



# Paying for Performance (P4P) in health care: time for caution!

---

By





Alan Maynard

email:[akm3@york.ac.uk](mailto:akm3@york.ac.uk)



# Outline





---

-  Why do we need P4P?
-  What is P4P?
-  Do incentives improve performance?
-  Overview: proceed with care!



# Why do we need P4P?

---

-  The evidence base for health care is poor?
-  There are large variations in clinical practice i.e. physicians give very different care to patients with similar characteristics and similar health needs
-  What is your medical error rate? Ten per cent or more?
-  Why is health care reform focused on organisational structures and care processes rather than on patient outcomes: do we make patients “better”?

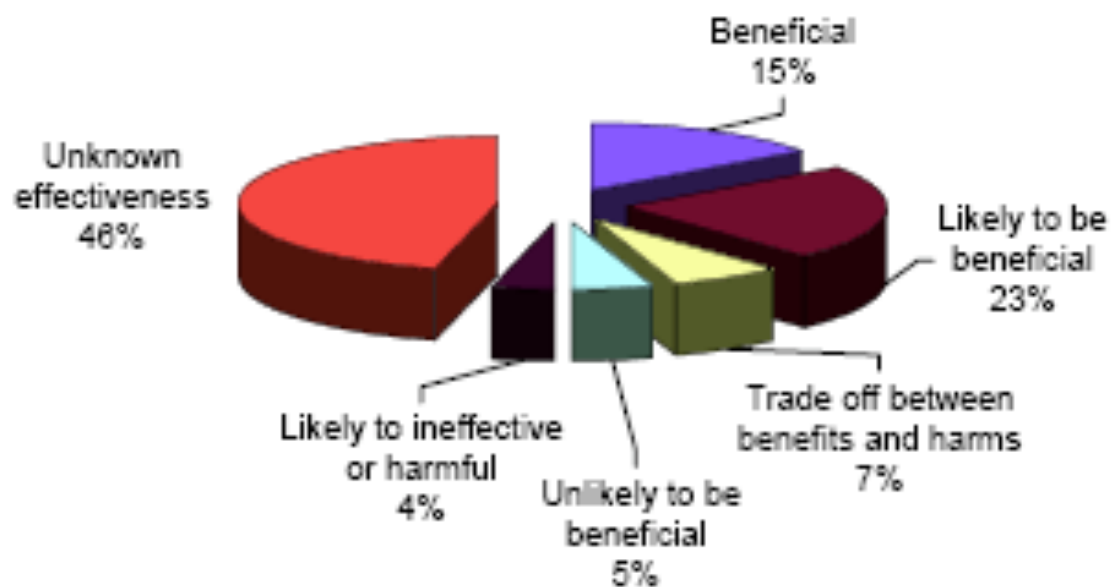
# Does health care improve health?



