Non-experimental evaluation of pay-for-performance: methods and findings

Matt Sutton
Professor of Health Economics
matt.sutton@manchester.ac.uk

Acknowledgements

- Co-investigators:
  - Soren Kristensen, Thomas Allen, Eleonora Fichera (Manchester)
  - Ruth McDonald (Nottingham), Martin Roland (Cambridge), Helen Lester (Birmingham), Ruth Boaden (Manchester)

- Funded by Department of Health Policy Research Programme and the NIHR Health Services and Delivery Research Programme (project number 08/1809/250)

- Disclaimer: The views and opinions expressed do not necessarily reflect those of the HS&DR Programme, NIHR, NHS or the Department of Health

Pay-For-Performance Schemes

- International trend towards linking provider revenue to achievement of quality indicators, called P4P
- P4P is being adopted despite little evidence
- Little clarity and consistency on what P4P involves

Design aspects of P4P schemes

- Coverage: patient groups and health conditions
- Payment for processes or outcomes
- Personal income or re-investment in patient care
- Size of the incentives
- Bonuses or penalties
- Payment schedule: targets, thresholds, linear schedules
- Linked to: absolute or relative performance
- Monitoring: self-report; independent data
- Frequency of assessment, payment and revision
- Supporting levers: feedback, shared learning, public reporting, choice

P4P in England

- The NHS in England has introduced various forms of pay-for-performance over the last nine years:
  - *Quality and Outcomes Framework (QOF),* from 2004/5
  - Advancing *Quality* (North West only, from October 2008)
  - *Commissioning for Quality and Innovation* (CQUIN), from 2009/10
  - *Best Practice Tariffs,* from 2010/11
  - Non-Payment Policies (e.g. hospital readmissions), from 2011/12

Types of P4P adopted in England

- *Quality and Outcomes Framework*
  - 150 new performance indicators for all 10,000 general practices
- *Advancing Quality*
  - regional quality tournament for 24 hospitals
- *Commissioning for Quality and Innovation Framework*
  - mandation of quality elements in local contract negotiations
- *Best Practice Tariffs*
  - quality-related price adjustments in national hospital payment system
- Non-payment policies
  - withholds of payments for specific examples of below-standard quality

Schemes

Aspect	QOF	AQ	CQUIN	BPT	NPP
Introduction	2004	2008	2009	2010	2011
Participants	Practices	Hospitals	Hospitals	Hospitals	Hospitals
Reward	Income	Budget	Revenue	Revenue	Revenue
Size	25% of total	4% of tariff	1.5% of total	5-34% tariff	100% tariff
Bonus	Bonuses	Bonuses	Penalties	Mixed	Penalties
Pay schedule	Linear	Lump sum	Target	Per-patient	Per-patient
Measurement	Absolute	Tournament	Absolute	Absolute	Absolute
Monitoring	Self-report	Self-report	Local agree	Admin. data	Admin. data
Frequency	Annual	Quarterly	Annual	Continuous	Continuous
Support	Weak public	Various	Weak public	None	None

- Two schemes each cost over £1bn p.a. (~1% of NHS)

Evaluation problems

- Counterfactuals / controls
  - Understanding the process of participation / assignment
  - Allowing for externalities / spillovers / contamination
- Availability and reliability of data / gaming
- Anticipation effects
- Expected duration

P4P in England

- The NHS in England has introduced various forms of pay-for-performance over the last nine years:
  - *Quality and Outcomes Framework (QOF),* from 2004/5
  - Advancing *Quality* (North West only, from October 2008)
  - *Best Practice Tariffs,* from 2010/11
  - *Commissioning for Quality and Innovation* (CQUIN), from 2009/10
  - Non-Payment Policies (e.g. hospital readmissions), from 2011/12

Advancing Quality

- First hospital P4P scheme to be introduced in the UK
- Based on US Hospital Quality Incentive Demonstration
- Adopted by all 24 NHS Acute Trusts in 1 of 10 regions

- Five patient groups:
  - pneumonia, CABG, AMI, heart failure, hip/knee replacements
- Performance on 28 quality indicators
- Simple tournament
  - top 6 Trusts received a 4% bonus on their tariff payments
  - next 6 Trusts received a 2% bonus on their tariff payments

Estimation of effect on mortality

- National hospital episodes data
- Deaths within 30 days of admission
- Risk-adjustment using age and sex, primary diagnosis, 31 co-existing conditions, admission type, residence type
- For patients admitted for:
  - 3 incentivised conditions (AMI, heart failure, pneumonia)
  - 6 non-incentivised, reference conditions
- Periods: 18mths before and after introduction
- 24 NW Trusts compared to 132 Trusts in rest of England
- 134,435 patients in NW versus 722,139 patients in RoE

Changes in unadjusted mortality

	North West			Rest of England		
	Before	After	**Change**	Before	After	**Change**
AMI	12.4	11.0	**-1.4**	11.0	10.7	**-0.3**
Heart failure	17.9	16.6	**-1.3**	16.6	16.1	**-0.6**
Pneumonia	28.0	25.9	**-2.2**	27.2	26.3	**-0.9**
Reference conditions	13.3	13.0	-0.3	11.7	11.0	-0.7

Mortality measured in percentage points.

