15tl7047
Articles
LWo
THELANCET-D-15-07047R3
S0140-6736(16)00276-2
Embargo: [add date when known]
This version saved: 15:37, 07-Apr-16

Long-term evidence for the effect of pay-for-performance in primary care on mortality in the UK: a population study

Andrew M Ryan, Sam Krinsky, Evangelos Kontopantelis, Tim Doran

Summary

Background Introduced in 2004, the UK's Quality and Outcomes Framework (QOF) is the world's largest primary care pay-for-performance programme. We tested whether the QOF was associated with reduced population mortality.

Methods We used population-level mortality statistics between 1994 and 2010 for the UK and other high-income countries that were not exposed to pay-for-performance. The primary outcome was age-adjusted and sex-adjusted mortality per 100 000 people for a composite outcome of chronic disorders that were targeted by the QOF. Secondary outcomes were age-adjusted and sex-adjusted mortality for ischaemic heart disease, cancer, and a composite of all non-targeted conditions. For each study outcome, we created a so-called synthetic UK as a weighted combination of comparison countries. We then estimated difference-in-differences models to test whether mortality fell more in the UK than in the synthetic UK after the QOF.

Findings Introduction of the QOF was not significantly associated with changes in population mortality for the composite outcome (–3·68 per 100 000 population [95% CI –8·16 to 0·80]; p=0·107), ischaemic heart disease (–2·21 per 100 000 [–6·86 to 2·44]; p=0·357), cancer (0·28 per 100 000 [–0·99 to 1·55]; p=0·679), or all non-targeted conditions (11·60 per 100 000 [–3·91 to 27·11]; p=0·143).

Interpretation Although we noted small mortality reductions for a composite outcome of targeted disorders, the QOF was not associated with significant changes in mortality. Our findings have implications for the probable effects of similar programmes on population health outcomes. The relation between incentives and mortality needs to be assessed in specific disease domains.

Funding None.

Department of Health Management and Policy, University of Michigan School of Public Health, Ann Arbor, MI, USA (A M Ryan PhD, S Krinsky MA); **Centre for Health Informatics, University of Manchester, Manchester, UK, and National Institute for Health Research, School for Primary Care Research, Radcliffe Observatory Quarter, Oxford, UK** (E Kontopantelis PhD); **and Department of Health Sciences, University of York, Heslington, York, UK** (T Doran MD)

Correspondence to:
Andrew Ryan, Department of Health Management and Policy, University of Michigan School of Public Health, Ann Arbor, MI 48109, USA
amryan@umich.edu

Introduction

Effective primary care can prevent illness and delay death by identifying and modifying risk factors,[1] diagnosing disease at an early stage,[2] and coordinating effective disease management.[3] In view of the low cost of diagnosis and treatment of disease in primary care compared with management of the complications of disease in acute settings, improvement of population health through enhanced primary care has tremendous potential to improve the value of health-care spending. Despite its importance, copious research suggests that high-quality primary care is underprovided.[4] This inadequate provision is due in part to health-care payment systems. In the USA, fee-for-service payment encourages high-intensity, procedure-based care rather than population-based patient management. In socialised health systems such as that in the UK, capitated payments encourage population-based approaches, but payments have traditionally been detached from quality of care.[5]

In response, many pay-for-performance initiatives have attempted to directly tie payment to quality of care. The UK's Quality and Outcomes Framework (QOF) is the world's largest pay-for-performance programme. It was introduced for all family practices in 2004, linking up to 25% of family practitioners' income to performance for more than 100 publicly reported quality indicators relating to management of chronic disease, organisation of care, and patient experience.[6] Notwithstanding an announcement that Scotland plans to eliminate the QOF in 2016, it currently remains in place across the whole of the UK. The magnitude of the financial incentives for clinical indicators varies substantially by disease area. For instance, in 2005, payments of up to £15 125 were available for the average family practice across 15 ischaemic heart disease and heart failure indicators, but only £1500 was available across two cancer indicators.