---

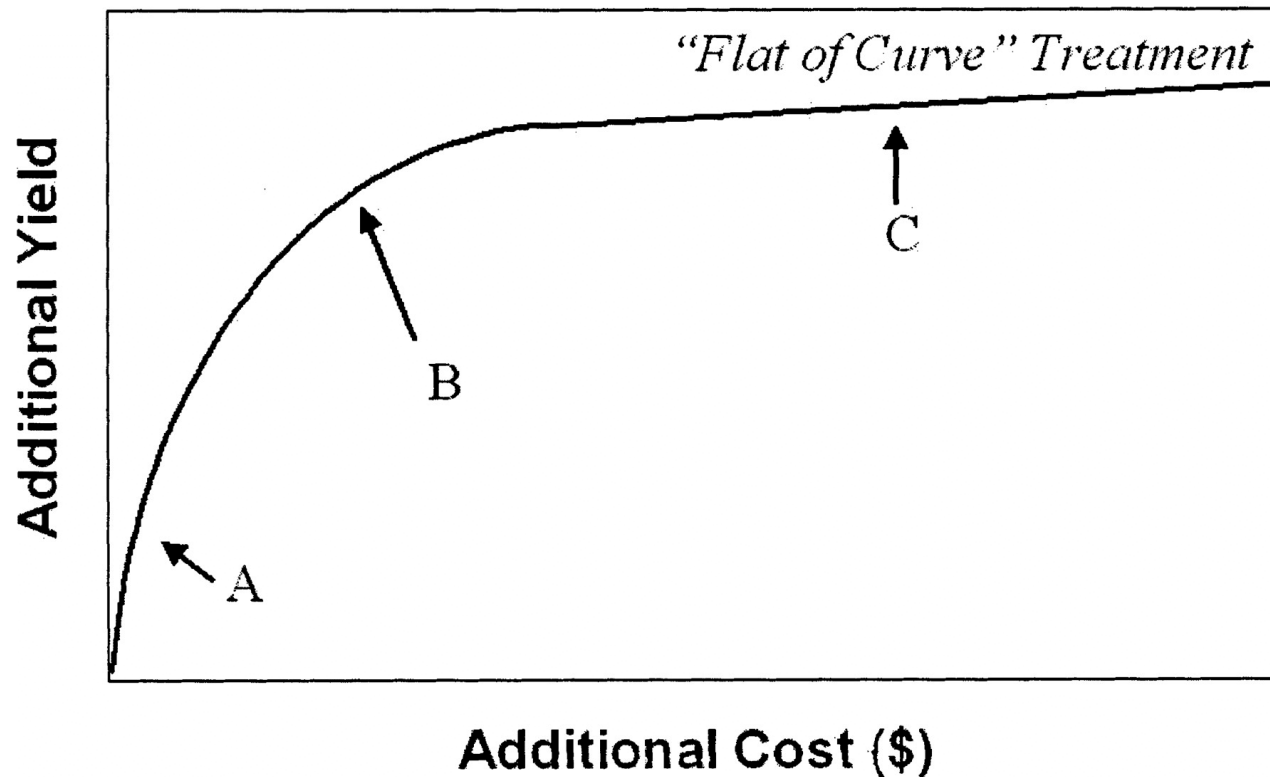
- “ I once asked a worker at a crematorium, who had a curiously contented look on his face, what he found so satisfying about his work. He replied that what fascinated him was the way in which so much went in and so little came out. I thought of advising him to get a job in the NHS, it might increase his job satisfaction, but decided against it. He probably gets his kicks from the visual demonstration of the gap between input and output. A more statistical demonstration might not have worked so well”
- Archie Cochrane, “Effectiveness and Efficiency” (1972)

Figure 1: Uncertainty about clinical effectiveness



Source: BMJ Publishing Group 2005<sup>13</sup>

# 'Flat of the Curve' Medicine?



Mark & Hlatky 2002, Fuchs 2004



# The causes of uncertainty about clinical effectiveness

---

- Not so much inadequate funding of R&D and clinical trials, more that the quality of research is poor.
- The problems of designing and reporting clinical trials e.g. the problem of “surrogate” end points, poor outcome measurement and biased reporting.
- What is the comparator?
- What patient groups are included in the trial?
- How long should you run the trial? Vioxx case



# Clinical practice variations

---

- With the evidence base so incomplete, variation in practice is perhaps unsurprising, given the “autonomy” of practitioners in health and social care.
- The variations literature is concerned with differences in activity and practice that are large and usually unexplained
- Of course to vary is normal, the issue is how much variation is efficient and acceptable?.
- Evidence of variations has existed for decades but no health care system, public or private, has managed them efficiently





# US variations in medical practice: the work of Wennberg

---

- The work of Wennberg: e.g his comparison of Medicare beneficiaries in Boston and New Haven where the demographic characteristics of the populations are similar. Adjusted rates of discharge, readmission, length of stay and reimbursement varied by 47, 29, 15 and 79%. Mortality rates were identical whilst expenditure in Boston nearly double that of New Haven (NEJM 1989 and Lancet 1987)



# Practice variations now

---

- US Medicare per capita spending in 2000 was \$10,550 per enrollee in Manhattan and \$4823 in Portland, Oregon. Differences are due to volume effects rather than illness differences, socio-economic status or price of services.
- “Residents in high spending regions received 60% more care but did not have lower mortality rates, better functional status or higher satisfaction” Fisher et al Annals in Internal Medicine(2003). Potential savings of 30% if high spenders reduce expenditure and provide the safe practices of conservative treatment regions? Fisher in NEJM, October, 2003



