P4P does work. It just needs using appropriately! Lessons from the UK (To P4P or not to P4P)

Dr Stephen Campbell
Health Sciences-Primary Care Research Group
University of Manchester

3 March 2011
Declaration of interests

• Co-lead UK Quality and Outcomes Framework - NICE External contractor

• Unrestricted research grant from NICE as the external contractor on QOF (2009-2013)

• Co-developed many of the indicators in the original QOF

• Not representing any organisation but my own views
What I’m going to cover

1. P4P update in the UK: what is QOF?
2. Factors that underpin P4P/QOF: IT; PHCT; money; registers
3. Evidence base: performance „improvement’ data
   • Changes to practice management
   • Incentivised areas / Non-incentivised areas
   • Which types of indicators work in P4P?
   • What happens to an indicator when P4P is removed?
4. Unintended consequences of QOF
5. Gaming / exception reporting
6. The value of piloting
7. Coordinated care for diabetes
8. Conclusions
1. UK Background: quality in the 1990s

Motive:
- Recognition of care variation by the medical profession
- Recognition of care variation by Government

Means:
- Development of methods of measuring quality
- Increasing computerisation of practices and electronic data record
- Rise of evidence based medicine

Quality improvement initiatives:
- National Service Frameworks for major chronic diseases
- Audit
The U.K.’s total expenditure on health was 6.6% of Gross Domestic Product (GDP) - 13.4% in the United States.

Per capita total health spending was only US$1,813 in the U.K., compared with US$2,387 in France, US$2,580 in Canada, US$2,780 in Germany, and US$4,540 in the United States.

U.K. health care infrastructure was outdated, not enough health professionals

Primary health care was under-resourced.

2000 NHS Plan for Reform and Investment. Over the next ten years, spending on the NHS increased by 43% in real terms

Total health spending increased to 8.7% of GDP by 2008, close to the OECD average of 9.0%
1. UK expenditure on health care since 1990

Political will to invest in the NHS underpinned by sustained economic growth
1. Quality of care in the UK improved between 1998 and 2003

- Quality was improving already
- So evidence of improvement pre-P4P.

Campbell et al. BMJ 2005; 331: 1121-1123
1. P4P: The Quality and Outcomes Framework

- Introduced in April 2004 for all practices in the UK
- Income dependent on performance against quality indicators
- Mostly clinical and mostly process measures rather than outcomes
- Each indicator allocated between 0.5 & 56 points
- Each point now earns an average practice £125 (200A$)
- Maximum of:
  - £79,800 (127,910 A$) per practice
  - £25,000 (40,072 A$) per physician
  - Achievement scores are publicly reported: www.qof.ic.nhs.uk

- 25% of practice income = P4P
Factors that underpin QOF

IT and Data

• QOF relies on accurate and reliable data extraction, as would any P4P Framework (esp as a payment mechanism)
• But also data entry: Business rules, READ codes, GPSS
• Registers: 1st rule!
• Timeframes: 15 months
• Precise denominators/numerators/exceptions
• Crucial to involve IT experts from the start…
Factors that underpin QOF

Infrastructure

- Practice management
- PHCT
- Nurses
- Computerisation / templates / protocols
- Recall
2. Why?

- As a basis for quality improvement: comparisons can stimulate and motivate change
- As part of pay for performance schemes (e.g. QOF)
- As part of regulation (e.g. of minimum standards)
- To assist purchasing (e.g. contracts which include minimum quality standards)
- To identify areas of need for future investment
- To inform service users

- Clear purpose and clear criteria by which to judge subsequent success
2. Purpose of QOF?!

- Intended consequences of the new contractual arrangements were to reward quality of care rather than numbers of registered patients, improve data capture and care processes, and to improve patient outcomes and doctor working conditions.

- Central dilemma? Is it a payment mechanism or a quality improvement scheme?
2. USA v. UK: P4P

- In a national survey in USA, 52% of HMOs (covering 81% of enrollees) report using pay for performance (Rosenthal 2006)

- Average of 5 performance measures per scheme. (UK QOF 86 clinical indicators!!)

