patients who were incorrectly identified as a part of AQ. From October 2009 onwards, NHS trusts in the North West were able to add unidentified patients to the patient list. The data are then merged at an individual level with the Quality Measures Reporter (QMR) which contains the information entered by the provider on the clinical process measures which each patient has been given, failed to be given or excluded from.

The Foundation Trust status of each Trust was obtained from the Monitor website. Ratings of each Trust's quality of care and financial management were taken from the Annual Health Check published by the Healthcare Commission. We use the ratings from 2007/8 and FT status at 2007, before AQ was introduced. Additionally, we use income deprivation indices extracted from The English Indices of Deprivation 2010. These are measured for lower-level super output areas (LSOAs) and represent the proportion of the area population receiving benefits from the State on the grounds of low income. We do not know the LSOA of residence for patients treated in the first four quarters. Throughout the period therefore, we used for each patient the average income deprivation of patients treated at the same hospital site. We examined the consequences of this aggregation in the seven quarters for which we do know the LSOA of residence of patients (see Discussion).

Overall, our merged data span eleven quarters from October 2008 until June 2011, totalling around 160,000 individuals and covers four health conditions: hip and knee replacement; pneumonia; acute myocardial infarction (AMI); and heart failure.

[1] Deliberately excluding eligible patients to achieve higher scores

Results

We present descriptive statistics in Table I, and Tables II and III give the estimated marginal effects yielded by models (1) and (2) respectively, and OPCS-4 procedure codes.

In our first model, the estimated marginal effects of the time-dummy variables reveal that, relative to the first period, achievement rates have increased over time and exclusion rates have decreased over time (Table II).

Patient Characteristics

Overall, individuals whose income deprivation score was relatively higher were less likely to be excluded and were also more likely to be in receipt of quality. A ten percentage-points increase in the deprivation score assigned to an individual is estimated to decrease the probability of being excluded by 0.78% and increases the probability of receiving quality by 0.42% (Table II).

However, our period-specific models reveal substantial changes within the overall AQ period: in the first period (October 2008 to September 2009), a ten percentage-points increase in the deprivation score of the hospital decreases the probability of being excluded by 3.9% and increases the probability of achievement by 2.9%. For the period October 2009 to March 2010, deprivation is not estimated to have had any significant effect. In the final period (April 2010 to June 2011), we find the (qualitatively) opposite result: the marginal effect of ten percentage-points increase in the deprivation score of the local area lowers the probability of achievement by 0.55% and increase the likelihood of exclusion by 0.59%.

As regards gender inequalities, males are marginally less likely to receive quality but the effect is modest. There are no significant differences in exception reporting rates between males and females; and we found no evidence of a time gradient in gender effects.

There are significant differences in achievement and exclusion rates between our age categories compared to the reference category (18-44 year olds). From the estimated marginal effect computed for model (1), we can see that, compared to their older counterparts, younger patients are less likely to be excluded and more likely to receive quality. The relationship between likelihood of achievement / exclusion and age is estimated to approximate a monotonically increasing function. Age inequalities increased substantially between the initial and second phase of AQ, before subsiding slightly in the final period.

As regards inequity across ethnic groups, we found that, compared to white patients, patients of mixed ethnicity were 1.9% less likely to be excluded and 1.3% more likely to receive quality. We included a category for those patients coded as ‘missing’ and individuals coded as such are estimated to be more likely to be excluded from quality (by 0.6%) and less likely to be in receipt of quality (by 0.9%). This may be

because being 'missing' is correlated with being in a health state in which individuals could not provide this information.

Provider Characteristics

Over the entire period, hospitals' which had previously acquired foundation trust (FT) status[2] were 0.7% more likely to exclude patients and FT status had no impact on the probability of achievement. However, our period-specific models reveal that the effects of FT status changed over time. Initially, FT status increased the probability of achievement by 1.9% and did not impact of the probability of exception reporting. However, between October 2009 and March 2010, FT hospitals were estimated to be 1.3% more likely to exclude patients; and there was no relationship between receiving quality and FT status. In the final period, the relationship between FT status and exclusion was unchanged; the probability of achievement was 1.5% reduced at a FT hospital. This may reflect an interaction with the trajectory of the AQ bonus structure.

Hospitals with 'good' fiscal positions were 1.1% more likely to exclude patients and 1.9% less likely to provide quality, compared to those with 'excellent' positions. Having a 'poor' fiscal position had no effect.

Hospitals that previously received a 'poor' quality score from the Healthcare Commission were 2.7% less likely to exclude patients and 2.7% more likely to provide quality than those in prior receipt of an 'excellent' score. This suggests that 'poorer quality' providers offered better care to AQ patients under AQ.