## Reimbursement rate for non-capitated Medicare per enrollee, 2006

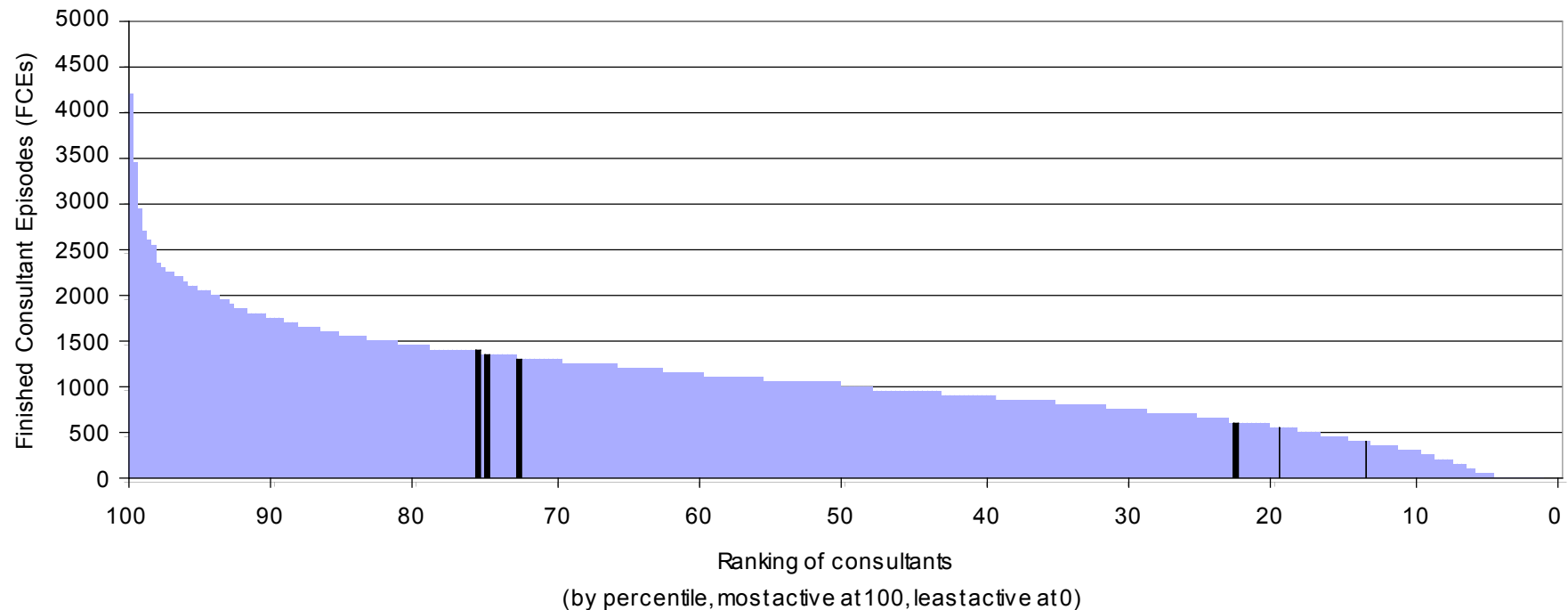
Hospital referral region	Medicare spending 2006 (\$)	Spending growth 1992-2006 (\$)	Annual growth rate 1992-2006 (%)
Manhattan NY	12114	4979	3.9
Los Angeles	10810	3707	3.0
Philadelphia	9665	3495	3.2
Boston	9526	3204	3.0
Nashville	8355	3048	3.3
Phoenix AZ	7890	2748	3.1
Atlanta	7363	2004	2.3
Seattle	7218	2379	2.9
Minneapolis	6705	2967	4.3

