Investigating the impact of the Quality and Outcomes Framework on the quality of primary care

Evan Kontopantelis  Tim Doran*  David Reeves
Jose Valderas  Martin Roland

National Primary Care Research and Development Centre
University of Manchester England

RSS, 9th September 2009

Outline

1 UK pay-for-performance scheme
   - The idea
   - Framework details

2 Methods
   - The design
   - The General Practice Research Database
   - Analysis

3 Results
   - Short term QOF effect (2004/05)
   - Long term QOF effect (2004/05-2006/07)
   - Summary
A P4P program kicked off in April 2004 with the introduction of a new GP contract.

- General practices are rewarded for achieving a set of quality targets for patients with chronic conditions.
- The aim was to increase overall quality of care and to reduce variation in quality between practices.

- The incentive scheme for payment of GPs was named Quality and Outcomes Framework (QOF).
- QOF is reviewed at least every two years.

Estimated cost of £1.8b, over 3 years (escalated to £2.8b).

- GP income increased by up to 25%, dependent on performance.
- 146 quality indicators.
  - Clinical care for 10 chronic diseases (76 indicators).
  - Organisation of care (56 indicators).
  - Additional services (10 indicators).
  - Patient experience (4 indicators).

- Implemented simultaneously in all practices (a control group was out of the question).
Performance.
Incentivised aspects keep improving but what about the non-incentivised ones?

- Quality scores for all QOF clinical indicators have been improving.
- They are only a small proportion of all clinical care.
- There are concerns that quality for non-incentivised aspects of care may have been neglected.
- How to measure performance on the non-incentivised aspects though?

Clinical indicators.
..and research questions.

- Two aspects to clinical indicators:
  - a disease condition (e.g. diabetes, CHD).
  - a care activity (e.g. influenza vaccination, BP control).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Incentivised (QOF)</th>
<th>Non-incentivised</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incentivised activity (QOF)</td>
<td>A: 30</td>
<td>B: 9</td>
</tr>
<tr>
<td>Non-incentivised activity</td>
<td>C: 7</td>
<td>D: 8</td>
</tr>
</tbody>
</table>

We aimed to compare the four classes on changes in quality from pre-QOF to post-QOF.
- Would class A indicators show most improvement?
- Would classes B and C show some ‘halo’ effects since they involve either a QOF condition or activity?
- Has quality for class D indicators declined?
GPRD holds event data for more than 270 English practices, from 1999 (488 active practices in 2007 and around 10m patients).

Database is broken down to numerous tables, because of the volume of data.

Final sample was 600,000 patients from 150 practices (over 24GB in size).

Allows us to:
- Investigate quality of care at the patient level.
- Model and measure non incentivised aspects of care.

Event files.
- **Clinical**: all medical history data (symptoms, signs and diagnoses).
- **Referral**: information on patient referrals to external care centres.
- **Immunisation**: data on immunisation records.
- **Therapy**: data relating to all prescriptions issued by a GP.
- **Test**: data on test records.

Lookup files.
- **Medical** codes: READ codes, 111,865 available.
- **Product** codes: 77,198 available.
- **Test** codes: 304 available.
Size of the tables prohibits looking at codes one by one.
Instead we use search terms to identify potentially relevant codes in the lookup tables and create draft lists.

**Example (Search terms for diabetes)**
- String search in Medical codes: 'diab’ ‘mell’ ‘iddm’ ‘niddm’.
- READ code search in Medical codes file: 'C10' 'XaFsp'.
- String search in Product codes file: 'insulin’ 'sulphonylurea’ ‘chlorpropamide’ ‘glibenclamide’.

Clinicians go through the draft lists and select the relevant codes.
Three sets of codes are created, that correspond to:
- QOF criteria.
- Conservative criteria.
- Speculative criteria.

Using the finalised code lists we search for events in the Clinical, Referral, Immunisation, Therapy and Test files.
The process.
Step 3: extracting the events of interest, for each indicator.

- For each indicator and year, we used the associated clinical rules and retrieved events to construct a % score:
  - (cases where indicator was met / eligible patients)*100%.

Example (C4 indicator, QOF year 06/07: 01Apr06-31Mar07)
- Percentage of patients with hypertension who have a record of blood glucose once since diagnosis.
- Denominator: Patients registered for the full year with the practice, diagnosed with hypertension at any point until the 31st of March 2007 (excluding resolved cases).
- Numerator: Patients in the denominator who have had at least one record of blood glucose, postdating the first (unresolved) hypertension diagnosis.

Issues to tackle.
Is the analysis ever easy...?