Research into the QOF suggests that the programme accelerated improvement for the incentivised indicators relative to preintervention trends in the 3 years after its implementation.[7] However, this improvement attenuated with time.[8] Conceptually, increased quality care through the QOF could reduce risk factors for acute events such as myocardial infarction and stroke, lowering associated mortality. Evidence from randomised controlled trials shows that 25 indicators in the QOF are associated with mortality reductions.[9] Attempts have been made to use this evidence to extrapolate quality improvement noted in the QOF to potential population mortality reductions.[9] However, evidence of efficacy from clinical trials might not translate into patient benefits in the real world of health-care delivery.[10] Additionally, measured quality improvement in the QOF could be partly driven by improved record keeping or manipulated performance statistics[11] and might not lead to improved outcomes. In

Research in context

Evidence before this study

We searched Embase and PubMed with the terms "quality", "outcomes framework", "incentive", "primary care", "mortality", "death*", "life expectancy", "primary health care", "value-based purchasing", and "pay-for-performance". We set no limit for language of publication and searched for articles published up to Dec 18, 2015. We reviewed all studies and trials, including observational studies and accompanying editorials. Investigators of many studies, predominantly observational, have examined the effect of incentive schemes on patient outcomes, but only one has addressed the effect of a national primary care pay-for-performance scheme on mortality. No previous cross-national studies have compared countries that implemented large-scale pay-for-performance with countries that did not.

Added value of this study

Investigators of previous studies have noted that financial incentive schemes can lead to slight improvements in incentivised aspects of care, but other aspects might be negatively affected, and the effects on patient outcomes are variable. We compared changes in mortality for disorders included in a major national primary care pay-for-performance programme in the UK with those in other high-income countries that did not introduce pay-for-performance. We noted no significant decrease in mortality in the UK after introduction of the incentive programme.

Implications of all the available evidence

Pay-for-performance might not be an effective method for improvement of population health. The costs and effectiveness of pay-for-performance programmes should be compared with other health system interventions to better understand than at present how resources can best be used to improve population outcomes.

this study, we compare changes in population mortality between the UK and countries that have not been exposed to large-scale pay-for-performance programmes to test whether the QOF is associated with broad improvements in population health.

See Online for appendix

Methods

Data sources

We used country-level, cause-specific mortality and population data from the WHO mortality database[12] between 1994 and 2010. For observations with missing population data (12% of country-years), we substituted population estimates from the US Census Bureau's international data base.[13] We used data for country characteristics from several sources, including the Penn World Tables,[14] the International Labour Organization key indicators of the labour market database,[15] the Standardized World Income Inequality Database,[16] and country health system classifications from Böhm and colleagues.[17]

Study design and population

We used a retrospective cohort design to test whether mortality for disease areas that were targeted by the QOF decreased more in the UK than in comparison countries in the 7 years after the QOF was introduced. Our comparison group consisted of countries that were classified in previous research as having a high-income epidemiological profile.[18] Of these countries, we excluded five because of extensive missing mortality data (Switzerland, Brunei, Andorra, Cyprus, and Malta), resulting in a set of 27 comparison countries.

Outcomes

Our primary outcome was age-adjusted and sex-adjusted mortality per 100 000 population for a composite outcome of disease areas that were targeted by the QOF from the beginning of the programme and had clinical evidence supporting the link between the incentivised indicators and mortality reduction.[9] These diseases were ischaemic heart disease, hypertension, stroke (including transient ischaemic attack), diabetes, chronic kidney disease, asthma, and chronic obstructive pulmonary disease (see appendix for detailed definitions of disease areas). Although chronic kidney disease was introduced into the QOF only in the third year of the programme, we included it in our composite outcome because of its association with diabetes and hypertension. In view of the interrelated nature of chronic disease, better care for a specific disease area (eg, diabetes) could contribute to reduced mortality for several causes of death (eg, diabetes, stroke, ischaemic heart disease, and chronic kidney disease). By considering composite mortality for several related disease areas, our outcome captures changes in mortality for causes of death that might not be directly related to improvement in care for the same disease area.

We assessed three secondary outcomes: age-adjusted and sex-adjusted mortality for ischaemic heart disease, cancer, and all causes of death that were not included in the primary outcome. We assessed ischaemic heart disease as a separate outcome because it is a leading cause of death in the UK and had the strongest financial incentives in the QOF (22% of total clinical incentives in the original programme).[6] We assessed cancer and all causes of death that were not included in the primary outcome as negative controls because we did not expect the QOF to substantially affect mortality for these causes of death. We assessed cancer separately because, by contrast with ischaemic heart disease, the cancer indicators in the QOF are not supported by clinical evidence for effect on outcomes and had weak financial incentives (<1% of total incentives).[9]

Causes of death that were not included in the primary outcome included some disease areas that were targeted by the programme, but these either lacked clinical evidence linking the incentivised indicator to mortality reduction (eg, cancer) or were only introduced in the third year of the programme (eg, atrial fibrillation).

