# Quality and Outcomes Framework

## Joint Executive Summary
## Reports to the Department of Health from
## the University of East Anglia & the University of York

**Corporate Author**: University of East Anglia

**Title**: Potential Population Health Gain of the Quality and Outcomes Framework. Report to Department of Health 2008

**Authors**:

*Dr. Robert Fleetcroft, Honorary Senior Lecturer, University of East Anglia*

*Dr. Sheetal Parekh, Research Associate, University of East Anglia*

*Dr. Nicholas Steel, Senior Clinical Lecturer, University of East Anglia*

*Dr. Louise Swift, Statistician, University of East Anglia*

*Dr Richard Cookson, Senior Lecturer, Department of Social Policy and Social Work, University of York*

*Professor Amanda Howe, Professor in Primary Care, University of East Anglia*

**Corporate Author**: University of York

**Title**: The GMS Quality and Outcomes Framework: Are the Quality And Outcomes Framework (QOF) Indicators a Cost-Effective Use of NHS Resources?

**Authors**:

*Mrs Anne Mason, Research Fellow, Centre for Health Economics, University of York*

*Mr Simon Walker, Research Fellow, Centre for Health Economics, University of York*

*Professor Karl Claxton, Professor of Economics, Centre for Health Economics, University of York*

*Dr Richard Cookson, Senior Lecturer, Department of Social Policy and Social Work, University of York*

*Dr Elisabeth Fenwick, Lecturer in Health Economics, Department of Public Health and Health Policy, University of Glasgow*

*Professor Mark Sculpher, Professor of Health Economics, Centre for Health Economics, University of York*

## Background

The new General Medical Service (GMS) contract was introduced into UK primary care in April 2004. The contract was supported by a significant investment, estimated to be £8 billion by the Department of Health in the first 3 years.[1] It includes a major pay-for-performance scheme for securing higher quality primary care, known as the Quality and Outcomes Framework (QOF), which rewards performance against criteria in 4 areas: clinical, organisational, patient experience and additional services. There were 10 clinical domains in the original contract, and this was revised to include a further 8 clinical domains in 2006.[1 2] Points are allocated on the basis of expected workload and quality of care.[1] Practices achieved a high standard of performance, with practices in England scoring an average of 959 points out of a possible 1050 in 2004/5, rising to 1011 points out of a possible 1050 points in 2005/6 excluding exception reported patients.[3] The revision to the contract in 2006 has increased the points allocated to clinical indicators to 665.[2] Points are not straightforwardly proportional to performance or performance improvement, but achievement is triggered at lower and upper target thresholds of attainment for each performance indicator. Payments are based on point achievement, adjusted for practice size and disease prevalence relative to national average values.

## Aims and Objectives

The Department of Health commissioned three related pieces of work from The University of Manchester, The University of York ('York') and The University of East Anglia (UEA). The University of Manchester sought to explore trends in the performance in a subset of indicators inside and outside the Quality Outcomes Framework (QOF). This work is ongoing and no further details are presented here.

The research by the UEA seeks to estimate the potential population health gain of the full implementation of the interventions in the new GMS contract. This work focussed on 'lives saved' as the measure of benefit. The York research aimed to assess the extent to which cost-effectiveness evidence supports QOF indicators that have a *direct therapeutic impact*. The benefits evaluated in this work were quality adjusted life years (QALYs), which cover mortality, morbidity and quality of life implications. The costs considered were those costs incurred by the NHS.

***Aims and objectives: University of East Anglia***

The research by the UEA aims to estimate the potential population health gain of the full implementation of the interventions in the new GMS contract, both in its original form (2003)[1] and its revised form (2006).[2] The number of lives saved is the chosen measure of health gain, as many of the clinical interventions in the new GMS contract have potential to save lives. The "potential" health gain is based on the difference between 100% indicator achievement and zero achievement; however this will be greater than the actual health gain as there was a substantial indicator achievement before the contract was introduced. Secondly, we aimed to construct the aggregate of the number of lives saved per 100,000 populations per year at two levels namely; an aggregate at the domain level and an overall aggregate of all clinical indicators.

***Aims and objectives: University of York***

The overall aim of the York project was to assess the extent to which available cost-effectiveness evidence supports those QOF indicators that have a *direct therapeutic impact*. Informed by this assessment, the cost-effectiveness of the existing QOF and the potential for improving its value for money was explored. These aims can be broken down into four inter-linking objectives:

(i) To set out an analytic framework to facilitate an assessment of the cost-effectiveness of providing financial incentives to change clinical practice.

(ii) To appraise the cost-effectiveness evidence relating QOF indicators with a potentially direct therapeutic impact.

(iii) To assess the cost-effectiveness of the use of financial incentives through QOF.

(iv) To explore how the 'optimality' of the current 'basket' of QOF indicators in the GP contract might be assessed.

## Methods

Literature reviews were used by both teams to identify relevant evidence. The UEA team searched for placebo-controlled trials providing clinical data on lives saved. Evidence on all clinical indicators in the QOF was sought. The York team searched for economic evaluations that reported both costs and QALYs, but the search for evidence was restricted to indicators expected to have a direct therapeutic impact. One key difference between the two teams' work is the comparison of interest: in the research by the York team, the comparator was 'usual practice'; in the research by UEA, the comparator was placebo. Importantly, whereas

the UEA team produced summary statistics of the overall impact of the QOF and of the clinical domains, the York team did attempt to aggregate findings at these levels.

***Methods: University of East Anglia***

The research includes four stages that mirror the aims of this project, these being search strategy, inclusion and exclusion criteria, sensitivity analysis and strategies for combining data.

*Search Strategy*

Two researchers (RF and SP) independently reviewed four sources for the highest level of evidence for lives saved in terms of absolute risk reduction for all cause mortality for each clinical indicator in the GMS contract 2003 and 2006 versions.[1 2] Sources used were the supporting documentation for the GMS contract,[1 2] the National Institute for Health and Clinical Excellence (NICE),[4] Clinical Evidence (CE) database[5] and the Cochrane Library.[6] A detailed search strategy is presented in Appendix A of the UEA report. The level of evidence was determined by using the classic "hierarchy of evidence" grading scale designed by the US Agency for Health Care Policy and Research (AHCPR 1992)[7] and adopted by the Cochrane Library.[6]

*Inclusion and exclusion of studies*

Studies included have the following characteristics:

- Clinical interventions were compared with a placebo control arm.
- Studies had clinical interventions that closely matched the clinical indicators.
- Studies had all cause mortality as an outcome.

