

How to identify when a performance indicator has run its course

In April 2011 eight clinical indicators will be removed from the UK Quality and Outcomes Framework. **David Reeves and colleagues** explain why they were chosen and suggest a rationale for future decisions

Increasing numbers of countries are using indicators to evaluate the quality of clinical care, with some linking payment to achievement.¹ For performance frameworks to remain effective the indicators need to be regularly reviewed. The frameworks cannot cover all clinical areas, and achievement on chosen indicators will eventually reach a ceiling beyond which further improvement is not feasible.^{2,3} However, there has been little work on how to select indicators for replacement. The Department of Health decided in 2008 that it would regularly replace indicators in the national primary care pay for performance scheme, the Quality and Outcomes Framework,⁴ making a rigorous approach to removal a priority. We draw on our previous work on pay for performance^{5,6} and our current work advising the National Institute for Health and Clinical Excellence (NICE) on the Quality and Outcomes Framework to suggest what should be considered when planning to remove indicators from a clinical performance framework.

First UK decisions

The Quality and Outcomes Framework currently includes 134 indicators for which general practices can earn up to a total of 1000 points. Negotiations between the Department of Health and the BMA's General Practitioners Committee last autumn led to an agreement to remove eight clinical indicators worth 28 points in April 2011 (table 1). The eight indicators are



all process measures and reward actions such as taking blood pressure or taking blood to measure cholesterol, glucose, or creatinine concentrations for people with relevant chronic diseases. The framework rewards the action itself rather than a clinically informed response to results or intermediate outcomes such as better control of blood pressure or cholesterol levels. It is therefore not surprising that achievement of these process indicators is high (median >95% and interquartile range <4.5%) with little change in

rates or variation across practices since 2005-6, the second year of the Quality and Outcomes Framework.

In many schemes, including the Quality and Outcomes Framework, providers can "except" certain patients from inclusion in the denominator figures for an indicator on grounds such as extreme frailty or contraindications to a specified drug. Exception reporting rates are also low for these eight indicators (median <5% and interquartile range <3%).

National achievement and exception rates for indicators that are to be removed from the Quality and Outcomes Framework in 2011⁷

Indicator (measurement of)	Condition	Median (interquartile range) achievement (%)			Median (interquartile range) rate of exceptions (%)			Paired indicator*
		2005-6	2006-7	2007-8	2005-6	2006-7	2007-8	
Blood pressure	Coronary heart disease	98.2 (96.7-99.3)	98.5 (97.2-99.5)	98.4 (97.1-99.4)	0.8 (0-1.8)	0.8 (0-1.7)	0.7 (0-1.6)	Yes
Haemoglobin A _{1c}	Diabetes	97.4 (95.1-98.8)	97.8 (95.9-99.0)	97.7 (96.0-98.9)	2.7 (1.4-4.4)	2.5 (1.3-4.2)	2.4 (1.2-3.9)	Yes
Blood pressure	Diabetes	98.8 (97.6-99.7)	99.0 (98.0-100)	98.9 (97.9-99.6)	1.1 (0.3-2.2)	1.0 (0.3-2.1)	1.0 (0.3-2.1)	Yes
Serum creatinine	Diabetes	96.7 (94.1-98.3)	97.4 (95.3-98.7)	97.4 (95.6-98.7)	1.9 (0.9-3.4)	1.7 (0.8-3.2)	1.6 (0.7-3.1)	No
Total cholesterol	Diabetes	96.4 (93.8-98.1)	96.9 (94.8-98.3)	96.8 (94.8-98.2)	2.2 (1.1-3.9)	2.1 (1.0-3.8)	2.0 (0.9-3.6)	Yes
Serum creatinine and thyroid stimulating hormone	Mental health (lithium)	100 (100-100)	100 (100-100)	100 (100-100)	0 (0-0)	0 (0-0)	0 (0-0)	No
Blood pressure	Stroke	97.4 (95.1-100)	97.7 (95.8-100)	97.6 (95.8-99.5)	1.3 (0-3.3)	1.2 (0-2.9)	1.1 (0-2.7)	Yes
Thyroid function	Hypothyroidism	96.8 (94.5-98.7)	96.7 (94.5-98.5)	96.5 (94.3-98.3)	0 (0-0.9)	0 (0-0.9)	0 (0-0.8)	No

*Paired indicators relate to control of the relevant measure—for example, the indicator that focuses on recording blood pressure in patients with coronary heart disease, is paired with another indicator that rewards on the basis of the proportion of patients whose last blood pressure reading was $\leq 150/90$ mm Hg.