Statistical analysis

To enable a direct comparison across countries, mortality for each country was standardised by age and sex to the 2004 UK population structure. We limited our calculation of population mortality to the population aged 74 years or younger in view of uncertainty about causes of death at older ages.[19] We standardised mortality to account for discontinuities associated with country-specific changes from the ICD9 to ICD10 classification systems for causes of death (appendix). Finally, we adjusted mortality to account for potential undercounting of deaths in the targeted disease areas due to use of ill-defined cause of death codes on death registrations (appendix).[20]

We used a difference-in-differences analysis to assess whether the QOF improved mortality in the UK to a greater extent than in comparison countries. The crucial assumption for difference-in-differences analysis is that the treatment and comparison groups have equivalent trends for the study outcome before the start of the intervention: the parallel trends assumption. This assumption implies that, without treatment, outcomes for the treatment and comparison groups would be expected to change at the same rate.[21]

However, for our study outcomes, the UK and comparison countries followed different trajectories before the QOF, violating the parallel trends assumption. To address this issue, for each outcome, we created so-called synthetic comparison groups, matching the UK with a weighted combination of non-exposed comparison countries based on preintervention mortality and other country characteristics.[22] Using a standard software routine (synth Stata module),[37] we derived weights to minimise the mean squared difference between the UK and comparison countries for age-standardised and sex-standardised mortality per 100 000 people in each year before the start of the QOF, gross domestic product per person, unemployment, the Gini coefficient of income inequality, and the type of health system.[18,23] We created separate synthetic comparison groups independently for each of the study outcomes. Preintervention levels and trends in mortality match almost exactly between the UK and the synthetic UK, validating our approach.

To test the effect of the QOF on mortality, we used linear regression to estimate the following equation for country j at time t:

	UK and high-income countries*		Synthetic controls†			
	UK	High-income countries‡	All QOF§	Ischaemic heart disease	Cancer	Non-QOF‖
Mortality per 100 000 people¶						
All QOF§	120·6 (29·5)	110·9 (35·4)	122·2 (37·1)	·· (··)	·· (··)	·· (··)
Ischaemic heart disease	79·6 (25·0)	60·9 (23·3)	·· (··)	81·0 (27·7)	·· (··)	·· (··)
Cancer	149·4 (13·8)	140·2 (22·2)	·· (··)	·· (··)	149·1 (28·7)	·· (··)
Non-QOF‖	276·7 (20·9)	280·9 (54·4)	·· (··)	·· (··)	·· (··)	272·9 (46·1)
GDP per person (US$; thousands)**	29·7 (3·8)	29·3 (11·5)	30·8 (12·0)	30·5 (10·3)	31·2 (11·1)	30·5 (11·0)
Unemployment	6·4 (1·6)	7·2 (3·6)	6·5 (3·4)	6·6 (2·9)	6·6 (3·3)	6·3 (4·3)
Gini††	34·7 (0·7)	31·5 (6·8)	34·2 (6·6)	34·3 (4·3)	33·3 (7·5)	33·4 (8·3)
Population (millions)	59·9 (1·2)	33·6 (57·4)	62·0 (101·4)	56·4 (109·4)	39·0 (87·5)	55·3 (81·5)
Countries	1	27	11	9	11	8
Countries by health system						
Etatist social health insurance	0	6 (22%)	2 (18%)	2 (22%)	1 (9%)	2 (25%)
National health insurance	0	5 (19%)	2 (18%)	2 (22%)	3 (27%)	1 (13%)
National health system	1 (100%)	7 (26%)	2 (18%)	2 (22%)	3 (27%)	1 (13%)
Private health insurance	0	1 (4%)	1 (9%)	1 (11%)	1 (9%)	1 (13%)
Social health insurance	0	3 (11%)	1 (9%)	0	0	1 (13%)
Unclassified	0	5 (19%)	3 (27%)	2 (22%)	3 (27%)	2 (25%)
Country-years	17	459	187	153	187	136
Imputed‡‡	1 (6%)	10 (2%)	4 (2%)	4 (3%)	4 (2%)	0