*Source: Fisher et al, NEJM, February 26<sup>th</sup>, 2009, page 851*

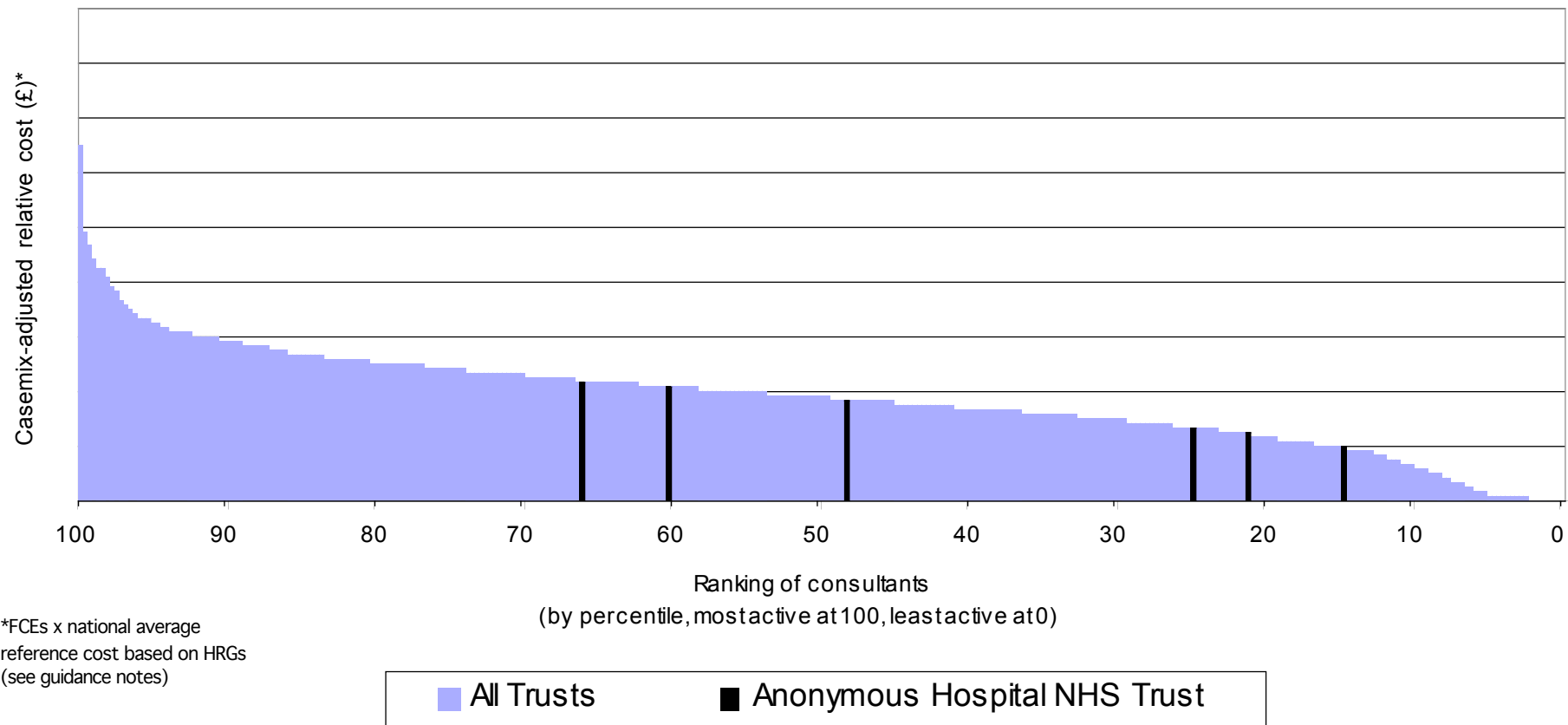
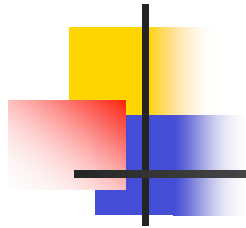
# Variations in English consultant activity

- Hospital Episode Statistics record activity by consultant in England (Wales has similar data)
- Rates of activity declining and very variable. What are the causes of declining activity? What are the causes of variation in activity?
- 📁 Data validity
- 📄 Public sector inefficiency
- 📄 Private sector problems

# Variation in activity in general surgery: FCEs



# Variation in activity in general surgery: HRG/cost adjusted



# Practice variations: why do they persist?






- “the amount and cost of hospital treatment in a community have more to do with the number of physicians there, their medical specialties and the procedures they prefer than the health of residents” Wennberg and Gittelsohn(1973 in the journal Science)
- Why are policy makers nationally internationally so slow at changing practices to reduce clinical practice variations and delivering care to patients in relation to the evidence base?
- Should we benchmark “best practice” and “induce” compliance across the NHS? E.g. mandate accurate PbR data?
- The issue of authority: neither the medical profession nor health care managers have the authority to measure and control practice variations