Ultimately, we need to know what will happen to performance if an indicator is replaced. Empirical evidence from the United Kingdom is limited

If we look at one of the eight indicators in more detail—the proportion of diabetic patients who have had their blood pressure measured in the previous 15 months—the reason for removal is clear. Performance has been extremely high and stable since 2005-6 both in terms of achievement (median around 99%) and exception reporting (around 1%), with low interpractice variation (table). Indeed, over 99% of practices scored maximum points (21% had 100% achievement), and the average remuneration for practices on this indicator was £374.40 (£3.1m nationally each year). These results strongly suggest that the ceiling has been reached in performance for this indicator and little can be gained from continuing to reward it.

However, the associated intermediate outcome indicator (the proportion of people with diabetes with blood pressure $\leq 145/85$ mm Hg in the previous 15 months) will remain in the 2011 framework. Although there have been moderate gains in performance since 2005-6, median achievement and exception rates in 2007-8 were 80% and 5.7% respectively.⁷ It may not be possible to reach the same level of performance for intermediate outcome indicators as for process indicators; however, around 10% of practices have attained achievement rates of $\geq 90\%$ and exception rates of $< 2\%$, showing that higher performance is possible. It would therefore be inappropriate to remove this indicator.

Criteria for removing indicators

Indicators that are candidates for removal from a framework should be identified largely on the basis of statistical criteria, with the final decision often determined by the context. Statistical criteria consider measures of performance as well as the economics of incentives. Economic analysis considers the net benefit of incentives by quantifying the costs of the indicator relative to the health benefits accrued. If the benefits outweigh the costs it is economically justifiable to continue to reward good performance. This approach is particularly suited to indicators that are associated with a direct therapeutic benefit but is less suited to process indicators such as measuring blood pressure, where it may be difficult to

attribute or quantify any resulting health benefit. We therefore suggest that economic analysis is not routinely used for process indicators.

The performance of an indicator should be assessed in at least five ways:

- Average rate of achievement
- Recent trend in achievement rate
- Extent and trend in variation of achievement rate
- Average rate and trend in exception reporting
- Extent and trend in variation of exception rate.

If the rates have skewed distributions, medians and interquartile ranges may be more appropriate measures than means and standard deviations.

The average reported achievement rate—the percentage of eligible patients for whom the indicator target has been achieved—should be

high. It is difficult to set one definition of high for all indicators because some, particularly intermediate outcomes, such as achieving low cholesterol levels, are unlikely to ever reach as high rates as, say, process indicators. One solution is to use a different empirical definition of high for each indicator—for example, by using the

achievement rates of the top 10% during the first year of the indicator's operation.

Examination of trends in performance can help identify indicators that have reached the limits of achievement. This is signalled either by consistently high performance or by a period of growth followed by a plateauing of the curve. Indicators for which improvement shows no signs of flattening off are less likely to be candidates for replacement. A variable pattern of improvement may signal a wider problem. A lack of change when there is clear room for improvement suggests a substantial mismatch between the magnitude of the incentive and the workload required.

Achievement rates will depend on a range of factors, only some of which will be under the control of providers.⁸ When factors outside providers' control have been allowed for, variation in achievement rates should be low. A wide variation in achievement suggests that many providers could substantially improve their performance.