Evidence for effectiveness for each clinical indicator in the new GMS contract was sought in each of the four sources independently by each author. Each study that had been identified by this process was then again reviewed by two researchers against the inclusion and exclusion criteria in developing the final list of appropriate studies. 28 studies were included in the final analysis.

*Sensitivity analysis*

Where several studies of similar level were identified for a particular clinical indicator, then the upper and lower risk reductions estimated by these studies were used to estimate an upper and lower limit to the population health gain. The 95% confidence intervals were given for each study included where this has been reported.

*Strategies for combining the data*

Where the absolute risk reduction (ARR) was not reported in a clinical trial, then this was calculated where possible from other measures of health gain such as relative risk reduction (RRR) and odds ratio (OR). This was combined with the UK population prevalence data to calculate the baseline risk. The risk reduction in mortality was approximated to a standardised year, with the assumption of a linear relationship between mortality and time. An adjustment was made for the prevalence of each condition to estimate the maximum health gain across a standardized population of 100,000 in terms of lives saved per year.

***Methods: University of York***

The research comprised four stages that mirrored the project objectives. First, an analytic framework was constructed to facilitate an assessment of the cost-effectiveness of providing financial incentives to change clinical practice. Second, a review of the literature and relevant websites was undertaken to identify and appraise the cost-effectiveness evidence relating to the therapeutic interventions covered in a subset of QOF indicators, namely those with a direct therapeutic impact. One or two studies were selected to provide data on costs and outcomes for each indicator, chosen on the basis of how closely they met the following criteria:

a. Set in the UK
b. Recently published
c. Reports both costs and quality-adjusted life-years (QALYs)
d. Addresses the indicator question (patient population, intervention and appropriate comparator)

One or two studies considered to best fit the indicator and providing the most relevant and robust cost and QALY data were selected for inclusion.

Third, economic modelling techniques were used to explore the potential value of existing QOF indicators in increasing the uptake of these therapeutic interventions. Cost and QALY estimates from the selected economic evaluations were used as parameters in the conceptual framework. QALYs were converted on to the monetary scale assuming a cost-effectiveness threshold of £20,000 per QALY (this is in accordance with the lower bound of NICE's stated cost-effectiveness threshold). These were combined with payment data, estimated for a hypothetical practice with average prevalence and practice size, to provide estimates of cost-effectiveness for a given QOF indicator. Five types of analysis were undertaken: the first four were based on QOF points and threshold values from 2004/05, but the fifth employed QOF

points and threshold values from 2006/07 (it should be noted that the indicators considered were those defined in the 2006/7 contract):

1. A 'base case' analysis chosen to approximately represent the indicator patient population and intervention, assuming a QOF payment duration of five years
2. Additional analyses of other related populations and/or interventions, also assuming a QOF payment duration of five years
3. A sensitivity analysis exploring the impact of extending payment duration to 10 or 15 years
4. A sensitivity analysis to test the assumption that QOF payments are purely covering the 'cost' of intervention using a 'best case' sensitivity analysis. This is only done for indicators whose cost-effectiveness is contingent upon achieving a relatively high change in utilisation, assuming a QOF payment duration of five years.
5. A second 'base case' analysis, using the same patient population and interventions employed in the first analysis, but based on QOF points and threshold values from 2006/07. Intervention costs were updated to 2006/07 values, but the cost-effectiveness threshold was held constant at £20,000 per QALY gained. QOF payment durations of five, ten and fifteen years were analysed.

Lastly, the issue of how the 'optimality' of the current set of QOF indicators in the GP contract might be assessed was considered.

## Results

Overall, the two pieces of research covered 30 indicators (Table 1). The University of East Anglia identified evidence for lives saved on 22 indicators in the 2003 contract and 19 indicators in the 2006 revised contract. The University of York identified cost-effectiveness evidence for 12 indicators in the 2006 revised contract. Evidence on both lives saved and cost effectiveness was available for nine indicators.

According to the findings from the University of East Anglia, the greatest numbers of lives saved were for influenza immunisation in patients with diabetes (DM18) or coronary heart disease (CHD12), and for controlling blood pressure in patients with hypertension (BP5). The University of York identified evidence for just one of these indicators: BP5 was found to have a net monetary benefit per patient of over £13,000. The University of York analysis suggested that the three most cost-effective indicators (in terms of net monetary benefit per patient) were use of ACE inhibitors or angiotensin receptor blockers for Chronic Kidney

disease (ChKD4; not assessed by UEA); anticoagulant therapy for atrial fibrillation; and beta-blockers for CHD.

***Results: University of East Anglia***

Evidence for lives saved was found for 19 indicators in the 2006 revised contract. A further 23 indicators were indirectly linked to a reduction in mortality, as they are necessary processes to deliver results for the directly linked 19 indicators. The number of potential lives saved by the 19 directly linked indicators ranged from 1.1 to 63.7 per 100,000 people per year.

In the 2003 GMS contract[1] the potential lives saved per 100,000 populations per year aggregated at the domain level are: 163.2 lives in coronary heart disease, 109.5 lives in diabetes, 53.6 lives in hypertension, 44.9 lives in stroke, 27.6 lives in chronic obstructive pulmonary disease, 11.6 lives in left ventricular dysfunction, and 8.8 lives in asthma (Table 2).

In the 2006 revised GMS contract[2] the potential lives saved per 100,000 populations per year aggregated at domain level are: 160.9 lives in coronary heart disease, 107.1 lives in diabetes, 48.2 lives in hypertension, 43.8 lives in stroke, 26.2 lives in chronic kidney disease, 25.0 lives in chronic obstructive pulmonary disease, 21.4 lives in atrial fibrillation, 11.6 lives in heart failure or left ventricular dysfunction and 10.9 lives in smoking cessation.

In the 2003 contract[1] there was potential for 415.77 lives saved per 100,000 per year (400.32-444.99) aggregated across all clinical indicators and domains. In the 2006 contract[2] this increased by 35.73 lives to a potential for 451.5 lives per 100,000 people per year saved (423.98-480.72) aggregated across all clinical indicators and domains.

***Results: University of York***

*Conceptual framework*

- When comparing two treatments, we need to consider the difference in both the quality adjusted survival duration (the health gain) and the health service cost between the two treatments.
- To compare treatment health gains with costs, health gains need to be converted onto a single commensurate scale, such as a monetary value. If the monetary value of a treatment's health gain outweighs the health service cost of the treatment then the treatment can be considered cost-effective, and may be considered for QOF.