# Patient safety 1

---




- Measuring error rates is difficult e.g. self reporting versus patient case records?
- The evidence base is incomplete:
  -  USA 3-5% of hospital admissions (Institute of Medicine 1999)
  -  UK :two retrospective English studies (Vincent et al, BMJ 2001, and Sari et al (2006)) :10%
  -  Australia: 16% (=10% if US criteria used)





# Patient safety 2

---

- US rates of 3-5% means that
  -  Medical errors kill 44,000-98,000 Americans each year
  -  Errors kill more Americans than motor vehicle accidents (43,458), or breast cancer (42,297) or AIDS (16,516)
  -  Medication errors alone kill nearly three times more Americans than 9/11



# Medical errors: "never" events

---

- Medication errors: wrong dose/wrong drug
- Wrong site surgery
- Pressure sores
- Infection: C.Diff, MRSA, central line infections, catheter associated urinary tract infections.
- Falls and trips.
- DVT prophylaxis
- Hand hygiene.....

# Lessons from the 19<sup>th</sup> century

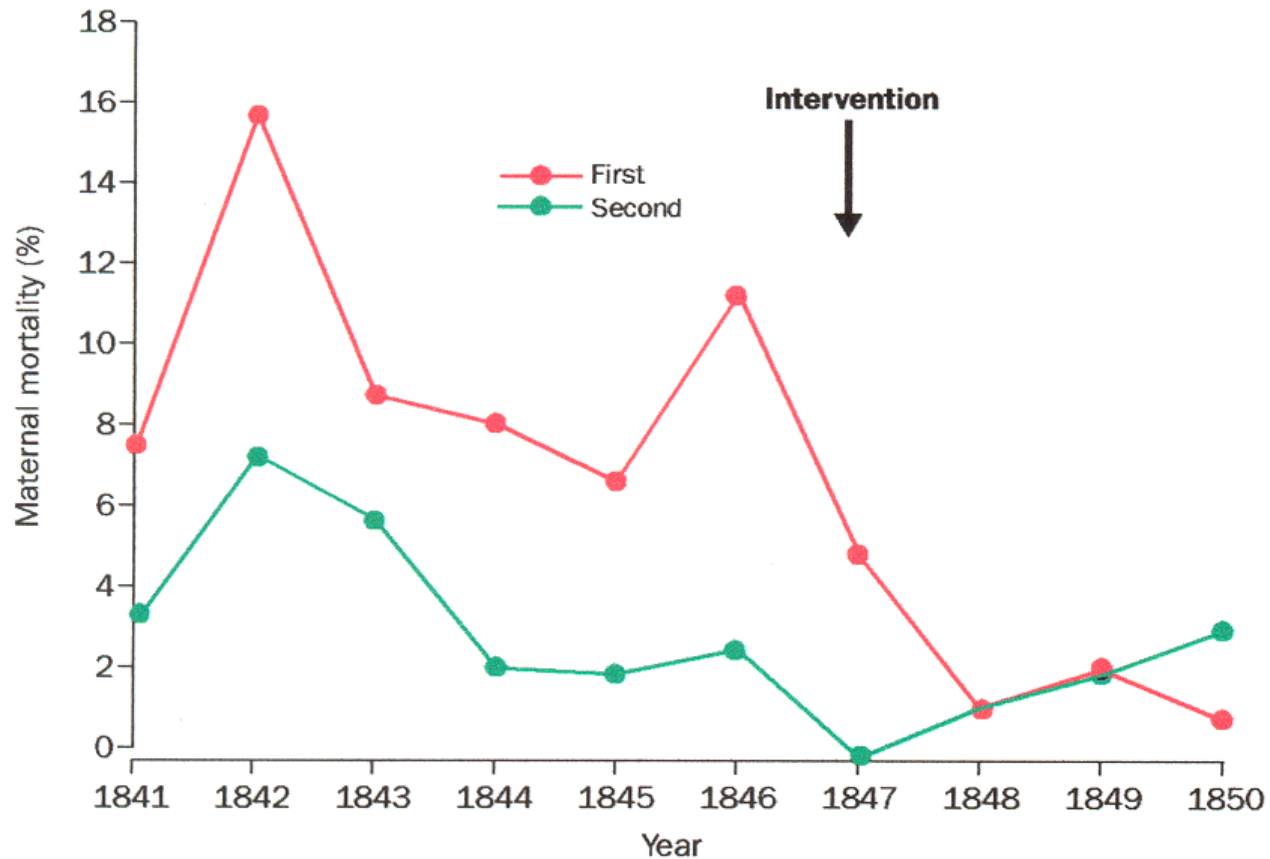
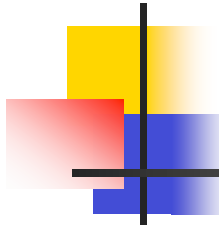