Data are mean (SD), n, or n (%). QOF=Quality and Outcomes Framework. GDP=Gross domestic product. *Simple (unweighted) mean. †Mean weighted by synthetic weights. ‡Excluding the UK. §Causes of death potentially sensitive to QOF incentives (ischaemic heart disease, stroke, asthma, chronic obstructive pulmonary disease, hypertension, diabetes, and chronic kidney disease). ‖All causes of death that were not included in the primary outcome. ¶For ages 0–74 years, standardised for age and sex to the 2004 UK population. **At constant 2005 US$. ††Gini coefficient of net income inequality. ‡‡Applicable to mortality only; no imputation done for covariates. Data taken from the WHO mortality database,[12] the US Census Bureau's international database,[13] the Penn World Tables 8.1,[14] the International Labour Organization key indicators of the labour market database,[15] the standardised world income inequality database,[16] and Böhm and colleagues.[17]

Table 1: **Characteristics of the UK and controls, 1994–2010**

$Mortality_{jt} = b_0 + (b_1 \times u_j) + (b_2 \times year_t) + (b_3 \times [UK_j \times post\text{-}QOF_t]) + e_{jt}$ where u_j is a vector of country fixed effects (including the UK), $year_t$ is a vector of year fixed effects, UK_j is a dummy variable showing that an observation is from the UK (showing the independent effect of the UK on mortality), post-QOF_t is a dummy variable that is equal to 1 after the QOF was implemented (2004–10), and e_{it} is the idiosyncratic error term. The coefficient b_0 is the intercept, b_1 captures the country-specific effects on mortality across the entire study period, b_2 captures secular year effects on mortality, and b_3 represents our estimate of the effect of the QOF. It shows the incremental difference in mortality between the UK and the comparison countries in the post-QOF period. We estimated this equation separately for each of our study outcomes. We weighted our observations by the weights that we used to construct the synthetic comparison groups. We assessed the significance of the effect of the QOF using the root mean-squared prediction error ratio test,[22] a non-parametric permutation test that is appropriate when the synthetic control method is used (appendix). We constructed 95% CIs that are implied by

	All QOF	Ischaemic heart disease	Cancer	Non-QOF
Argentina	0	0	0	0
Australia	0·158	0	0·059	0
Austria	0	0	0	0
Belgium	0	0	0	0
Canada	0	0	0	0
Chile	0·109	0	0·049	0·135
Denmark	0	0	0·197	0
Finland	0	0	0	0
France	0	0	0	0
Germany	0	0	0	0
Greece	0	0	0	0
Iceland	0	0·001	0·028	0
Ireland	0·248	0·358	0·257	0·058
Israel	0	0·046	0·012	0
Italy	0	0	0·032	0
Japan	0	0	0	0·14
Luxembourg	0·033	0	0	0·009
Netherlands	0·065	0·05	0	0·337
New Zealand	0	0·193	0	0
Norway	0	0	0	0
Portugal	0·029	0·009	0	0
Singapore	0·028	0·066	0·146	0·12
South Korea	0·115	0	0	0
Spain	0·04	0	0·025	0·114
Sweden	0	0	0	0
USA	0·166	0·184	0·109	0·087
Uruguay	0·007	0·094	0·084	0
Countries with non-zero weights	11	9	11	8

Values denote weights assigned to each country. QOF=Quality and Outcomes Framework.