- For now, we assume that any implementation payment is a payment per patient treated and is purely an incentive payment to the GP as it does not contribute towards the treatment's cost.
- Some individuals will already be receiving the incentivised treatment before any payment is introduced, but following the introduction of the implementation payment the NHS will receive a lower incremental net monetary benefit for these individuals as the cost for them is increased but there is no extra health benefit. However, for patients who were previously not treated there is the additional health benefit of the treatment but also the additional cost of both the treatment and the implementation payment.
- The maximum value of perfect implementation is the average value of the gains per eligible patient (including those who are already receiving it, for whom there is no gain and also those who are not receiving it for whom the gain is the monetary value of the health gain less the cost). This provides the maximum value for any per patient implementation payment (which would be paid per treated patient and not just per patient who was previously not treated). If the implementation payment exceeds this maximum value, then the average monetary benefit per patient would be negative.
- When implementation payments are made they will not always lead to full implementation (i.e. not all eligible patients will receive the incentivised intervention). However, if the increase in the proportion of patients treated can be estimated then the change in incremental net monetary benefit to the average patient (taking account of the implementation payment) can be calculated and if this is negative the QOF payment would not be cost-effective.
- Extension 1: The QOF payment as costs to GP plus an incentive payment. The QOF payment is intended to incorporate both the cost of the intervention to the GP and an incentive payment. This means that the payment is not a pure incentive payment and cannot be completely separated from the cost of the intervention to the health service. Instead, to avoid double counting, the cost of the intervention to the GP should be calculated and using this we can calculate the minimum (the cost of the intervention to the GP) and maximum (the value of actual implementation plus the cost of the intervention to the GP) payments for a QOF indicator.
- Extension 2: Implementation payment is put towards cost. When the proportion of the implementation payment that is put towards cost is unknown, definitive statements of minimum and maximum payments that would be considered cost-effective cannot be made. However, if the implementation payment (including the

part which will be put towards cost) is less than the actual value of implementation then this is a sufficient condition for the payment being cost-effective.

- Extension 3: Ranking QOF indicators. When considering new indicators, it is possible to examine which are potentially the most cost-effective. This can be done by comparing the value of perfect implementation of different indicators, with the indicators with larger values being potentially more cost-effective, and hence such indicators should be prioritised.
- Extension 4: Ranking QOF indicators when the total number of indicators is constrained. When only a certain number of indicators can be chosen, the use of the patient value of perfect implementation may be insufficient to determine the most cost-effective choice of indicators. Instead the population value of perfect implementation should be considered.

*Payment calculations*

For the 2004/05 analyses, there was a wide variation in the annual QOF per-patient payment: the smallest non-zero payment was £0.13 (for an achievement level of 40 to 45% for ChKD4), and the highest payment was £87.79 (for an achievement level of 60 to 75% for MH5). In the 2006/07 analysis, the annual QOF per-patient payment ranged from £0.22 (for an achievement level of 40 to 45% for ChKD4) to £73.04 (for an achievement level of 75 to 80% for CHD11).

*Literature mapping review*

Data on 335 studies covering 20 indicators were abstracted for the literature mapping review. Following Department of Health advice, studies on four indicators relating to influenza immunisation were accorded a lower priority (N=73). The remaining 262 studies were classified the studies as D (definite; N=22); P (possible; N=129); U (unclear; N=28); or N (not relevant; N=83) and those coded as D or P were ordered or retrieved electronically (N=151). Five papers were unavailable within the project timeframe. The remaining 146 papers were abstracted under the relevant indicator headings, with some papers informing multiple indicators. Economic evaluations providing cost and outcomes data best fitting the indicator were selected for the analyses, supplemented with data from authors where appropriate.

*Characteristics of included indicators*

Most (N=9) of the 16 indicators considered for the analysis were for 'treating' (category 5), with others covering 'measuring' (category 3: N=3), 'offering therapy' (category 4: N=1) or

achieving intermediate outcomes (N=2). No indicator met the criteria for category 1 (structural) or category 2 (diagnostic).

Analysis for two of the 16 indicators (DEP1 and MH5) proved infeasible given an absence of available data. Analysis of two other indicators (CHD8 and DM20) was contingent upon supplementary data from study authors, which were not provided. Of the remaining 12 indicators, 10 were based on evidence that had a publication year of or after 2000, eight were based on studies reporting QALYs as the outcome measure and eight were based on UK evidence. Six of the indicators were based on studies meeting all four criteria, i.e. were published in or after the year 2000, were based in the UK, reported QALY data and addressed the indicator question. Based on the criteria listed above, three indicators were classified as 'robust', two as indicative/robust and seven as 'indicative' only.

Regarding the level of relationship of the indicator evidence to the *indicator population*, none was classed as having 'low', one was classified as 'medium/low', six as 'medium', one as 'medium/high' and six as 'high'. For level of relationship of the indicator evidence to the *indicator intervention*, none was classed as having 'low', one was classified as 'medium/low', five as 'medium', one as 'medium/high' and five as 'high'.

*Indicator cost-effectiveness*

Indicator cost-effectiveness varied by baseline uptake (where baseline uptake is the level of utilisation that would have been experienced if QOF had not been introduced) . Where baseline uptake was 30%, the mean absolute change in utilisation needed for an indicator to be cost-effective was 0.5%. In other words, if, prior to QOF introduction, GP practices were currently meeting the indicator for 30% of their patients, then on average a change of only 0.5 percent would be needed for the indicator to be cost-effective. In the 2004/05 analyses, this value ranged from just over 0% to 2% and as the analysis was constructed using discrete intervals for change in utilisation, the precise value of the minimum change necessary to achieve cost-effectiveness could not be specified. For baseline uptake levels of 60%, the mean absolute change in utilisation needed for an indicator to be cost-effective was 4.6% (range: just over 0% to around 20%). For baseline uptake levels of 90% (i.e. when, prior to QOF, all practices are already meeting the QOF target for 90% of the indicator population), the mean absolute change in utilisation needed for an indicator to be cost-effective was 1.7% (range: just 0.1% to around 7.5%). This figure is lower than the figure for a baseline uptake of 60% because two of the indicators (CHD 11 and SMOKING2) could never be cost-effective under a scenario where baseline uptake was 90% and the estimate of the mean was based on data

from just 10 indicators. However, as they were able to contribute data toward estimate of the mean statistic for a 60% baseline, this estimate was higher.