Figure 2. Maternal mortality rates in the First and Second Clinic at the Lying-In Women's Hospital, Vienna, before and after hand hygiene in chlorinated lime had been introduced in May, 1847. Rates have been calculated according to numbers given in reference 22.



# Lunacy Act 1845

---

- All managers of psychiatric institutions were required to evaluate the success of their institutions by reporting annually patient outcomes in relation to four criteria.
- Were the patients:
  -  Dead?
  -  Recovered?
  -  Relieved?
  -  Unrelieved?
- Failure to collect these data incurred fines for physicians of £2
- These data were collected by all psychiatric institutions until 1948 and by some acute hospitals such as St Thomas's and the London




# Patient reported outcome measures (PROMs)

---

- Measuring changes in physical and psychological functioning or quality of life before and after medical care e.g. before a hip replacement and 3 months after it.
- Using quality of life measures translated into dozens of languages and used in thousands of clinical trials e.g. EQ5D ([www.euroqol.org](http://www.euroqol.org) )
- Comparing quality of life before and after health care, does the patient get better functioning?

# Measuring Patient Outcomes in the English NHS







<b>Procedure</b>	<b>Condition-specific</b>	<b>Generic</b>
Primary Unilateral Hip Replacement	Oxford Hip Score	EQ5D
Primary Unilateral Knee Replacement	Oxford Hip Score	EQ5D
Groin Hernia Repair	None	EQ5D
Varicose Vein Procedures	Aberdeen Varicose Vein Questionnaire	EQ5D
Plus a standard set of patient-specific questions in all cases		

*Source: DH Operating Framework, Guidance on the routine collection of patient-reported outcome measures, Department of Health 2007*



# Pay for performance (P4P)

---

- P4P for hospitals
  -  US Premier-Medicare programme
  -  UK CQUIN
  -  “Never” events
- P4P for individual practitioners
  -  UK general practitioners (primary care)
- Does P4P improve efficiency?



# US Premier-Medicare

---

- Hundreds of self selected hospitals giving care to Medicare patients (mostly citizens over 65 years of age)
- Focuses on five areas: acute myocardial infarction, heart failure, community acquired pneumonia, coronary artery by-pass grafts and hip and knee replacement surgery.
- Hospitals measure performance in relation to 33 agreed quality indicators.
- Composite quality score ranks hospitals
- Best decile gets 2% bonus on Medicare tariff payments; second best decile gets an additional 1%. Poorest decile to lose 2% and second worst decile to lose 1%





# English Commissioning for Quality and Innovation












---

- Purchasers of NHS care (Primary Care Trusts (PCTs)) setting standards for care. Mixture of input standards, process measures derived from Premier and outcomes (PROMs)
- In 2009-10 all providers (public and private hospitals providing care to NHS patients) have to give PCTs baseline data on performance.
- Failure to provide these data will reduce tariffs by 0.5% in 2009-10
- Performance will then be regulated with possible loss of 3-4% in later years if providers fail to meet purchaser requirements



# CQUIN: an example







---

-  Movement towards consultant obstetrician presence on labour ward
-  Movement towards midwife: delivery staffing ratios in “Safer Childbirth”
-  Increasing percentage of mothers breastfeeding on discharge home
-  Improving and sharing personalised care plans.
-  Use of Common Assessment Framework (CAF) for vulnerable children.
-  Improving the care of children and young people with diabetes mellitus.
-  Implementation of NICE clinical guideline 68 (stroke and TIA)
-  Improving end of life care (e.g. the Liverpool pathway)
-  Hip and knee best practice bundle
-  Acute myocardial infarction best practice bundle
-  Care and compassion, in particular avoidance of malnutrition and pressure sores
- Patient reported outcome measures (PROMs)
- Baseline measurement in 2009-10 with risk of loss of 0.5% of tariff, with change from 2010 onwards and up to 4% of tariff income at risk