- Indicator classes are imbalanced.
- Three different types of activities:
  - clinical processes related to measurement (PM/R).
  - clinical processes related to treatment (PT).
  - intermediate outcome measures (I).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Incentivised (QOF)</th>
<th>Non-incentivised</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PM/R</td>
<td>PT</td>
</tr>
<tr>
<td>Inc. activity (QOF)</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>Non-inc. activity</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

- Quality of care was already improving (prior to QOF).
- The ceiling has been reached for certain ‘easy’ indicators.
The approach. Interrupted Time Series.

- ITS multi-level multivariate regression analyses, allowed us to compare the four indicator classes on the basis of two outcome measures:
  - The level difference between the estimated pre- and post-QOF trends, in 2004/05.
  - The change in slope from the pre- to the post-QOF trend.

The main analysis used logit-transformed scores.
Regression models run with `xtmixed` in STATA (mle).
Practices were treated as crossed with indicators and random intercepts were used.
Created variable combining class & type into 9 ind groups.
Analysed 3 models, for each of the outcome measures:
  - Uncontrolled
  - Controlled pre-QOF slope & mean of each indicator - RE*.
  - Controlled pre-QOF slope & mean of each indicator - FE.

Using the estimated means for the 9 indicator groups we compared across classes on: (i) PM/R types only (ii) PT types only (iii) PM/R & PT types combined.
The QOF impact

Type PM/R indicators of class A significantly increased in level above expectation post-QOF.

Type PM/R indicators of classes B & D significantly decreased in level below expectation post-QOF.

Similar results for classes A & D (but not B) using the combined process type indicators.
UK pay-for-performance scheme

Methods

Results

Change in level immediately following QOF.

Comparing ordered means across classes, using the Newman-Keuls procedure.

- For PM/R indicators all four classes were significantly different from each other in mean level change, bar B & D.
- For PT indicators there were no significant differences.
- For combined P indicators, pairs A & C and B & D were not significantly different, but A & C had significantly greater level changes than B & D.

<table>
<thead>
<tr>
<th>Indicator group</th>
<th>Change in level immediately following QOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-I</td>
<td>0.32</td>
</tr>
<tr>
<td>A-PM/R</td>
<td>1.23</td>
</tr>
<tr>
<td>C-PM/R</td>
<td>0.45</td>
</tr>
<tr>
<td>B-PM/R</td>
<td>-0.74</td>
</tr>
<tr>
<td>D-PM/R</td>
<td>-1.37</td>
</tr>
<tr>
<td>A-PT</td>
<td>0.24</td>
</tr>
<tr>
<td>B-PT</td>
<td>-0.33</td>
</tr>
<tr>
<td>A-P</td>
<td>0.68</td>
</tr>
<tr>
<td>C-P</td>
<td>0.35</td>
</tr>
<tr>
<td>B-P</td>
<td>-0.54</td>
</tr>
<tr>
<td>D-P</td>
<td>-0.83</td>
</tr>
</tbody>
</table>

Ordered Means

- For the A-I group.
- For all classes of PM/R indicators, bar class C.
- For all classes of PT indicators.
- For all classes of the combined P indicators, bar class C.
Change in slope from pre- to post-QOF.
Comparing ordered means across classes, using the Newman-Keuls procedure.

<table>
<thead>
<tr>
<th>Ordered Means</th>
<th>A-I</th>
<th>A-PM/R</th>
<th>B-PM/R</th>
<th>D-PM/R</th>
<th>A-PT</th>
<th>D-PT</th>
<th>B-PT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-I</td>
<td>-0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-PM/R</td>
<td>0.06</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-PM/R</td>
<td>-0.17</td>
<td></td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-PM/R</td>
<td>-0.23</td>
<td></td>
<td></td>
<td>***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-PT</td>
<td>-0.17</td>
<td></td>
<td>ns</td>
<td></td>
<td>A-P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-PT</td>
<td>-0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-PT</td>
<td>-0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-PT</td>
<td>-0.21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-P</td>
<td>-0.05</td>
<td>**</td>
<td>**</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-P</td>
<td>-0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-P</td>
<td>-0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-P</td>
<td>-0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- For the PM/R, only class C differed significantly from the other classes (in mean slope change).
- For the PT indicators there were no significant differences.
- For the combined P indicators, only class C emerged as significantly different.

Conclusions.
...results are still being interpreted!

- Short term:
  - Only the PM/R indicators within class A showed greater than expected improvement.
  - PM/R indicators in classes B and D demonstrated significantly lower than expected gains.
  - Class C (QOF-incentivised patient group) PM/R indicators significantly out-performed classes B and D.

- Long term:
  - All types and classes, with the exception of C-PM/R, experienced a significant drop in the rate of improvement.
  - QOF had an immediate ‘halo’ effect on non-QOF activities for patient groups targeted by QOF (class C).
Thank you

Not too much information, I hope!

Comments and questions:
e.kontopantelis@manchester.ac.uk