The 2006/07 QOF incorporated an increase in point value (which would be expected to make indicators less cost-effective), but also a raising of the lower threshold for payment (which would be expected to make indicators more cost-effective at lower levels of utilisation). As these two factors work in different directions, the net effect upon QOF indicators will be variable. Findings from the 2006/07 analysis were broadly similar to those from the 2004/05 analysis, with the magnitude of absolute changes in utilisation relative to baseline being key in determining the cost effectiveness of the indicators. Compared with the 2004/05 analyses, the most noteworthy change in indicator cost-effectiveness from the 2006/07 analysis relates to the DM21 indicator (diabetic retinopathy screening). For any level of baseline uptake and magnitude of absolute change in utilisation, the DM21 indicator is not cost-effective under the 2006/07 payments and thresholds. This contrasts starkly with the cost-effectiveness results under 2004/05 payments and thresholds, where only very small absolute changes in utilisation are required for the indicator payments to be a cost-effective use of resources. However, this finding results from the costs for the intervention being inflated from 2004/05 prices to 2006/07 prices while there is no increase in health benefits; as a result, the net monetary benefit of retinopathy screening is negative and thus the procedure being incentivised is not a cost-effective use of resources *per se* assuming a cost-effectiveness threshold of £20,000 per QALY gained, let alone when a extra payment is given to GPs to incentivised its provision.

With the exception of DM21, when baseline utilisation is 30%, then the minimum absolute change in utilisation required under the 2006/07 QOF is 0%, whereas under the 2004/05 QOF it could be up to 2%. Under the 2006/07 QOF, GP practices only achieve points once they reach a threshold utilisation rate of 40%, whereas under the 2004/05 QOF the minimum threshold utilisation rate was lower (25%) for some indicators. Since no payment is received when achievement is less than 40% and if the treatment targeted by the indicator is cost effective *per se* (i.e. has positive net monetary benefit per patient), then no change is required for the indicator to be cost-effective. When baseline utilisation is 45%, the required absolute change in utilisation in the 2006/07 QOF is generally equal to or lower than the corresponding values for 2004/05. The only exception is DM21, which is not cost-effective under any scenario in the 2006/07 analysis.

For baseline utilisation levels of 60%, 75% and 90% the absolute change in utilisation required for the indicator payment to be a cost-effective use of resources under the 2006/07 QOF is greater than or equal to that required under the 2004/05 QOF. This is due to the

payments per patient being larger under the 2006/07 QOF and also the costs of the interventions being higher due to inflation, thus reducing the net monetary benefit of a specific absolute change in the utilisation rate.

In general, when baseline implementation rates are high, larger absolute changes in utilisation are required for payments to be cost-effective. This is because, at high baseline uptake levels, more patients are already receiving and benefiting from the incentivised treatment but GPs receive extra money for these patients for whom there are no additional health benefits. Hence larger changes in utilisation are needed if payments given to GPs for patients already being treated are to be offset by the monetary value of the additional health gain to patients who were previously untreated.

For a given baseline utilisation rates, the 2006/07 QOF indicators typically require larger QOF-induced absolute changes in utilisation rates compared to the 2004/05 QOF indicators. This is unsurprising given the higher price per point as well as the inflation in health care costs. However, there may be exceptions to this general rule as the cost-effectiveness of hypothetical changes is dependent on the baseline utilisation rate and whether the intervention being assessed is dominant (i.e. has additional health benefits and is cost saving when compared to the practise that will be displaced in its favour).

*Results of the sensitivity analysis*

Two one-way sensitivity analyses were undertaken. QOF payment duration, which was assumed to be five years in the base case analysis, was increased first to 10 years and then to 15 years. Under both the 2004/05 and 2006/07 analyses, for the indicator to be cost-effective at longer payment durations, greater absolute changes in utilisation are required for given baseline uptake levels. However, there are exceptions to this general finding. In the 2004/05 QOF, increasing the duration of payments in two indicators, CHD10 (beta-blockers for post MI patients) and ChKD4 (ACEis for patients with chronic kidney disease), had little effect on the change in utilisation required for the indicator to be cost-effective. In both cases, the indicators remained cost-effective provided at least small (non-zero) changes resulted from the introduction of the QOF payment. In the 2006/07 QOF, findings for CHD10 and ChKD4 were similar to the analyses for 2004/05.

## Discussion

The University of East Anglia notes that the clinical indicators in the new GMS contract have potential for significant health gain, and explain why actual health gain realised is less than

the potential health gain. Additional issues considered include the strengths and limitations of the study and implications for further research.

The University of York considered a number of issues, including selection of topics /indicators; the cost-effectiveness of indicators with a direct therapeutic effect; options for extending the economic analysis; and possible avenues for further research.

***Discussion: University of East Anglia***

The clinical indicators in the new GMS contract have potential for significant health gain. In the original form of the contract (2003) there was potential for 415.77 lives being saved per 100,000 populations in one year. In the revised contract (2006) this raised to 451.5, an increase of 35.73. This difference in the potential to save lives from the original GMS contract to the revised GMS contract 2006/07 was largely due to the inclusion of the Kidney Disease and the Atrial Fibrillation indicators. The greatest number of lives saved by domain was in the CHD and the DM indicators, accounting for more than half of all lives saved across all indicators. These diseases were more common and have a number of clinical interventions that are effective. By indicator, influenza immunization carries the greatest potential for lives saved, followed by treatment for primary prevention for hypertension. Influenza immunization was already incentivised before the QOF was introduced, thus the health gain of the QOF will be smaller than projected. Smoking had less health gain than expected by the authors. There are at least two reasons for this: firstly, smoking cessation therapy in the GMS contract had been targeted only at patients with other clinical conditions - thus the large pool of patients who would benefit from primary prevention was not included in the new GMS contract clinical indicators. Records 22 includes recording of smoking habit in patients over 15 years of age, but does not incentivise smoking cessation. Secondly, patients who had suffered a stroke, heart attack or diabetes had a low prevalence of smoking in the contract data - presumably as many would have already stopped smoking. Smoking cessation for primary prevention would have significantly increased the potential for lives saved if it were included in the new GMS contract.