Table 2: **Composition of synthetic UK**

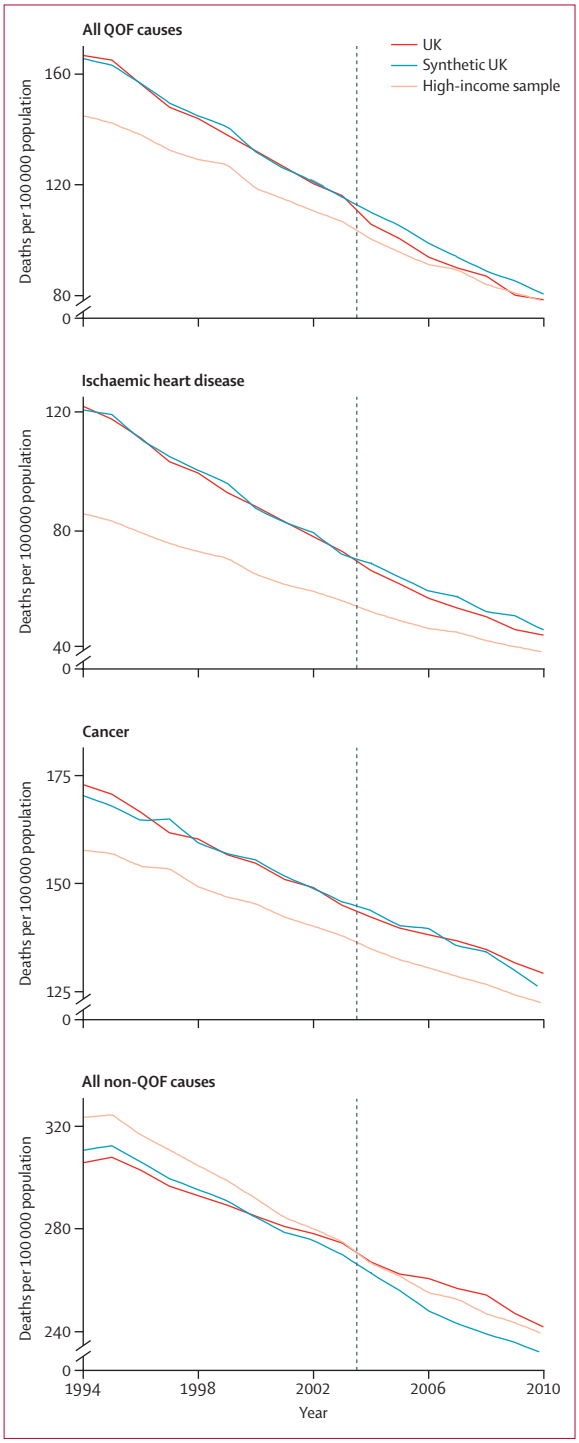

Figure: **Age-adjusted and sex-adjusted mortality in the UK and comparison countries**
The dashed vertical line denotes the period immediately preceding the start of the QOF. QOF=Quality and Outcomes Framework.

these permutation test p-values using the approach described by Altman and Bland.[24] For the sake of comparison, we also report significance from parametric *t* tests based on standard errors that were robust to country-level clustering. Our analysis used multiple imputation to account for missing data (see the appendix for detailed procedures).

To show how the results based on the synthetic control model differ from those from traditional difference-in-differences models, we estimated regression models using the full unweighted set of high-income countries. The model is identical to the equation listed above, with the exception that we included the time-varying matching variables as covariates in the regression. We did all analyses using Stata version 13.

Role of the funding source

There was no funding source for this study. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Before the start of the QOF, the UK had higher age-standardised and sex-standardised mortality per 100 000 population than did the combined set of comparison countries for the composite outcome (120·6 [SD 29·5] per 100 000 population *vs* 110·9 [35·4] per 100 000), ischaemic heart disease (79·6 [25·0] per 100 000 *vs* 60·9 [23·3] per 100 000), and cancer (149·4 [13·8] per 100 000 *vs* 140·2 [22·2] per 100 000), and slightly lower mortality for causes of death that were not related to the QOF (276·7 [20·9] per 100 000 *vs* 280·9 [54·4] per 100 000; table 1). However, the mortalities of the UK and the synthetic UK were nearly identical. Other country characteristics, consisting of gross domestic product per person, unemployment, and the Gini coefficient of income inequality, tended to be closer between the UK and the synthetic UK than between the UK and all comparison countries. 2% of country-year observations were imputed.

For each study outcome, the countries that made up the synthetic UK showed some similarities and some differences (table 2). South Korea, New Zealand, and Denmark each contribute heavily to a single outcome (South Korea to the composite outcome, New Zealand to ischaemic heart disease, and Denmark to cancer). After the QOF was introduced, age-standardised and sex-standardised mortality for the composite outcome, ischaemic heart disease, and cancer decreased at a similar rate for the UK and the synthetic UK (figure). By contrast, mortality for all disorders that were not targeted by the QOF decreased at a slower rate in the UK than in the synthetic UK.

Table 3 shows the difference-in-differences estimates of the association between the QOF and population mortality. All estimates are interpreted as a change in mortality per 100 000 population compared with mortality if the QOF had not been implemented. Estimates from the standard difference-in-differences specification show that the QOF was associated with lower mortality for the composite outcome (–12·81 per 100 000 people [95% CI –17·42 to –8·21]; p<0·0001), ischaemic heart disease (–16·38 per 100 000 [–20·32 to –12·44]; p<0·0001), and cancer (–2·64 per 100 000 [–5·13 to –0·15]; p=0·038) than that without the QOF, but significantly higher mortality for all non-targeted causes of death (11·33 per 100 000 [5·12–17·54]; p=0·0008). However, these estimates are biased by non-parallel trends between the UK and the entire set of high-income countries. Estimates from the synthetic comparison specification in which trends are parallel between the UK and the synthetic UK show that the QOF was not significantly associated with mortality for the composite outcome (–3·68 per 100 000 [–8·16 to 0·80]; p=0·107), ischaemic heart disease (–2·21 per 100 000 [–6·86 to 2·44]; p=0·357), cancer (0·28 per 100 000 [–0·99 to 1·55]; p=0·679), or all non-targeted disorders (11·60 per 100 000 [95% CI –3·91 to 27·11];