- Rewards for reaching fixed threshold dominate; only 23% reward improvement

- 5-7% of physician pay (UK QOF: ~25%!!)
3. Improved organisation of care

- Increased computerization
- Better organised care & systematic protocol driven care
- Greater use of templates improves recording and outcomes of care
- Recall means better follow up

- Role of nurses may be changing
  
  McDonald, Lester and Campbell, Soc Sci Med 2009; 68(7);1206-1212
Nurses: case study

• The majority of chronic disease management associated with delivering standards in the UK QOF is undertaken by nurses (McGregor et al 2008; Gemmell et al 2009)
• Practices that employ more nurses perform better in a number of clinical domains measured by the QOF (Griffiths et al 20010).
• The new GMS contract has given practice nurses increased autonomy and job satisfaction (Maisey et al 2008).
• However, discontent and resentment about how the financial gains are targeted at equity principal partners only (McGregor et al 2008; Campbell et al 2008) and salaried GPs (Lester et al 2009).
• PNs can feel isolated and unsupported especially in smaller practices (O’Donnell et al 2010).
• Practice nurse work is changing to reflect a more medical, and less caring, template driven orientation to service delivery (McDonald et al 2009).
3. Achievement for 50 ‘stable’ clinical indicators

**Reported achievement**

**Overall, stable indicators**

<table>
<thead>
<tr>
<th>Year</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004/5</td>
<td>84.9%</td>
</tr>
<tr>
<td>2005/6</td>
<td>89.2%</td>
</tr>
<tr>
<td>2006/7</td>
<td>91.0%</td>
</tr>
<tr>
<td>2007/8</td>
<td>90.9%</td>
</tr>
<tr>
<td>2008/9</td>
<td>90.8%</td>
</tr>
</tbody>
</table>

**Stable indicator:** continuously incentivised under the QOF since 2004-5.
3. Protecting patients against inappropriate treatment: Exception reporting

- Physicians may exclude („exception report”) patients for whom targets are inappropriate
- Discretionary exception reporting:
  - treatment not tolerated
  - patient terminally ill, exceptionally frail or has supervening condition
- Non-discretionary exception reporting:
  - patient refuses intervention or repeatedly refuses to attend
  - patient newly registered or diagnosed
  - service not available to practice
Concerns over inappropriate use of exception reporting have led to calls for the provision to be amended (Audit Commission 2011).

Cheating…!

- Exception reporting rates low (median rate 6 percent) [Doran 2008], (stayed <5.5%) with little evidence of widespread fraud [Simpson 2007]. Still at 5%
- However, a minority of practices have achieved higher scores by excluding unusually large numbers of patients [Doran 2006, Gravelle 2008]
- Overt use as a means of ‘gaming’ admitted by GPs in two of 27 participating pilot practices (Campbell et al In Press).

Most practices value and use exception reporting as a clinical safeguard to quality individual patient care within an evidence-based but largely population level and inflexible framework.

“I think you do need scope to exception report patients because people are really quite individual and unique, and I think, you know, in the end there is always an element of clinical judgment.” (GP, 29)

“The only rule I think we have is that we feel that every exempt reporting needs to be defensible. So if you ask me why did you exempt this patient I should have an answer…even if you ask me three years later I should be able to look at a note and say that’s why.”(GP, 17)

Evidence base

- QOF presumes “average” patients
- GPs’ wishes to espouse and practice holism and act as patient advocates.
- Concern that the financial pressure to meet the QOF targets (25-30%) may risk patient’s agenda being ignored.
- Patient-centred care requires both patients’ and doctors’ differing priorities to be negotiated to achieve ideal outcomes.
- Safeguard against the inappropriate or over-treatment of patients. Individually tailored care rather than focusing on generalised population based care (Heath 2007; 2009).
Clinical performance – patient evaluation

Figure 1: Mean Scores for (a) Clinical Quality at the Practice Level for Coronary Heart Disease, Asthma, and Type 2 Diabetes, 1998 to 2007 and (b) for patient evaluations of communication with their physician, access to care and continuity of care, 1998 to 2007

Two aspects to clinical indicators:

- a disease condition (e.g. diabetes, CHD).
- a care activity (e.g. influenza vaccination, BP control).