Difference-in-differences

Health condition	Between-Region Difference in Differences	Triple Difference
Reference conditions	0.3 (-0.4 to 1.1)	-
Incentivised conditions	**-0.9** (-1.4 to -0.4)	**-1.3** (-2.1 to -0.4)
AMI	**-0.3** (-1.0 to 0.4)	**-0.6** (-1.7 to 0.4)
Heart failure	**-0.3** (-1.2 to 0.6)	**-0.6** (-1.8 to 0.6)
Pneumonia	**-1.6** (-2.4 to -0.8)	**-1.9** (-3.0 to -0.9)

Risk-adjusted mortality measured in percentage points (95% CI)

Further analyses

- No significant differences in patient discharge to institutions
- Trends in mortality were similar in the North West to the rest of England before introduction of the scheme
- Similar results when exclude the south of England
- Results unaffected by controlling for baseline mortality and changes in patient volumes
- Accompanying qualitative evaluation suggests it was regional collaboration and broad range of quality improvement activity that explains the impact

Best Practice Tariffs

- 'Best Practice Tariffs' introduced into the activity-based financing system for four types of intervention in 2010/11
- Three models:
  - "Incentivise daycase" model, applied to removal of gall bladder
  - "Paying for best practice" model, applied to stroke & hip fracture
  - "Pathway" model, applied to cataracts
- Quality bonuses, revenue caps and withholds

Prices for gall bladder removal

Year	Without complications	With complications
2007/08	£1,777 (<70years)	£2,328 (>69years)
2008/09	£1,837 (<70years)	£2,371 (>69years)
2009/10	£1,365	£2,131
2010/11	£1,694 (daycase) £1,369 (inpatient)	£2,164