	All QOF		Ischaemic heart disease		Cancer		Non-QOF	
	DD	SC	DD	SC	DD	SC	DD	SC
Estimate	–12·81	–3·68	–16·38	–2·21	–2·64	0·28	11·33	11·60
p value*	..	0·107	..	0·357	..	0·679	..	0·143
95% CI implied by p value	..	–8·16 to 0·80	..	–6·86 to 2·44	..	–0·99 to 1·55	..	–3·91 to 27·11
p value from *t* test†	<0·0001	0·682	<0·0001	0·766	0·038	0·894	0·0008	0·083
95% CI from *t* test	–17·42 to –8·21	–23·23 to 15·88	–20·32 to –12·44	–18·86 to 14·44	–5·13 to –0·15	–4·38 to 4·94	5·12 to 17·54	–2·01 to 25·22
Difference in trend‡	–1·49	–0·12	–2·16	0·01	–0·75	–0·33	2·01	1·05
Countries	28§	12¶	28§	10¶	28§	12¶	28§	9¶
Country-years	476§	204¶	476§	170¶	476§	204¶	476§	153¶

Coefficients for model covariates are provided in the appendix. DD=difference-in-differences. SC=Difference-in-differences with synthetic control. *From permutation test. †Based on standard errors robust to country-level clustering. ‡Difference in preintervention linear trend between treatment and comparison group, 1994–2003. §Includes UK and all comparison countries. ¶Includes UK and all comparison countries with non-zero weights.

Table 3: **Estimates of the effect of the QOF on age-adjusted and sex-adjusted mortality per 100 000 population**

p=0·143). When the synthetic comparison group was used, statistical inference was consistent (p>0·05 for all outcomes) for the parametric and non-parametric significance tests. Extensive sensitivity analysis substantiated our main results (appendix).

Discussion

Our results show that introduction of the QOF in the UK was not significantly associated with changes in population mortality for disease areas that were targeted by the programme. We recorded that the QOF was also not significantly associated with changes in mortality for disease areas that were not targeted by the programme. Extensive research into pay-for-performance programmes has yet to show clear patient benefits.[25–27] The QOF is a unique programme, covering almost the entire UK population through an enormous investment: about £5·86 billion (US$9 billion) was invested in incentive payments during the first 7 years of the QOF, with additional billions invested in programme administration and information technology support. Although research suggests that the QOF improved quality for incentivised activities in its early years, albeit at the possible expense of quality for some non-incentivised aspects of care,[7] this improvement seemed to attenuate with time.[8] Findings from some studies suggest that the QOF led to better intermediate outcomes[28] and fewer emergency hospital admissions for some disorders, including ischaemic heart disease, than without the QOF.[29] As a result, the QOF might reasonably be expected to have reduced population mortality.[30] Extrapolations on the basis of improvements for the incentivised indicators estimated that the QOF reduced mortality by 11 deaths per 100 000 people in 2004.[9] Yet, evidence suggests that local variation in quality performance in the QOF was not associated with mortality within England.[31] By comparing mortality between the UK and comparable countries that were not exposed to national-scale pay-for-performance, our study provides the first cross-national evidence for the effects of pay-for-performance on population health.

The apparent failure of such a large and sustained programme to reduce mortality suggests that faults might exist in the general approach of use of financial incentives to improve population outcomes or in the specific design of the QOF. Possible explanations include reported improvements because of improved recording or gaming, an absence of a direct effect of incentivised aspects of care on mortality, insufficiently large or mistargeted financial incentives, setting of suboptimum clinical targets, and insufficiently challenging achievement thresholds.[30] Alternative programme designs might have resulted in greater reductions in population mortality than with the QOF design. The effects of improved primary care on mortality could possibly be slight in comparison with other factors, including socioeconomic determinants, and could not be noted in population-level data. The QOF could possibly have improved non-fatal outcomes, but we were not able to note these outcomes in our study.