Three indicator classes, in terms of incentivisation:

- (A) Condition & process incentivised within QOF (28 ind)
- (B) Condition or process incentivised (13 ind)
- (C) Neither condition nor process incentivised (7 ind)

Doran, Campbell et al BMJ (In Press)
Incentivised v. non-incentivised

- Incentivised conditions - The rate of improvement was not sustained and improvement appeared to plateau after 04/05.
- QOF did not generate positive spill-overs to other activities & appears to have had a negative impact on non-inc ones for otherwise inc conditions.

Doran, Campbell et al BMJ (In Press)
Which types of indicators work in P4P?

Process indicators: % of patients with hypertension with BP recorded in last 15 months (98%)

Treatment: % of patients with CHD on Statin (82%)

Intermediate outcomes: % of patients with diabetes with HBA1c <7 (54%)

Depends where you put the threshold
## Criteria for retiring indicators

<table>
<thead>
<tr>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reported achievement</td>
</tr>
<tr>
<td><strong>Average rate</strong></td>
</tr>
<tr>
<td><strong>Variation</strong></td>
</tr>
<tr>
<td><strong>Historical trends</strong></td>
</tr>
<tr>
<td>Exception reporting</td>
</tr>
<tr>
<td><strong>Average rate</strong></td>
</tr>
<tr>
<td><strong>Variation</strong></td>
</tr>
</tbody>
</table>

Reeves, Doran.. Campbell, Lester et al. BMJ, 2010
How do you remove indicators and then what happens?

• Very limited UK evidence base: data not collected after removal – no strategy
• Kaiser Permanente in California – up to 8 years of data including 4 „shared’’ indicators with QOF
Red dot: incentive off, Green dot: incentive on

Hypertension Control (systolic<140), ages 20 and up

Diabetic Retinopathy Screening, ages 31 and up

Cervical Cancer ages 21-64

Diabetes Glycemic Control (<8%), ages 18-75

Lester, Campbell et al BMJ 2010
Unintended consequences

- Change in professional values: meet the target
- Changes to practice nurse roles: continuity, professionalism, responsibility. But finance?
- Salaried doctor roles
- Less holistic approach /measure fixation/ tunnel vision
A general practitioner reflecting upon his career... Started 30 years before

“There can be no doubt that the relations between the public and the family doctor were more cordial than they are at the present time” (doctor-pat relationship)

“Changing the (GP) was less frequent than it is now” (continuity of care)

Complains “of what is now too rare, the old feeling of a family doctor...the old friend...who knows the kind of stuff his flock is made of, is gone”. (personal care)

Dr James Clarke

1874
New ways of working

Stage 1 – Collation of Information: Based on NICE guidelines

Stage 2 – Prioritisation: Advisory Committee

Stage 3 – Indicator Piloting

Stage 4 – Validation & Publication

Stage 5 – QOF Changes

24 months
## Indicator testing protocol

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Summary / recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarity</td>
<td>● RAND Appropriateness Method</td>
</tr>
<tr>
<td>Acceptability (interviews)</td>
<td>● Risks, issues, relative impact and uncertainties</td>
</tr>
<tr>
<td>Feasibility (IT-NHSIC)</td>
<td>● ‘Technically feasible’ in current GP systems and supported by current methods of data extraction for QOF,</td>
</tr>
<tr>
<td>Reliability (IT-NHSIC)</td>
<td>● Reproducible in test pack</td>
</tr>
</tbody>
</table>
| Implementation (interviews with staff and patients) | ● Baseline and final (evidence of sensitivity to change);  
● Exception reporting / gaming;  
● Changes in practice organisation; potential barriers;  
● Unintended consequences |
| Changes existing QOF indicators | ● Summary of any suggested changes                                                        |
| Changes in wording            | ● Summary of any suggested changes                                                        |
| Cost effectiveness            | ● Summary of evidence of cost effectiveness                                               |
| Overall NEC recommendation    | ● 1) no major barriers/risks/issues/ uncertainties  
● 2) some barriers/risks/issues/uncertainties but okay  
● 3) major barriers/risks/issues/uncertainties preclude it |
Acceptability - implementation