Actual health gain realised by the introduction of QOF will be less than the potential health gain for several reasons. (1) There was a significant baseline activity in primary care before the implementation of the new GMS contract.[8-10] (2) It is unrealistic to expect 100% of patients on a register to receive indicated care. Patients will appropriately be excluded by exception reporting. With the exception reporting rates in 2006/7, the potential maximum health gain is reduced by 13% to 393.7 lives per 100,000 people saved in one year. (3) Increase in achievement may in part be due to better record keeping, rather than actual

improved performance. (4) In patients with multiple conditions, the beneficial effect of interventions may be less (or more) than additive. (5) Some practices may on occasion game results, rather than actually improving patient care. (6) There is an opportunity cost of the financial spend allocated to QOF. (7) This analysis only reports the effects of QOF on activities within the QOF. The potential effect on activities outside QOF (either positive or negative) is not explored. Finally, (8) there are potential side effects of drugs used in QOF indicators, which will on occasion cause harm. Lives saved were only one measure of primary care quality, however it was an important one and currently is more widely available than other measures such as QALYs. Health gain represented a possible additional criterion to be used when allocating points to indicators and conditions in future revisions of the QOF.

*Strengths of the Study*

Our study reviewed four robust sources of evidence of which three are regularly updated online. These sources were well known sources of evidence. The evidence was independently searched by two researchers. (RF and SP) The independent search increased the robustness of the evidence sought. The included studies were then quality assessed using the quality grading scale designed by AHCPR[7] which has been adopted by the Cochrane Library.[6] Wherever possible we selected the highest level of evidence. Secondly, the measure of health gain chosen was all cause mortality. This measure of health gain was widely available and was a measure of quality also used by the World Health Organization and the Department of Health in performance indicators. Mortality was a relevant and important outcome for many interventions in the new GMS contract.

*Limitations and weaknesses*

The limitations specific to each individual indicator have been mentioned in each individual caveats section. There were other limitations general to all indicators in the new GMS contract. Firstly the measure of quality used (lives saved in one year) was narrow. This measure has direct evidence for 19 indicators and are linked to a further 23 indicators, therefore they only cover 42 out of the 80 clinical indicators. There were many aspects of primary care quality which cannot be captured by lives saved, for example access to services and palliative care. Unfortunately there was as yet sufficient evidence in terms of QALYs across the clinical spectrum covered by the new GMS contract. There are other benefits which would not be captured by either measure, such as improved recording of diseases and prevalence. Secondly trial participants often did not represent the characteristic of patients who were present in general practice. This was most marked in the evidence cited for CKD3, where patients from the evidence base included some patients with other chronic diseases. Thirdly the level of evidence for some indicators was less than grade 1a. For example

evidence for the effectiveness for influenza immunization was based on the evidence from cohort studies, which was subject to selection bias. However it would not have been possible to design RCTs effectively for interventions such as influenza immunization, due to ethical issues of the existing evidence and the beneficial effect of herd immunity on patients who were not immunized. There was a lack of research into important areas for some of the indicators, such as cholesterol and diabetes. We identified only one study for this indicator which was terminated early and therefore the study did not have appropriate power to detect a mortality difference if one existed. Finally, many diseases are undergoing a change in prevalence, with for example coronary heart disease falling and diabetes increasing. Changes in disease prevalence would alter the potential population health gain for particular indicators and the nature and pathogenicity of the diseases themselves may change over time. For example some of the increase in prevalence of type 2 diabetes may be due to early diagnosis (and therefore a lower baseline risk in terms of morbidity and mortality).

*Further Implications for Research*

Further research is needed in areas of the new GMS contract indicators to complete the evidence base for lives saved. This specific study addressed only the indicators mentioned in the GMS contract, and a systematic approach to evaluating evidence for lives saved across the whole spectrum of primary care will give a more complete picture of potential for lives saved. More research is needed across the spectrum of primary care to evaluate other measures of health gain, such as QALYs. This will enable the development of GP contract to include the new indicators to maximize health gain. Quality can be used as one of a number of factors (including baseline performance and difficulty of clinical tasks) in determining the ideal financial incentive for each indicator.

***Discussion: University of York***

*Selection of topics*

Our analysis focused on indicators with a direct therapeutic benefit because we expected there to be little relevant evidence on other types of indicators. We therefore explored the potential cost-effectiveness of only a small subset of QOF indicators. Although there is good reason to believe that evidence on the costs and effects of other indicators is likely to be very limited, we have only formally confirmed this for two indicators (DEP1 and MH5). However, the absence of relevant evidence does not imply that these indicators are not cost-effective.

Structural, diagnostic and measurement indicators would be particularly difficult to address because typically no therapeutic intervention is defined. For example, setting up a register of

patients with a given disease does not define what intervention(s), if any, will ultimately be delivered. This is an issue that could, in principle, be explored by a *de novo* prospective study, of an experimental or quasi-experimental nature, where a proportion of patients/practices received the change and others acted as controls.

In some clinical domains there may be no therapeutic indicators. It could be argued that, in this situation, adding a therapeutic indicator might help assess the impact of some of the non-therapeutic ones as it would provide a clue as to their ultimate clinical implications. However, there is a risk that the therapeutic target could become too narrow, hindering GPs' attempts to adapt treatment to particular patients' needs. While non-therapeutic targets may reflect genuine quality of care, their implications for costs and outcomes are difficult to measure.

*Cost-effectiveness of indicators with a direct therapeutic effect*

On the basis of the available evidence on the costs and benefits of a given therapeutic intervention, relative to one or more comparators, it has been possible to show what absolute change in the utilisation would be necessary for an indicator to be considered cost-effective given the QOF payment. This absolute change is conditional on the baseline utilisation (the level of utilisation if QOF had not been introduced (the counter-factual)) because, if this baseline utilisation is high, the additional cost of QOF payment can be spread over fewer additional patients and there is less scope for the indicator to be cost-effective. For those indicators where some evidence on costs and benefits was available, all interventions under the 2004/05 QOF were shown to be cost-effective, whereas just one indicator in the 2006/07 analysis was not cost-effective. Whether the QOF payments themselves represent potential value for money depends on baseline utilisation and change in utilisation following the introduction of the QOF payments. Although some of the necessary changes in utilisation to make the QOF potentially cost-effective appear modest, even these changes need to be demonstrated empirically.

However, there are a number of important caveats. Firstly, for many indicators, the cost and QALY evidence from the literature is subject to major uncertainties. In addition, some studies relate to non-UK patients and practice, and it is not clear how generalisable this evidence is for the UK. There are also analytical uncertainties: for example, the discount rates used in many studies were different to those currently recommended for the UK. These uncertainties are shown in Table 3 where only two of the indicators had estimates of cost-effectiveness that were considered robust. The evidence presented here should therefore be interpreted with considerable caution.