# Medical errors: "never" events

---

- Since October 2008 Medicare has refused to reimburse hospitals for some never events (marked with \* below)
-  Medication errors: wrong dose/wrong drug
-  Wrong site surgery/item left in patient after surgery
-  Pressure sores (\*)
-  Infections: C.Diff, MRSA, central line infections, catheter associated urinary tract infections (\*).
-  Falls and trips.
-  DVT prophylaxis
- Why should purchasers pay for never events?



# P4P for individual clinicians

---

- UK: the “quality outcomes framework” (QOF) identified a group of “preventive” interventions.
- Each intervention was given a level of points, each worth £75 (now £125)
- Performance measured and points awarded to the practice/group of practitioners
- Cost an additional £1 billion




## United Kingdom Quality Outcomes Framework (QOF) for primary care, 2004

<b>Disease</b>	<b>Performance indicator</b>
Asthma	% of patients with asthma who have had an asthma review in the previous 15 months
Cancer	% of patients with cancer reviewed within 6 months of confirmed diagnosis
Chronic obstructive pulmonary disease (COPD)	% of patients with COPD with diagnosis confirmed by spirometry and reversibility testing
Coronary heart disease (CHD)	% of patients with CHD whose last blood pressure measurement was 150/90 mm Hg or less
Diabetes	% of patients with diabetes whose last blood pressure measurement was 145/85 mm Hg or less
Hypertension	% of patients with hypertension with last blood pressure measurement was 150/90 mm Hg or less
Hypothyroidism	% of patients with hypothyroidism with thyroid function tests recorded in the previous 15 months
Mental health	% of patients with severe long-term mental health problems reviewed in the preceding 15 months

# GP Contract quality framework

## A: Clinical indicators 2004



---

■ CHD:	121
■ Stroke:	31
■ Cancer:	12
■ Hypothyroidism:	8
■ Diabetes:	99
■ Hypertension:	105
■ Mental health:	41
■ COPD:	45
■ Epilepsy:	16
■ <i>Total:</i>	<i>550</i>



# Problems with the GP-QOF

---



## Cost



Were the most cost effective interventions prioritised?: the evidence base for the interventions was imperfect (Fleetcroft and Cookson, JHSRP, 2006)



Using the National Institute for Health and Clinical Excellence (NICE) to identify more efficient interventions to prioritise



Changing targets in fee for service systems disrupts clinical practice and creates opposition



What was the opportunity cost i.e. what was given up by practitioners?



What was the basis of the weighting/points allocation: population health gain or administrative burden to doctors?



Conclusion: swift change in behaviour but was it efficient?



# Does P4P improve efficiency?

---

- US Premier: evidence of improved performance in relation to use of agreed processes of care. No evidence of cost savings. No evidence of improved mortality outcomes (Ryan, HSR, 2009; Ryan, Health Economics, 2009)
- Other US evidence e.g. Mullen, Frank and Rosenthal (NBER paper 14886, 2009) failed to find evidence that P4P in a large network of HMOs resulted in a major quality improvement
- UK CQUIN and PROMs: no evaluation data
- GP-QOF expensive but was it efficient?





# Policy and research challenges

---

- Does reforming organisational structure lead to improved adherence to evidence based processes of care, and does this improve patient outcomes?
- Is it more efficient to incentivise hospitals (Premier/CQUIN) or to incentivise individual practitioners?
- Are the incentives used too small to have an effect?
- Are penalties (loss of income) more efficient than bonuses (gains in income)?



# Overview

---

- Behaviour affected by incentives
- But what are the relative roles of career development opportunities, management and incentives in altering clinical behaviour?
- Managers usually lack authority to change clinical behaviour even if they have the comparative cost, activity and outcome data to identify poorly performing outliers
- Clinical and managerial collaboration is important but the role of financial incentives has to be better evidenced