Average rates of exceptions and variations

in these rates should be low. It would be inappropriate to replace an indicator for which a large proportion of patients with the condition are excepted without first determining the reason for the high level of exceptions, including the possibility that the indicator had poor face validity and was not seen as useful in clinical practice. Indeed, one of the next tasks of the external contractor team employed by NICE to help develop the framework is to look in detail at indicators with high exception reporting as well as high achievement. One such indicator is the percentage of patients with newly diagnosed angina who are referred for exercise testing or specialist assessment. What constitutes a low exception rate may vary by indicator, but a good indication that a practical limit has been reached is low variation in the rate between practices.

Even if an indicator satisfies the statistical criteria for removal, contextual factors, which consider the wider framework in which the indicator is operating, may make removal inappropriate. Contextual factors include policy considerations such as maintaining an appropriate balance of indicators across disease domains; stakeholder perspectives such as concerns of health professionals about additional workload and reliance on incentives⁹; and concerns of patients and user groups about perceived loss of prioritisation of their condition. Indeed, the public seems to perceive inclusion in the framework to be important for good care. When the Department of Health invited ideas for inclusion in the framework in 2007, for example, 153 were received in five weeks, 52% of which came from national disease societies or local patient groups.

Another consideration is circumstances affecting the validity of the indicator, such as changes in evidence. For example, the framework currently contains an indicator for the percentage of patients taking lithium who have a record of plasma concentrations in the therapeutic range within the past six months. However, a recent National Patient Safety Agency alert reported 567 dosage errors for lithium, five of which occurred in primary care.¹⁰ The indicator may need to be reviewed in the light of this and 2006 NICE guidance that lithium concentrations should “normally” be measured every three months.¹¹

Consequences of replacement

We have proposed selecting indicators for replacement on the basis of each indicator's recent history of achievement and exception



reporting rates. Underlying this approach is an assumption that this provides a good guide to the future performance of that indicator. The approach, however, provides no actual estimates of future performance, nor any measure of the degree of uncertainty in the forecast. This is something that we plan to explore in the near future.

Ultimately, we need to know what will happen to performance if an indicator is replaced. Empirical evidence from the United Kingdom is limited. There is some conflicting evidence from performance on two indicators removed for contextual reasons from the Quality and Outcomes Framework in 2006. In a sample of 150 practices in which performance was tracked, immunisation of asthmatic patients against influenza showed a substantial reduction in achievement rates after it was removed from the framework, but there was no such reduction for checking lithium concentrations in patients taking the drug. This process indicator, however, was paired with an intermediate outcome indicator (the percentage of patients taking lithium with a record of lithium levels in the therapeutic range within the past six months), which remained in the framework.

Strategies to minimise the risk of harm from removal might include a gradual reduction of the payments for achieving indicators or, as in the above example, initially to remove half of a paired indicator, so that the removed process is still incorporated as part of a linked intermediate outcome indicator.

Finally, removed indicators need to be monitored within the framework. A new centrally managed tool for extracting the necessary data from practices' clinical computing systems—the General Practice Extraction Service—is due to be operational in England by 2011.¹² In the meantime, large scale general practice databases that allow interrogation of electronically captured patient consultation data, such as the General Practice Research Database, QRESEARCH, and the Health Improvement Network, could be used to identify general trends.

David Reeves senior research fellow

Tim Doran clinical research fellow

Jose M Valderas clinical lecturer

Evangelos Kontopantelis research associate, National Primary Care Research and Development Centre, Manchester M13 9PL

Paul Trueman director, York health economics consortium, York Health Economics Consortium, University of York, York YO10 5NH

Matt Sutton professor of health economics, Health Methodology Research Group, Manchester M13 9PL

Stephen Campbell senior research fellow, National Primary Care Research and Development Centre, Manchester M13 9PL
Helen Lester professor of primary care, National Institute for Health Research School for Primary Care Research, Manchester M13 9PL

Correspondence to: H Lester
Helen.lester@manchester.ac.uk

Accepted: 16 February 2010

Contributors and sources: The ideas in this paper arose from thinking, researching, and reading around the experience of developing the Quality and Outcomes Framework. All the authors contributed initial ideas and to the serial drafts and agreed the final submission. HL is guarantor.