The second caveat relates to the assumed duration of the QOF payments. From a policy perspective, there is no indication of the period of time over which these payments will be paid. As our base-case, we have assumed a five-year payment duration. Sensitivity analyses are also presented for 10 and 15 years. A third caveat relates to the uncertainty regarding what the QOF payments were intended to reflect. In principle, the payments could provide a 'pure incentive' payment with the assumption that the resource implications of providing the intervention are covered from other sources. Alternatively, the payment could be entirely concerned with covering the cost of the intervention with no incentive element intended. A third option is that the payment reflects a mix of covering cost and providing an incentive but in unknown proportions. The base-case analyses presented here all assume that payments wholly represent an incentive payment. As such, this represents a 'worst case scenario' for the cost-effectiveness of the indicators as 100 percent of the QOF payment is additional to the resource cost of the intervention and needs to be justified by the health gains the intervention achieves. As a sensitivity analysis, a 'best case scenario' is explored for selected indicators where the QOF payment is assumed wholly to relate to the cost of the intervention.

A fourth caveat is that the cost-effectiveness threshold used in this report of £20,000 per QALY and may not be appropriate. If the threshold used is inappropriate the estimates may under or over value the net monetary benefits of the interventions which will have consequences for the changes in utilisations required for cost-effectiveness at given baseline utilisation levels. The value of the threshold is an empirical question; however, given the current lack of direct evidence, and the use of this threshold by NICE we consider the threshold of £20,000 per QALY to be the most appropriate threshold for these analyses.

*Extending the economic analysis*

It needs to be emphasized that the analysis presented in this report relates to the *potential* cost-effectiveness of the QOF indicators: it shows, for a series of baseline levels of uptake, the magnitude of change in utilisation of a given intervention as a result of QOF that needs to be achieved in order for the payment to be considered good value.

From a policy perspective, it may be attractive for the analysis to to show how the existing indicators could be improved in terms of their points and threshold levels. However, based on existing evidence, the only improvement that could be recommended would relate to any indicators that incentivise interventions that are clearly not cost-effective and probably should not be used in the NHS. No such indicator was found in the 2004/05 analysis, and just one indicator was identified in the 2006/07 analysis (however, this result should be interpreted

with caution as it is based on an extrapolation of the result based on several assumptions which may not hold). The reason why this indicator was no longer cost-effective under the 2006/07 QOF was due to the balance of costs and benefits for the intervention, rather than to the construction of the incentive system. The finding highlights problem of incentivising interventions that are of borderline cost effectiveness. To improve existing indicators that incentivise cost-effective treatments, changing the points and threshold levels would require evidence on the sensitivity of utilisation of an indicator in response to a change in payment. No such evidence currently exists. In terms of the duration for QOF payments, the 2006/07 analysis showed that if payment levels are maintained long-term without adjusting these for observed changes in utilisation, the incentive payment may not always be cost-effective in practice. Therefore, actual changes in utilisation should be monitored and a flexible approach to payment levels adopted in order to ensure that value for money is preserved.

What about interventions that are currently outside the QOF but may be candidates to be so in the future? There is a strong case to use the economic framework presented here to assess the potential value of possible additions to the QOF. In part, these candidate interventions might be identified through the various activities of the National Institute for Health and Clinical Excellence such as clinical guidelines, technology appraisal and the development of public health interventions and programmes. If suitable evidence exists, the potential value of these interventions could be assessed, using assumed QOF payments in the same way that existing indicators are considered in this report. Some evidence of how the QOF payment generally affects uptake would be the preferred basis for an extension of the QOF.

*Further research*

There are a number of options for further research. The first option relates to the evidence base associated with the costs and benefits of the interventions within the QOF (or candidates for future inclusion). To provide robust evidence, additional primary data collection would be necessary for many indicators, most notably those relating to structural, diagnostic or measurement activities. A range of studies could be considered depending on the existing uncertainties. In terms of resource implications, these might range from small-scale surveys of clinical practice (e.g. to establish the range of therapeutic interventions by GPs that stem from the existence of a structural change or information from a diagnostic or measurement intervention) to large-scale cluster randomised trials.

There should also be a role for more extensive synthesis and modelling based on existing evidence. Given the time constraints of this project, we have sought evidence on the costs and benefits of interventions from existing economic evaluation studies. With more time, it

may be possible to assess the cost-effectiveness of some indicators not addressed here or to improve the estimates for those that have been considered. This could be done by developing cost-effectiveness models based on existing clinical and epidemiological evidence together with appropriately elicited expert opinions. This work may be particularly worthwhile for those indicators commanding a relatively large number of QOF points and for which cost-effectiveness is most uncertain. This work would be a useful precursor to more extensive primary data collection by highlighting those where, based on existing evidence, the uncertainty is extensive and potentially highly costly.

The precision of the potential cost-effectiveness estimates presented here has not been assessed. That is, the focus of the work has been on *mean* costs and effects, and the precision of those estimates has not been formally considered. It may be that further research is justified even for those indicators for which the cost-effectiveness estimates here are viewed as robust because the precision of those estimates is low. A framework to prioritise additional research to increase the precision of cost-effectiveness has recently been proposed[11] and would be a valuable guide in identifying, prioritising and designing additional research of this type.

## Conclusions

The UEA analysis suggests that full implementation of the quality indicators in the original GMS contract (2003)[1] can be expected to result in a potential 415.77 lives being saved per 100,000 populations per year. However the actual number of lives saved was lower for two reasons. Firstly, there was a significant baseline activity in primary care before the implementation of the new GMS contract.[8-10] Secondly, less than 100% of the target population received the intervention due to exclusions due to exception reporting and less than full implementation of the contract. Full implementation of the revised GMS contract (2006)[2] increases the potential lives saved to 451.5 per 100,000 populations per year. This equates to approximately 270,155 lives saved in one year for the whole of the United Kingdom population.

Lives saved are a direct outcome of 19 of the 2006 GMS contract indicators, and an indirect outcome of a further 23 indicators. Health gain represents a possible additional criterion to be used when allocating points to indicators and conditions in future revisions of the Quality Outcomes Framework (QOF). Limitations of this study include the use of mortality as a relatively narrow measure of health gain, the assumption of linearity between mortality reduction and time, an incomplete evidence base for some interventions in the GMS contract,

the changing prevalence of disease and the assumption that patients in the clinical trials are representative of the UK population. Further research includes developing other measures of quality such as QALYs[12] for as many clinical indicators as possible to inform development and weighting of both current and new indicators in future revisions of the GMS contract.