Discussion

Overall, the probability of receiving quality at an AQ hospital was around 7% higher in April 2011 (final quarter) compared to the first quarter; and over the same period, the probability of being excluded decreased by around 3%. This indicates that both 'population achievement' and 'reported achievement' increased under AQ.

With regard to the actual distribution of quality, we found that quality gains were not equally distributed across patients or provider types. We found that hospitals a*chievement rates* vary by the characteristics of both patients and hospitals, and there are clear and substantial inequalities across age and deprivation. In the case of deprivation, these inequalities have only emerged since April 2010. Achievement rates have also varied depending on the characteristics of providers.

We found evidence of gaming of exception reporting. Whilst it can be expected that legitimate exclusions are correlated with the characteristics of the population served by the hospital (and our results are consistent with this), exclusions should not be correlated with the characteristics of providers. Nonetheless we have shown that foundation trust status, quality score and hospitals' fiscal positions are correlated with rates of

[2] Foundation trust hospitals have greater autonomy (particularly in relation to financial management)

exception reporting. These characteristics might plausibly impact on providers' costs or have other effects but we would not expect that these factors would impact of the rate of exception reporting.

The exception reporting practices of hospitals have changed over time (Table III) and, in particular, exception reporting / quality inequalities across deprivation have emerged in the absence of positive financial incentives (April 2010 to June 2011). Foundation trust hospitals appear to have performed relatively well in the first two periods for which bonuses were rewards, and contrastingly poorly when the carrot was removed and replaced with a stick (April 2010 to June 2011). FT hospitals have greater financial autonomy than non-FT hospitals and thus may consequently be more sensitive to changes in the incentive structure. They may be a useful yardstick of providers' likely responses to changes in the financial rules.

AQ was imposed on all hospitals in the North West and thus could not self-select into the incentive scheme. Whilst this eliminates a source of sample selection bias, it also means that there were no non-intervention cases in the North West ruling out a quasi experiment within the North West.

As previously noted, we use income deprivation indices extracted from The English Indices of Deprivation 2010 but we do know the LSOA of residence of patients. This measure of deprivation is not ideal as it does not reflect variation in deprivation within populations treated at the same hospital.

Conclusion

This paper is a unique contribution to the current evidence base relating to the impact of P4P on equity. We use, for the first time, data from the AQ initiative in which all hospitals were required to participate. We use, unlike many studies, data at the level of the individual and our dataset contains over seven hundred thousand observations on 160,000 individuals. We have used multinomial logistic regression to reflect the fact that there are three outcomes to which individuals are assigned.

Whilst the evidence base relating to the effects of P4P on quality is substantial and expanding, questions relating to which groups receive quality and which groups are excepted from quality have received less attention. In this paper, we sought to answer two key questions: have patients' probabilities of being excluded from, or achieving, performance metrics depend on their characteristics and the characteristics of the hospitals that treat them; and have these probabilities have changed over time and have quality differences have narrowed or widened over time. Our results suggest that the answer to both these questions is yes: receipt of quality and exclusion from AQ depends on both the characteristics of patients and providers; and quality differences have, for the most part, either remained stable over time or worsened. The correlations we found between exception reporting and provider characteristics may be the consequence of gaming by providers.

The reduction of health inequalities is an explicitly stated objective of the NHS. P4P schemes have been adopted in further reforms to payment mechanisms intended to improve quality. Whilst P4P schemes may have had quality gains (as in the case of AQ), it has also been posited that one of their unintended consequences might be an increase in health inequalities across population groups. We find no evidence that inequalities across population groups have been reduced at all; and we find clear evidence of inequalities in age and deprivation that have increased or emerged over time. On this basis, policymakers may wish to note that whilst introducing P4P might bring about quality improvements, it could also unintentionally bring about a worsening of existing health inequalities.

References

Alshamsan R, Majeed A, Ashworth M, Car, J & Millett, C (2010), *Impact of pay for performance on inequalities in health care: systematic review*, Journal of Health Services Research and Policy, 15(3): 178-84.

Crawley D, Ng A, Mainous III AG, Majeed A & Millett, C (2009), *Impact of pay for performance on quality of chronic disease management by social class group in England,* Journal of the Royal Society of Medicine, 102: 103-7.

Department of Health (2003), *Tackling health inequalities: A Programme for Action*, London: DH

Department of Health (2000), *The NHS Plan: a plan for investment, a plan for reform*, London: DH

Doran T, Fullwood C, Kontopantelis E & Reeves D (2008), *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, 372: 728-36.

Gravelle H, Sutton M & Ma A (2010), *Doctor behaviour under a pay for performance contract: treating, cheating and case finding?,* The Economic Journal, 120: 129-56.