Competing interests: All authors have completed the unified competing interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare (1) HL, TD, SC, DR, EK, MS, and PT have an external contract with NICE for work on removing indicators from the Quality and Outcomes Framework; the views expressed are those of the authors and do not necessarily represent the views of NICE or its independent QOF Advisory Committee. (2) No financial relationships with commercial entities that might have an interest in the submitted work; (3) No spouses, partners, or children with relationships with commercial entities that might have an interest in the submitted work; (4) No non-financial interests that may be relevant to the submitted work.

Provenance and peer review: Not commissioned; externally peer reviewed.

- Lindenauer PK, Remus D, Roman S, Rothberg M, Benjamin E, Ma A, et al. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med* 2007;356:486-96.
- Ganz D, Wenger NS, Roth C, Kamberg CJ, Chang JT, MacLean CH, et al. The effect of a quality improvement initiative on the quality of other aspects of health care: the law of unintended consequences? *Med Care* 2007;45:8-18.
- Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Short and long-term effects of pay-for-performance on the quality of English primary care: interrupted time series analysis. *N Engl J Med* 2009;361:368-78.
- Department of Health. *Developing the Quality and Outcomes Framework: proposal for a new, independent process*. London: DH, 2008.
- 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. *N Engl J Med* 2007;357:181-90.
- Lester HE, Majeed A. The future of the Quality and Outcomes Framework. *BMJ* 2009;338:5-6.
- NHS Information Centre. Quality and Outcomes Framework. www.ic.nhs.uk/statistics-and-data-collections/audits-and-performance/the-quality-and-outcomes.
- Giuffrida A, Gravelle H, Roland M. Measuring quality of care with routine data: avoiding confusion between performance indicators and health outcomes. *BMJ* 1999;319:94-8.
- Hartzband P, Groppman J. Money and the changing culture of medicine. *N Engl J Med* 2009;360:101-3.
- National Patient Safety Agency. Safer lithium therapy. 2009. www.nrls.npsa.nhs.uk/resources/type/alerts/?entryid45=65426.
- National Institute for Clinical Excellence. *Bipolar disorder: the management of bipolar disorder in adults, children and adolescents, in primary and secondary care*. Clinical Guideline 38. NICE, 2006.
- NHS Information Centre. General practice extraction service. www.ic.nhs.uk/services/in-development/general-practice-extraction-service.

Cite this as: *BMJ* 2010;340:c1717

FROM BMJ.BLOGS

The value of an intercalated BSc



Helen Jaques asks why more medical students in the United Kingdom don't take the opportunity to study for an intercalated Bachelor of Science degree as part of

their medical degree. There are numerous advantages—"a key benefit of doing an intercalated BSc is the edge it gives in the job market: having a BSc is worth between 2 and 6 extra points in the UK foundation programme, which is the route through which medical students get their first junior doctor job after graduating," she writes. "Undergraduates who do an intercalated BSc also do better in subsequent medical exams than those who don't do the extra year." And yet only 30% of students choose to do an intercalated BSc and that number seems to be declining.

Vidhya Alakeson writes about the task the US government now faces to put the health reform bill into practice. "It is time for the hard work of implementing the pages of provisions included in the health reform bill that was passed by Congress last month," she says. "While the media coverage has rightly focused on the historic extension of health insurance to cover most of the 47 million uninsured Americans, the legislation contains many other provisions that are worth a closer look, particularly in light of current priorities within the NHS."

Tracey Koehlmoos was surprised on a recent visit to Jakarta to see a 25 foot poster of the Marlboro Man. Smoking is big problem in Indonesia. "We switched tables at more than one restaurant when someone lit up right next to us," she says. "People smoke on the streets, at bus stops and in restaurants. I learned that some 64% of men in Indonesia smoke—that is approximately 57 million people."

► Read these articles and other BMJ blogs at <http://blogs.bmj.com/bmj>