**Conclusions and recommendations**

The University of York presented a conceptual framework to assess the cost-effectiveness of QOF indicators in a way that reflects the cost of the intervention, the health effects it achieves and the cost of the incentive payment. Based on available evidence on the cost-effectiveness of those interventions with a direct therapeutic effect, it was possible to estimate the potential cost-effectiveness of twelve of these indicators. This is shown, for a given baseline utilisation, in terms of the absolute change in uptake necessary to justify the QOF payment. Once evidence emerges on the actual change in utilisation achieved by QOF, the potential cost-effectiveness of these indicators can be translated into actual changes.

**Table 1: Overview of findings: UEA and York assessments**

| Indicator | Intervention | UEA | | | York | | | |
|---|---|---|---|---|---|---|---|---|
| | | Assessed | Lives saved per 100,000 population p.a. *2003* | Lives saved per 100,000 population p.a. *2006* | Assessed | Incremental QALYs per patient treated *2006* | Incremental costs per patient treated *2006* | Net monetary benefit of indicator treatment per patient treated *2006* |
| AF3 | Anticoagulant | Y | N/A | 21.4 | Y | 1.465 to 2.2 | -£1,162 to -£16,922 | £45,162 to £46,222 |
| Asthma5 | Smoking cessation advice/referral | Y | 8.8 | See Smoking2 | N | See Smoking2 | See Smoking2 | See Smoking2 |
| BP3 | Smoking cessation advice/referral | Y | 5.4 | See Smoking2 | N | See Smoking2 | See Smoking2 | See Smoking2 |
| BP5 | Hypertension, BP,150/90 in past 9 months | Y | 48.2 | 48.2 | Y | 0.7 | £751 | £13,249 |
| CHD4 | Smoking cessation advice/referral | Y | 2.4 | See Smoking2 | N | See Smoking2 | See Smoking2 | See Smoking2 |
| CHD6 | BP<150/90 | Y | 11.3 | 11.3 | N | N/A | N/A | N/A |
| CHD8 | Cholesterol < 5 mmol | Y | 15.8 | 15.8 | Y | Not estimable | Not estimable | Not estimable |
| CHD9 | Aspirin | Y | 24.8 | 24.8 | Y | 0.0066 | -£30 | £162 |
| CHD10 | Beta blocker | Y | 45.9 | 45.9 | Y | 1.89 | £234 | £37,566 |
| CHD11 | ACE/ARB | Y | 1.5 | 1.5 | Y | 0.08 | £488 | £1112 |
| CHD12 | Influenza immunization | Y | 61.6 | 61.6 | N | N/A | N/A | N/A |
| ChKD3 | BP<140/85 | Y | N/A | 26.2 | N | N/A | N/A | N/A |
| ChKD4 | ACE/ARB | N | N/A | N/A | Y | 0.8076 to 1.5308 | -£31,811 to -£32,906 | £49,058 to £62,427 |
| CS1 | Cervical screening | N | N/A | N/A | Y | 0.137 | £68 | £2,672 |
| COPD5 | Smoking cessation advice/referral | Y | 2.6 | See Smoking2 | N | See Smoking2 | See Smoking2 | See Smoking2 |
| COPD8 | Influenza immunization | Y | 25.0 | 25.0 | N | N/A | N/A | N/A |
| DEP1 | Case finding for depression | N | N/A | N/A | Y | Not estimable | Not estimable | Not estimable |
| DM4 | Smoking cessation advice/referral | Y | 2.4 | See Smoking2 | N | See Smoking2 | See Smoking2 | See Smoking2 |
| DM6/DM20 | HbA1c<7.4 | Y | 26.5 | 26.5 | N | Not estimable | Not estimable | Not estimable |
| DM7 | HbA1c <10 | Y | 7.4 | 7.4 | N | N/A | N/A | N/A |
| DM12 | BP<145/85 | Y | 13.5 | 13.5 | N | N/A | N/A | N/A |
| DM15 | Proteinuria/ microalbuminuria on ACE | Y | 3.4 | 3.4 | Y | 0.7210 | -£9,662 | £2,4081 |
| DM18 | Influenza immunization | Y | 63.7 | 63.7 | N | N/A | N/A | N/A |
| DM21 | Diabetic retinal screening | N | N/A | N/A | Y | 0.4865 | £9,750 | -£21 |
| LVD/HF3 | ACE/ARB | Y | 11.6 | 11.6 | Y | 0.21 | £25 | £4175 |
| MH5 | Lithium | N | N/A | N/A | Y | Not estimable | Not estimable | Not estimable |
| Smoking2 | Smoking cessation advice/referral | Y | In domains | 10.9 | Y | 0.0157 to 0.0451 | £11 to £90 | £303 to £812 |
| Stroke4 | Smoking cessation advice/referral | Y | 1.1 | See Smoking2 | N | See Smoking2 | See Smoking2 | See Smoking2 |
| Stroke9/Stroke12 | Antiplatelet/ anticoagulant | Y | 15.8 | 15.8 | Y | 0.17 | £371 | £3029 |

| Indicator | Intervention | UEA | | | York | | | |
|---|---|---|---|---|---|---|---|---|
| | | Assessed | Lives saved per 100,000 population p.a. *2003* | Lives saved per 100,000 population p.a. *2006* | Assessed | Incremental QALYs per patient treated *2006* | Incremental costs per patient treated *2006* | Net monetary benefit of indicator treatment per patient treated *2006* |
| Stroke10 | Influenza immunization | Y | 28.1 | 28.1 | N | N/A | N/A | N/A |
| **Total, adjusted for double counting** | | **Y** | **415.8** | **451.5** | **N** | N/A | N/A | N/A |

Notes:

1. UEA comparisons are relative to placebo; York comparisons are relative to usual care
2. AF3: York ranges are with and without co-morbidities (all patients aged 65)
3. ChKD4: York ranges are for non-diabetic vs. diabetic patients
4. Smoking2: York ranges are for brief advice only vs. brief advice +NRT