Institute of Medicine (2001), *Crossing the Quality Chasm: A New Health System for the 21st Century*, Washington DC: National Academy Press

Karve AM, Ou FS, Lytle BL & Peterson ED (2008), *Potential unintended financial consequences of pay-for-performance on the quality of care for minority patients*, American Heart Journal, 155: 571-6.

McGovern MP, Boroujerdi MA, Taylor MW, et al. (2008), *The effect of the UK incentive-based contract on the management of coronary heart disease in primary care*, Family Practice, 56: 277-9.

McLean G, Sutton M & Guthrie B (2006), *Deprivation and quality of primary care services: evidence for the persistence of the inverse care law from the UK Quality and Outcomes Framework*, Journal of Epidemiology and Community Health, 60: 917-22.

Sigfrid LA, Turner C, Crook D & Ray S (2006), *Using the UK primary care Quality and Outcomes Frameowork to audit health care equity: preliminary data on diabetes management*, Journal of Public Health, 28: 221-5.

Simpson CR, Hannaford PC, Lefevre K & Williams D (2006), *Effect of the UK incentive-based contract on the management of patients with stroke in primary care*, Stroke, 37: 2354-60.

Sutton M & McLean G (2006), *Determinants of primary medical care quality measured under the new UK contract: cross sectional study*, BMJ, 332: 277-99.

Van Herck P, De Smedt D, Annemans, L, Remmen R, Rosenthal, M & Sermeus, W (2010), *Systematic Review: Effects, design choice and context of pay-for-performance in health care*, BMC Health Services Research, 10: 247.

Victora CG, Vaughan JP, Barros FC, Silva, AC & Tomasi E (2000), *Explaining trends in inequities: evidence from Brazilian child health studies*, Lancet, 356(9235): 1093-98.

World Health Organization (2000), *World Health Report 2000: Health Systems Improving Performance*, Geneva: World Health Organization

World Health Organization (2008), *Closing the Gap in a Generation: Health Equity through Action on the Social Determinants of Health. Final Report of the Commission on Social Determinants of Health*, Switzerland, Geneva: World Health Organization

TABLE I

Variable	Description	Mean			
		All period	*Oct-08 to Sept-09*	*Oct-09 to Mar-10*	*Apr-10 to June-11*
Exclusion		0.48	0.49	0.47	0.47
Success		0.46	0.43	0.46	0.48
Failure		0.06	0.08	0.07	0.05
Age	*Patient age at beginning of episode*	72.27	72.26	72.04	72.36
Male	*Dummy variable for male*	0.510	0.513	0.509	0.508
Ethnicity	*Broad ethinic categories defined by NHS*				
white		0.847	0.702	0.910	0.911
mixed		0.001	0.001	0.001	0.001
asian		0.018	0.014	0.017	0.020
black		0.004	0.003	0.004	0.004
other		0.006	0.004	0.005	0.007
missing		0.125	0.276	0.062	0.057
Foundation Trust	*Dummy variable for foundation trust status **as at year 2007***	0.310	0.281	0.324	0.322
Income deprivation	*Patients area income deprivation aggregated to Hospital location*	0.181	0.179	0.181	0.182
Quality rating	***2007** Quality of care score (HC)*				
Excellent		0.309	0.331	0.294	0.300
Good		0.577	0.575	0.593	0.571
Fair/Poor		0.115	0.094	0.113	0.128
Financial management rating	*Financial rating of NHS Trusts **in 2007***				
Excellent		0.467	0.450	0.480	0.473
Good		0.274	0.267	0.259	0.283
Fair/Poor		0.259	0.283	0.261	0.244
Observations		**148070**	**45526**	**29438**	**73106**

TABLE II

	Exclusions		Success	
	ME	**z-stat**	**ME**	**z-stat**
Period (Oct-08 as base category)				
Jan-09	-0.0274***	(-5.57)	0.022***	(4.75)
Apr-09	-0.036***	(-7.37)	0.033***	(7.27)
Jul-09	-0.038***	(-7.85)	0.038***	(8.29)
Oct-09	-0.051***	(-10.81)	0.057***	(12.8)
Jan-10	-0.055***	(-11.69)	0.069***	(15.52)
Apr-10	-0.064***	(-13.29)	0.082***	(18.13)
Jul-10	-0.057***	(-11.84)	0.081***	(17.83)
Oct-10	-0.052***	(-11.02)	0.077***	(17.45)
Jan-11	-0.050***	(-10.68)	0.082***	(18.5)
Apr-11	-0.057***	(-11.97)	0.090***	(19.97)
Male	-0.0006	(-0.4)	-0.003*	(-1.98)
Age Band (18-44 as base category)				
45-64	0.038***	(11.11)	-0.030***	(-9.07)
65-74	0.065***	(19.48)	-0.051***	(-15.22)
75-84	0.090***	(27.62)	-0.073***	(-22.72)
85+	0.127***	(37.82)	-0.107***	(-32.14)
Ethnicity (White as base category)				
Mixed	-0.019**	(-3.16)	0.013*	(2.21)
Asian	-0.009	(-1.92)	0.006	(1.23)
Black	-0.008	(-0.68)	0.007	(0.79)
Other	-0.031***	(-4.03)	0.031***	(4.14)
Missing	0.006*	(2.21)	-0.009***	(-3.22)
Foundation Trust Status	0.007***	(3.71)	-0.005	(-1.94)
Financial management rating (Excellent as base category)				
Good	0.011***	(5.88)	-0.019***	(-10.02)
Fair/Poor	0.001	(0.31)	-0.0004	(-0.19)
Quality rating (Excellent as base category)				
Good	0.003*	(2.17)	-0.014***	(-9.48)
Fair/Poor	-0.027***	(-10.12)	0.027***	(10.69)
Income deprivation	-0.078***	(-5.06)	0.042**	(2.83)
Observations	**769,904**			
Pseudo R-squared	**0.3049**			