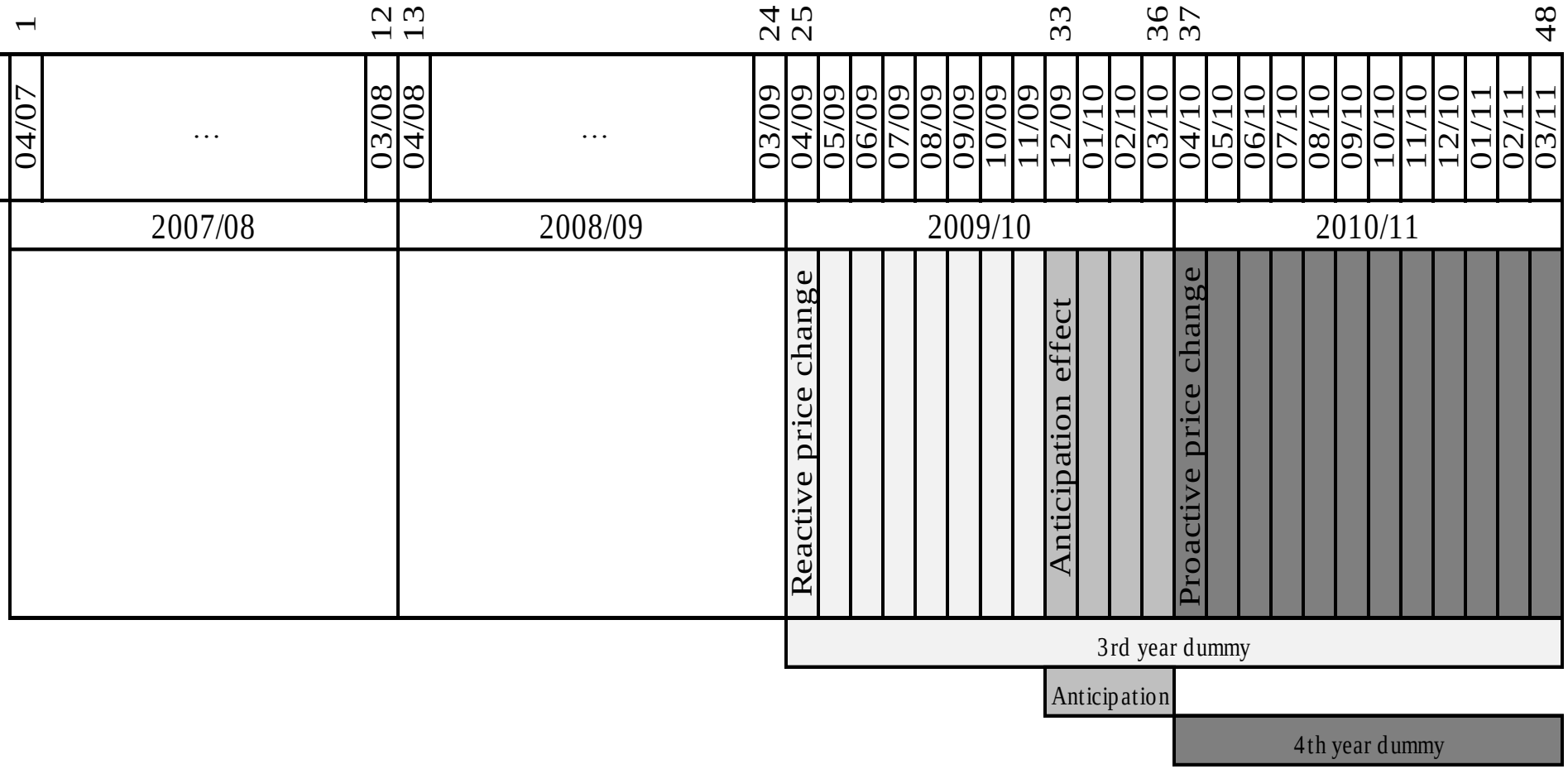
Methods

- National hospital episode records - before (2007/8–2009/10) and after (2010/11) introduction

- Gall bladder BPT introduced nationwide

- We used patients undergoing a comparable basket of procedures as controls:
  1. List of procedures recommended for daycase treatment
  2. Procedures in similar range of recommended % as gall bladders
  3. Tests of pre-intervention trends for each outcome of interest
  4. Overall basket with similar Laspeyres price index

Timeline and definition of DiDs



Indicators examined

- Proportion performed as daycase
- Proportion planned as a daycase
- Average age of patients treated
- Proportion of male patients
- Average number of co-morbidities
- Proportion of reversions to open surgery
- Readmission rate
- Proportion of stays over 1 day
- Death rate
- Volume of operations
- Average waiting time

Estimated effects of daycase BPT

	(1)	(4)	(6)
	Daycase prop.	Readmissions	Waiting time
Gall bladder	-0.149***	0.005***	20.60***
	(-48.60)	(10.70)	(41.42)
DiD 3rd year	0.005	0.001	1.96*
	(0.83)	(0.64)	(2.29)
Anticipation	0.028**	-0.001	4.76**
	(2.78)	(-0.42)	(3.26)
DiD 4th year	0.063***	-0.000	5.80***
	(8.33)	(-0.05)	(5.89)
Constant	0.320***	0.009***	34.26***
	(58.60)	(11.73)	(36.72)
Observations	16214	15875	12029
R^2	0.638	0.058	0.677

“Paying for best practice” model

- Additional payments for meeting quality standards
- Decision on whether stroke and hip fracture BPTs would operate was negotiated locally
- We surveyed providers on whether they could receive BPTs
- Activities are only measured in clinical audits

- We analysed outcome indicators that could be derived consistently before and after BPTs for all patients:
  - mortality within 30 days
  - readmission within 30 days
  - return to usual place of residence within 56 days
- Difference-in-differences on eligible versus ineligible

Findings

- Introduction of stroke BPT had no impact

- Hip fracture BPT associated with:
  - 4.0% point larger increase in surgery within 48 hours
  - 0.7% point larger decrease in the mortality rate
  - 2.1% point larger increase in returns home within 56 days

- Why?
  - different tariff structures (only paid for hip fracture if all criteria met)?
  - differences in underlying quality trends?
  - other supporting or competing initiatives (e.g. audits)?

CQUIN contract negotiations

- Two national indicators (0.3% of revenue):
  - Patients screened for VTE risk
  - Patient experience
- Option for region-wide indicators - mandated or ‘suggested’
- Locally-negotiated element:
  - Target levels of quality required
  - Additional topics and indicators
  - Revenue weights (to sum to 1.5% in total)

Variations in local content of CQUIN

Sector	Content	Mean	St.Dev.	Min.	Median	Max.
Acute care	Indicators	18.4	9.4	3	16	52
	Topics	11.3	4.4	2	11	25

- In CQUIN schemes for 151 hospitals
  - there were 92 distinct topics
  - there were 1,546 unique indicators
  - 70% had no baseline measures of performance

Impact of inclusion of topic on outcomes

- Analysed nine indicators (covering patient safety, elective surgery outcomes, emergency readmissions, and returns to usual residence)
- Compared *changes* in outcomes in hospitals with a particular topic in their CQUIN scheme with *changes* in outcomes in hospitals that did not
- Inclusion of patient safety and hip fracture topics was associated with significantly better outcomes before CQUIN
- Only one instance of association with better outcomes

Concluding remarks

- One English region imported a P4P scheme from the US
  - translated it to NHS context
  - was associated with substantial reduction in mortality
  - a quality improvement programme supported by financial incentives
- Findings from evaluations of other P4P schemes in England are consistent with the inconsistent evidence-base
- This emphasises that policy-makers should:
  - pilot P4P schemes
  - design them carefully
  - support them with complementary mechanisms
  - evaluate them properly, to build up an evidence-base
  - be prepared for disappointing evaluation results

Non-experimental evaluation of pay-for-performance: methods and findings

Matt Sutton
Professor of Health Economics
matt.sutton@manchester.ac.uk