**Table 2: University of East Anglia. Health gain league table, estimated lives saved per 100,000 populations per year by clinical domain for 2003[1] and 2006 contract[2]**

| Clinical domain | 2003 contract: lives saved | 2006 contract: lives saved |
|---|---|---|
| Coronary heart disease | 163.2 | 160.9 |
| Diabetes | 109.5 | 107.1 |
| Hypertension | 53.6 | 48.2 |
| Stroke | 44.9 | 43.8 |
| Chronic obstructive airways disease | 27.6 | 25.0 |
| Asthma | 8.8 | NA |
| Heart failure | 11.6 | 11.6 |
| Atrial fibrillation | NA | 21.4 |
| Chronic kidney disease | NA | 26.2 |
| Smoking | (included in other domains) | 10.9 |

**Table 3: University of York: summary of cost-effectiveness results, by indicator; analyses employing 2004/05 QOF points and threshold values**

| Indicator | Category[a] | Robustness[b] | Approximate absolute change in utilisation from baseline required for indicator to be cost-effective[c] | | | UK-based | QALY-based | Post 2000 | Relationship to indicator population | Relationship to indicator intervention | Key additional analyses[e] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Baseline: 30% | Baseline: 60% | Baseline: 90% | | | | | | |
| AF3 - 65-year-old with co-morbidities | 5 | Indicative | 0% | 0.1% | 0.2% | X | √ | √ | Medium | High | Intervention is less cost-effective in 85-year-old patients with same co-morbidities |
| AF3 - all 65-year-olds | 5 | Indicative | 0% | 0.1% | 0.2% | X | √ | √ | Medium | High | Intervention is less cost-effective in 85-year-olds |
| BP5 | 6 | Indicative | 0.1% | 0.2% | 0.3% | √ | √ | √ | Medium | Medium | Beta blockers and diuretics least cost-effective options, particularly when annual risk of diabetes is high |
| CHD8[f] | 6 | | | | | | | | | | |
| CHD9 | 5 | Robust | 0.6% | 4.5% | 7.5% | √ | √ | √ | Medium | High | |
| CHD10 | 5 | Indicative | 0.1% | 0.1% | 0.1% | √ | X | X | Medium | Medium /Low | |
| CHD11 | 5 | Robust | 2.0% | 20.0% | never cost-effective | √ | √ | √ | High | Medium | |
| ChKD4 - Non-diabetics | 5 | Indicative | 0% | 0.1% | 0.1% | X | X | √/X | High | High | |
| ChKD4 - Diabetics | 5 | Indicative | 0% | 0.1% | 0.1% | X | X | X | Medium | Medium | |
| CS1 | 3 | Indicative | >0.0% | >0.0% | >0.0% | √ | X | √ | Medium/High | Medium/High | When impact on quality of life is taken into account, screening intervals of 3 years or under are not cost-effective |
| DEP1[d] | 3 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| DM15 | 5 | Indicative | 0.1% | 0.3% | 0.4% | X | X | √ | Medium | Medium | |
| DM20[f] | 5 | | | | | | | | | | |

| Indicator | Category[a] | Robustness[b] | Approximate absolute change in utilisation from baseline required for indicator to be cost-effective[c] | | | UK-based | QALY-based | Post 2000 | Relationship to indicator population | Relationship to indicator intervention | Key additional analyses[e] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Baseline: 30% | Baseline: 60% | Baseline: 90% | | | | | | |
| DM21 | 3 | Indicative | 0.2% | 1.5% | 2.5% | √/X | √ | √/X | Medium | Medium | When screening every five years is used as the comparator, the intervention is less cost-effective |
| HF3 | 5 | Indicative | 0.4% | 3% | 3.5% | X | √ | X | High | Medium | |
| MH5[d] | 6 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| SMOKING2 | 4 | Indicative | 1.5% | 20% | never cost-effective | √ | √ | √ | Medium/Low | High | Indicator more cost-effective for younger patients and for NRT + brief advice vs. brief advice alone |
| STROKE12 | 5 | Robust | >0.0% | 0.5% | 0.9% | √ | √ | √ | High | High | Cost-effectiveness increases when treatment effects on non-vascular deaths taken into account |

[a] 1 =structural; 2 =diagnostic; 3=measuring; 4=offer therapy; 5=give therapy; 6=intermediate outcome
[b] For definition, see section 3.2
[c] Results for 'base case' analysis only (see text under each indicator for details); estimates are approximate, as analysis uses bandings to estimate required change in utilisation.
[d] Analyses for DEP1 and MH5 were not feasible
[e] Only those additional analyses which have a major effect on results are detailed
[f] Analyses for CHD8 and DM20 are pending supplementary data from authors

## References


1. Department of Health. Department of Health. Investing in General Practice: The New General Medical Services Contract London. 2003.
2. Hakin BMH. *Revisions to the GMS contract 2006/07*. England: NHS Employers, 2006.
3. The Information Centre for Health and Social Care. Available at http://www.ic.nhs.uk/statistics-and-data-collections/audits-and-performance/qof Accessed 18/05/07. 2006.
4. National Institute for Health and Clinical Excellence. NICE Guidance. 2007; Available at http://guidance.nice.org.uk Accessed in 2006/07. 2007.
5. BMJ. BMJ Clinical Evidence 2007; Available at: http://clinicalevidence.com.ceweb/resources/useful_links.jsp Accessed 2006/07. 2007.
6. Cochrane Collaboration. The Cochrane Library 2007; Available at http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME?CRETRY=1&SRETRY=0 Accessed in 2006/07. 2007.
7. Agency for Health Care Policy and Research. Acute pain management: operative or medical procedures and trauma, Part 2. *Clin Pharm* 1992;11(5):391-414.
8. Fleetcroft R, Cookson R. Do the incentive payments in the new NHS contract for primary care reflect likely population health gains? *Journal of Health Services & Research Policy* 2006;11(1):27-31.
9. Steel N, Maisey S, Clark A, Fleetcroft R, Howe A. Quality of clinical primary care and targeted incentive payments: an observational study. *Br J Gen Pract* 2007;57(539):449-54.
10. Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. Quality of primary care in England with the introduction of pay for performance. *New England Journal of Medicine* 2007;357(2):181-90.
11. Fenwick E, Claxton K, Sculpher M. The value of implementation and the value of information: combined and uneven development. *Medical Decision Making* 2008;28(1):21-32.
12. Bandolier. "Evidence based thinking about health care" 2007. Available at http://www.jr2.ox.ac.uk/bandolier/band24/b24-7.html. 2007.