Our study has various limitations. Variation in coding practices and classification systems can introduce random error into international comparative analysis of mortality statistics.[20] Although we accounted for ill-defined codes in vital registration systems and statistically adjusted discontinuous mortality that resulted from countries switching from the ICD9 to ICD10 coding system, country-specific variation in cause of death coding will remain. However, unless idiosyncrasies in mortality data and coding practices aligned with the start of the QOF, they would not be expected to bias our results.

The large difference in estimates between our traditional difference-in-differences and synthetic control specifications shows the challenges associated with identification of the effects of the QOF in the context of steep mortality decreases in the UK before the programme's introduction. Additionally, the QOF was not the only reason why mortality would have changed in our study period: various other changes in population risk factors and medical care occurred in both the UK and the comparison countries.[32] Although no other financial incentive programmes were implemented that came close to matching the scale of the QOF, various programmes were implemented in countries in the study sample. Yet, our preferred synthetic control approach, which succeeded in creating comparison groups that closely matched the UK, accounted for differences in background improvements, helping to isolate the effect of the QOF. Although we cannot compare the effect of the QOF with no intervention in the comparison groups, we can say that the QOF did not seem to generate incremental improvements in mortality when compared with the general improvements—perhaps encouraged by payment policy or other reforms—that were noted in comparable countries. By analogy with terminology from clinical trials, we do not interpret our results as a comparison between treatment and no treatment study arms, but rather between treatment and usual care.

Our analysis might have been underpowered to detect effects of the QOF on mortality. The substantial variance of population mortality estimates could have decreased our ability to detect significant effects. Although use of non-parametric permutation tests for statistical inference was necessary because of violations of standard statistical assumptions,[22] these tests are conservative. Our permutation tests could therefore possibly have failed to reject the null hypothesis when the QOF did in fact reduce mortality. However, even the smaller variance estimates from the standard difference-in-differences models would not have affected our study inferences, and our results were not sensitive to our use of parametric or non-parametric methods for statistical inference.

Finally, a longer study period than the one we used could possibly be needed for risk factor reduction and

disease management in primary care to be manifested in population mortality. However, evidence from trials[33] and natural experiments[34] suggests that pharmacological interventions and improvements in risk factors result in striking reductions in mortality for disorders such as ischaemic heart disease during short timeframes.[35] Any effect on mortality of a large-scale intervention such as the QOF would therefore probably become evident within the first 7 years of implementation.

Programmes that use payments to physicians and health-care institutions to incentivise high-quality care have a strong foothold in several countries and are especially well developed in the UK and USA. Our research raises questions about whether pay-for-performance in other settings is a viable method to improve population health. For example, as a result of the Patient Protection and Affordable Care Act in the USA, pay-for-performance programmes have proliferated through Medicare.[36] Although design flaws might have undermined the effect of the QOF,[30] if a programme with the size and scope of the QOF was not associated with statistically greater reductions in population mortality than without the QOF, less ambitious programmes than the QOF—such as those in the USA—might be even less likely to reduce mortality. Pay-for-performance programmes will continue to develop, hopefully in ways that incentivise higher value care than without them.[30] The costs and effectiveness of pay-for-performance with other health system interventions should be explored to better understand how resources can best be used to improve population health.

Contributors

All authors conceived and designed the study, analysed and interpreted data, and drafted and critically revised the report. SK acquired data.

Declaration of interests

The authors declare no competing interests.

Acknowledgments

We thank Amy VanderZanden of the Institute for Health Metrics and Evaluation for providing detailed tables supporting the cause of death groupings used in this report and Kristin Maurer for providing research assistance for this study.

References

1 Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. *Mil Quart* 2005; **83:** 457–502.

2 Bellamy D, Smith J. Role of primary care in early diagnosis and effective management of COPD. *Int J Clin Pract* 2007; **61:** 1380–89.

3 Flanagan DE, Cox P, Paine D, Davies J, Armitage M. Secondary prevention of coronary heart disease in primary care: a healthy heart initiative. *QJM* 1999; **92:** 245–50.

4 McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *N Eng J Med* 2003; **348:** 2635–45.

5 Leatherman S, Berwick D, Iles D, et al. The business case for quality: case studies and an analysis. *Health Aff (Millwood)* 2003; **22:** 17–30.

6 Roland M. Linking physicians' pay to the quality of care—a major experiment in the United Kingdom. *N Eng J Med* 2004; **351:** 1448–54.

7 Doran T, Kontopantelis E, Valderas JM, et al. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. *BMJ* 2011; **342:** d3590.

8 Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of pay for performance on the quality of primary care in England. *N Engl J Med* 2009; **361:** 368–78.