Palliative care

Indicator as piloted:

- The percentage of patients on the palliative care register who have a preferred place to receive end-of-life care documented in the records

- Anxiety over the rigidity of the stipulated timeframes which are too prescriptive

- Perceived potential harm to patients: timing of raising preferred place of death is subjective and patient specific

- Undue focus on one isolated question from a multifaceted and complex issue → measure fixation
EVIDENCE BASE FOR P4P
P4P programs should:

(1) select and define P4P targets on the basis of baseline room for improvement,

(2) make use of process and (intermediary) outcome indicators as target measures,

(3) involve stakeholders and communicate information about the programs thoroughly and directly,

(4) implement a uniform P4P design across payers,

(5) focus on both quality improvement and achievement,

(6) distribute incentives to the individual and/or team level.

Money, money, money!

- The financial incentives must be large to be effective (Epstein AM, 2009)
- Without new money, some physicians may lose funds and are unlikely to embrace the new payment systems

- Overall findings are mixed: some strongly positive results. Not compelling
- 4 studies showed unintended effects such as adverse selection and improved documentation rather than delivery of care
- Data are lacking: P4P can work; P4P may fail

Christianson et al for HF 2008

- P4P creates short term improvement
- Impossible to attribute to P4P due to background noise
Using indicators in P4P: Conclusions

- Small gains at large cost  (Scott, Aus Health Review 2008; Christianson et al, Health Foundation, UK, 2008)

- “make haste slowly”
Aim: to reduce the level of teenage pregnancy

- Target GPs – *wrong level*. There is a limited amount that GPs can do to reduce teenage pregnancy. They could be more proactive about offering contraceptive advice to teenagers, but many in this age group are not even in regular contact with their general practice.

- Primary care trust (Medicare Local) – *right level*. Broader responsibilities, may link with local authorities to provide better information in schools, or may providing proactive care through school nurses. PCTs may also review overall provision of contraceptive services and provide funding to other providers, e.g. Brook Advisory Clinics.

- **Health system level**
## Levels of application for clinical indicators

<table>
<thead>
<tr>
<th>General practice level</th>
<th>RACGP level</th>
<th>Health system level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease specific</td>
<td>Disease specific</td>
<td>Disease specific indicators</td>
</tr>
<tr>
<td>Preventive care</td>
<td>Preventive care</td>
<td>Preventive care</td>
</tr>
<tr>
<td>G practice improvement</td>
<td>G practice improvement</td>
<td>G practice accreditation</td>
</tr>
<tr>
<td>G practice accreditation</td>
<td>G practice accreditation</td>
<td>Health system improvement</td>
</tr>
</tbody>
</table>

| Quality and safety     | Quality and safety | Quality and safety |

**Clinical indicators and the RACGP: Policy endorsed by the 51st RACGP, Council 5 May 2009**
Marmot Indicators for Local Authorities in England: locus of control

Marmot Indicators for Local Authorities in England

*Fair Society, Healthy Lives: The Marmot Review* report was published in February 2010 ([www.marmotreview.org](http://www.marmotreview.org)). The report included some suggested indicators to support monitoring of the overall strategic direction in reducing health inequalities. The London Health Observatory and the Marmot Review Team have produced baseline figures for some key indicators of the social determinants of health, health outcomes and social inequality that correspond, as closely as is currently possible, to the indicators proposed in *Fair Society, Healthy Lives*.