TABLE III

	Oct-08 to Sept-09				Oct-09 to Mar-10				Apr-10 to June-11			
	Exclusions		Success		Exclusions		Success		Exclusions		Success	
	ME	z-stat	ME	z-stat	ME	z-stat	ME	z-stat	ME	z-stat	ME	z-stat
Male	-0.002	(-0.69)	-0.001	(-0.31)	0.002	(0.49)	-0.004	(-1.43)	-0.001	(-0.55)	-0.003	(-1.51)
Age Band (18-44 as base category)												
45-64	0.029***	(4.43)	-0.015*	(-2.3)	0.049***	(6.85)	-0.042***	(-5.97)	0.038***	(8.07)	-0.034***	(-7.34)
65-74	0.054***	(8.22)	-0.029***	(-4.45)	0.077***	(10.98)	-0.065***	(-9.25)	0.067***	(14.34)	-0.057***	(-12.35)
75-84	0.075***	(11.74)	-0.048***	(-7.67)	0.098***	(14.39)	-0.087***	(-12.67)	0.095***	(21.02)	-0.082***	(-18.41)
85+	0.118***	(18)	-0.089***	(-13.63)	0.135***	(19.1)	-0.118***	(-16.7)	0.129***	(27.74)	-0.113***	(-24.63)
Ethnicity (White as base category)												
Mixed	-0.023*	(-2)	0.015	(1.41)	0.007	(0.56)	-0.006	(-0.45)	-0.028***	(-3.36)	0.020*	(2.47)
Asian	0.004	(0.39)	-0.002	(-0.26)	-0.00002	(0)	0.003	(0.26)	-0.018**	(-2.89)	0.011	(1.76)
Black	-0.022	(-0.96)	0.029	(1.34)	0.006	(0.26)	-0.013	(-0.59)	-0.006	(-0.45)	0.006	(0.53)
Other	-0.034*	(-2.01)	0.036*	(2.1)	-0.012	(-0.7)	0.011	(0.65)	-0.040***	(-4.11)	0.041***	(4.18)
Missing	-0.023***	(-4.57)	0.020***	(4.08)	0.012	(1.92)	-0.012	(-1.95)	0.018***	(4.14)	-0.022***	(-5.14)
Foundation Trust Status	-0.002	(-0.55)	0.019***	(5.24)	0.013**	(3.07)	-0.010*	(-2.44)	0.013***	(4.83)	-0.015***	(-6.09)
Financial management rating (Excellent as base category)												
Good	0.034***	(8.36)	-0.029***	(-7.48)	0.006	(1.42)	-0.016***	(-3.78)	0.0003	(0.13)	-0.013***	(-5.16)
Fair/Poor	0.008*	(2.09)	0.007	(1.8)	0.008	(1.87)	-0.011**	(-2.57)	-0.005	(-1.66)	-0.002	(-0.75)
Quality rating (Excellent as base category)												
Good	0.017***	(5.85)	-0.026***	(-9.73)	-0.018***	(-5.07)	0.003	(0.81)	0.002	(0.82)	-0.012***	(-5.78)
Fair/Poor	-0.016**	(-2.98)	0.028***	(5.53)	-0.043***	(-7.36)	0.041***	(7.43)	-0.029***	(-8.24)	0.024***	(6.83)
Income deprivation	-0.393***	(-12.12)	0.290***	(9.53)	0.009	(0.29)	-0.060	(-1.9)	0.059**	(2.82)	-0.055**	(-2.7)
Observations	**235,255**				**154,002**				**380,647**			
Pseudo R-squared	**0.2827**				**0.3140**				**0.3187**			