9 Fleetcroft R, Parekh-Bhurke S, Howe A, Cookson R, Swift L, Steel N. The UK pay-for-performance programme in primary care: estimation of population mortality reduction. *Br J Gen Pract* 2010; **60:** e345–52.

10 Singal AG, Higgins PD, Waljee AK. A primer on effectiveness and efficacy trials. *Clin Transl Gastroenterol* 2014; **5:** e45.

11 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:** 274–84.

12 WHO. WHO mortality database. http://www.who.int/healthinfo/statistics/mortality_rawdata/en/ (accessed April 9, 2015).

13 US Census Bureau. International data base. July, 2015. https://www.census.gov/population/international/data/idb/informationGateway.php (accessed April 9, 2015).

14 Feenstra RC, Inklaar R, Timmer MP. The next generation of the Penn World Table. *Am Econ Rev* 2015; **105:** 3150–82.

15 International Labour Organization. Key indicators of the labour market (KILM) 2015. http://www.ilo.org/global/statistics-and-databases/research-and-databases/kilm/lang--en/index.htm (accessed May 28, 2015).

16 Solt F. 2009. The Standardized World Income Inequality Database, Harvard Dataverse, version 14. https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/11992 (accessed May 28, 2015).

17 Böhm K, Schmid A, Götze R, Landwehr C, Rothgang H. Classifying OECD healthcare systems: a deductive approach. *TransState Working Papers* 2012; **165:** 1–83.

18 Murray CJ, Ezzati M, Flaxman AD, et al. GBD 2010: design, definitions, and metrics. *Lancet* 2012; **380:** 2063–66.

19 Nolte E, McKee CM. Measuring the health of nations: updating an earlier analysis. *Health Aff (Millwood)* 2008; **27:** 58–71.

20 Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012; **380:** 2095–128.

21 Ryan A, Dimick J. Methods for evaluating changes in health care policy: the difference-in-differences approach. *JAMA* 2014; **312:** 2401–402.

22 Abadie A, Diamond A, Hainmueller J. Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *J Am Stat Assoc* 2010; **105:** 493–505.

23 Gini C. Concentration and dependency ratios. *Rivista di Politica Economica* 1997; **87:** 769–89.

24 Altman DG, Bland JM. How to obtain the confidence interval from a p value. *BMJ* 2011; **343:** d2090.

25 Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med* 2006; **145:** 265–72.

26 Flodgren G, Eccles MP, Shepperd S, Scott A, Parmelli E, Beyer FR. An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes. *Cochrane Database Syst Rev* 2011; **7:** CD009255.

27 Houle SK, McAlister FA, Jackevicius CA, Chuck AW, Tsuyuki RT. Does performance-based remuneration for individual health care practitioners affect patient care?: a systematic review. *Ann Intern Med* 2012; **157:** 889–99.

28 Dusheiko M, Gravelle H, Martin S, Rice N, Smith PC. Does better disease management in primary care reduce hospital costs? Evidence from English primary care. *J Health Econ* 2011; **30:** 919–32.

29 Harrison M, 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.

30 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.

31 Kontopantelis E, Springate DA, Ashworth M, Webb RT, Buchan IE, Doran T. Investigating the relationship between quality of primary care and premature mortality in England: a spatial whole-population study. *BMJ* 2015; **350:** h904.

32 Ford ES, Capewell S. Proportion of the decline in cardiovascular mortality disease due to prevention versus treatment: public health versus clinical care. *Ann Rev Public Health* 2011; **32:** 5–22.

33 LaRosa J, He J, Vupputuri S. Effect of statins on risk of coronary disease: a meta-analysis of randomized controlled trials. *JAMA* 1999; **282:** 2340–46.

34 Zatonski W, Campos H, Willett W. Rapid declines in coronary heart disease mortality in eastern Europe are associated with increased consumption of oils rich in alpha-linolenic acid. *Eur J Epidemiol* 2008; **23:** 3–10.

35 Capewell S, O'Flaherty M. Can dietary changes rapidly decrease cardiovascular mortality rates? *Eur Heart J* 2011; **32:** 187–89.

36 Ryan AM, Damberg CL. What can the past of pay-for-performance tell us about the future of value-based purchasing in Medicare? *Healthcare* 2013; **1:** 42–49.

37 Abadie A, Diamond A, Hainmueller J. SYNTH: Stata module to implement synthetic control methods for comparative case studies. http://econpapers.repec.org/software/bocbocode/s457334.htm (accessed April 6, 2016).