The indicators are shown below in spine chart format. The spine charts contain the following indicators:

- Male life expectancy
- Female life expectancy
- Slope index of inequality (SII) for male life expectancy
- Slope index of inequality (SII) for female life expectancy
- Slope index of inequality (SII) for male disability-free life expectancy
- Slope index of inequality (SII) for female disability-free life expectancy
- Children achieving a good level of development at age 5
- Young people who are not in education, employment or training (NEET)
- People in households in receipt of means-tested benefits
- Slope index of inequality for people in households in receipt of means-tested benefits

To view the spine chart for your area please select your region and your local authority from the drop down menus below.

P4P: Not a magic bullet or a panacea

- Pay-for-performance/QOF: only as one part of a multiple systems based strategy for quality improvement – not a panacea.
- Find out the baseline
- Mix of organisational/clinical/patient experience schemes and accreditation
- Mix of financial and professional incentives
- Training; curricula, certification: nurses and PHCT/Allied
- There will be unintended consequences!
- Cost effectiveness - limited data!
Balanced scorecard / approach

- P4P
- Public reporting
- Accreditation
- Practice level/ML/ State/ health system
- Regulation
- Investment

- MACRO-MESO-MICRO
Balanced scorecard / approach

Test average: 53.52
39 Test centuries
12363 Test runs
Test wins as captain: 47
3 World cup wins
3 Ashes series loses
AUSTRALIA?
To be an ideal P4P indicator, the clinical issue in question should be:

- **Common**: high prevalence, high impact and common on practice level (Diabetes not MS?)
- **Have significant morbidity and/or mortality**
- **Recognised gap** between actual and potential performance; so baseline is known
- **Evidence base** to create indicators/targets
- **Evidence base**: process leads to better outcomes
- **IT logical /internally consistent**: extractable from all systems
- **Directly under control of every practice or medicare local?**
- **Equal access** to expanded services
- **Free of obvious unintended consequences**
- **Does attribution matter?**

Based on Campbell and Lester 2010
Policy window

- Decisiveness and political will are important
- UK QOF introduced without piloting
- Improving Primary Health Care for All Australians
- Medicare Locals Guidelines
Kingdon’s model (1995) examined how issues get onto the policy agenda and become translated into policy and identified three streams:

- Problem: policy objectives
- Policy: proposals, strategies
- Politics: bargaining, compromise, negotiation and political will

When all three streams flow together a policy window, (a short lived or temporary opportunity for pursuing significant policy change) exists

“Appreciative setting” policy-makers show readiness to respond to an agenda item (Vickers 1995).

However, the subsequent agenda item must be seen as amenable to a policy intervention
Implementing indicators

What needs to be in place to implement indicators/incentives?

- Exworthy and Powell (2004) suggested implementation is influenced by the three streams of:
  - **Policy**: goals and objectives
  - **Process**: technical and political feasibility
  - **Resources**: financial and human capacity including PHCT - nurses
Coordinated care for diabetes

Are the following in place?:

- Goals, policy objectives/proposals    ✓?
- Technical feasibility / IT            X?
- Political feasibility / will          ✓?
- Financial and human capacity          ✓?
Improving PHC for all Australians

- Having the right workforce
  - Structure / process / outcomes // training, primary care trained and employed nurses with career pathways// allied health

- Having the right infrastructure
  - Structure

- Identifying and addressing gaps in local services
  - Structure / process / HCA / asking patients / outcomes

- Making sure the HS works for patients & providers
  - Asking patients and staff / structure-process-outcome
  - P4P: a role to play

- Timely and local access / responsiveness / less use of hospitals
Medicare locals

- Who does your community serve? HNA
- What are their needs? Local QOF/ask patients
- Who will deliver care? PHCT
- Local access to services they need / extended hours: don’t sacrifice everything for access
- Meet local needs: evidence of “success” may be different?
- Out of hours/deputising: structure-process-outcome
- Work closely with hospital networks – indicators on info exchange, planning, budgeting, etc
- Accountability: S-P-O / contractual indicators
Coordinated Care for Diabetes pilot: July 2011

• Enrolment period on Diabetes Pilot
  • Must correspond to indicators: i.e. BP checked in last 12 months must have 12 month enrolment

• Clinical indicators based on evidence
  • Find out baseline

• Involve the practice team – inc GP and PN in PHCT
Australia: key lessons and recommendations

• **Diabetes Coordinated Care Pilot:**
  - Workforce issues: involve nurses with responsibility: prof incentive
  - Learn from German enrolment and DMPs etc
  - Monitor non-incentivised areas & emphasise guidelines

• **Culture of improvement:** not in UK. <10% of income

• **P4P for longer term health improvement:** Limit size, focus on clinical indicators: a cycle of piloting and removal as one part of wider multiple strategy. One part of wider strategy

• **Meet the needs of patients:** Ask patients!

• **QOF has created a climate of expectation that improvement or targeted areas MUST have a financial incentive**
  - Professional and reputational incentives?
So...if not P4P (on its own)...then what...?

- P4P
- Professional incentives
- Penalties/regulation: police minimum standards and safety
- Ask patients!
- Investment for Improvement
  - Nurses, allied staff, buildings, computers etc
Health care reform is political!
New reforms: UK!!

• New government plans....
• 2011 White paper
• £80billion of NHS budget -GP commissioning and Consortia
• Any Willing Provider…
• No PCTs, no SHAs
• Commissioning Board
• Focus on “outcomes”
• Centre for Commissioning (CfC): Good Commissioning Guides (i.e. commissioning diagnostic services)
Since 1991: family practitioner committees, health authorities, GP fundholders, total purchasing consortiums, GP multifunds, primary care groups, primary care trusts, and external commissioning support agencies. Yet…

“little evidence to suggest that any of these organisational structures for commissioning are better or worse than others, or that the proposed new consortiums will work any better than the current arrangements.”

• King’s Fund: GPs who are successfully involved in practice based commissioning should be given real rather than indicative budgets for some services and their performance monitored closely (2011)
• PBC: mixed results. lack of time, resources and personnel, and difficult relationships with the PCT (Checkland 2009)
• Deaths from heart attacks and cancer falling, despite lower spending on health than other EU countries (Appleby BMJ).
• “Lansley developed a lot of these ideas when the NHS budget was still growing significantly” (Prof Chris Ham)
• YouGov poll 2011; 27% support GPs using private companies to provide NHS services and 50% oppose the measure
Thank you very much for listening!

stephen.campbell@manchester.ac.uk

Dr Stephen Campbell, Senior Research Fellow, Health Sciences – Primary Care, Williamson Building, University of Manchester, Oxford Road, Manchester, M13 9PL
Tel: 00441612757655
NHS outcomes framework, identifying key indicators that will judge the service across five „domains‟.

Effectiveness
- Preventing people from dying prematurely
- Enhancing quality of life for people with long term conditions
- Helping people to recover from episodes of ill health or following injury

Patient experience
- Ensuring that people have a positive experience of care

Safety
- Treating and caring for people in a safe environment and protecting them from avoidable harm
Each domain follows a similar structure with:

- a single or small number of overarching indicator(s), allowing the Secretary of State for Health to track the progress of the NHS as a whole in delivering outcomes across the breadth of activity covered by the domain;

- a small set of improvement areas where the NHS Commissioning Board will be tasked with delivering better outcomes because the evidence suggests that significant improvement or health gain is possible; and

- a supporting suite of NICE Quality Standards setting out what high quality care looks like for a particular pathway of care.

Across the five domains there are 10 overarching indicators, 31 improvement areas and 51 indicators in total.
Outcomes??

I'm talking about outcomes.

Why do I feel this urgent need to talk about outcomes?

It's not about theory or ideology, it's about people's lives!

But I hear you ask: isn't it too much all at once?

Does it have to be so fast? Does it have to be so soon?

I'm afraid it does.

Because it was all carefully worked out when we were in opposition!