NATIONAL INSTITUTE FOR HEALTH AND CLINICAL EXCELLENCE

QUALITY AND OUTCOMES FRAMEWORK (QOF) INDICATOR DEVELOPMENT PROGRAMME

Review of QOF indicators

QOF indicator area: Diabetes

Date of Primary Care QOF Indicator Advisory Committee meeting: 13th December 2012

Contents

Introduction

NICE has been asked to consider developing a Quality and Outcomes Framework (QOF) composite indicator (see Types of composite indicators below) covering 9 recommended care processes for people with diabetes. This request follows the publication of the National Diabetes Audit, which suggested that although QOF records show that achievement of the individual care processes is high, only around 53% of people with type 2 diabetes receive all 9 care processes.

This briefing note outlines the factors to be considered when developing a composite indicator of care processes for people with diabetes.

Background

NICE recommends that people with diabetes have the following annual checks:

- body mass index (BMI) measurement
- blood pressure
- smoking status
- blood glucose levels (HbA1c)
- urinary albumin test
- serum creatinine test
- cholesterol levels
- eye check (retinopathy screening)
- foot check.

These recommendations cover process measures rather than measures of treatments or intermediate outcomes. However, for some of these recommendations the 2012/13 QOF includes only measures of intermediate outcomes.

In early 2012, the National Diabetes Audit reported that in 2009/10 54% of people with type 2 diabetes received all 9 of these checks. This is a substantial improvement on the figures from 6 years earlier, when only 5% received all 9 checks.

Achievement of these 9 checks was reported to be much higher in QOF. The national points achievement in QOF 2009/10, averaged across the 9 processes, was 97.5% and the average underlying achievement was 93.9% of eligible patients.

Subsequent analysis has shown that at individual indicator level, the differences between the QOF and the National Diabetes Audit figures are

largely a reflection of differences in the code clusters used. In particular, the definition of QOF indicator DM13 (microalbuminuria testing) may have led to differences in the results.

The proportion of people with diabetes who receive all 9 care processes is unknown because the National Diabetes Audit does not reflect the definitions used in the QOF and in clinical practice. Patient-level data would be needed to obtain an accurate estimate of this figure.

One small study from Scotland suggests that achievement against all 9 care processes, using the QOF definitions, is around 58% [1]. However, the practices involved in this study may not be representative of Scotland or the UK, thereby limiting the generalisability of the findings.

Types of bundled indicators

The term 'bundled' has been used when describing a number of different types of indicators. This can lead to confusion about what is meant by the term 'bundled' indicator. For the purpose of this briefing paper we have used the term 'composite indicator'.

One example of what may be considered a composite indicator in the QOF is Cancer 3:

The percentage of patients with cancer, diagnosed within the preceding 18 months, who have a patient review recorded as occurring within 6 months of the practice receiving confirmation of the diagnosis

In this example of an 'unspecified review indicator', the indicator measures the delivery of a review that is intended to contain several elements, but these elements are not directly specified in the indicator description and underlying indicator definition, and are therefore not measured (although they are discussed in the supporting QOF guidance). The composite nature of the indicator is therefore implied and is not explicit in the indicator wording.

The QOF Committee has previously 'unbundled' an 'unspecified review indicator' in to its constituent parts in the context of mental health.

A second example in the QOF is DEM 4:

The percentage of patients with a new diagnosis of dementia recorded between the preceding 1 April to 31 March with a record of FBC, calcium, glucose, renal and liver function, thyroid function tests, serum vitamin B12 and folate levels recorded 6 months before or after entering on to the register.

In this example of an ‘all or nothing’ indicator, a set of specified tests are *all* required to be carried out to be successful against the indicator: failure to carry out one of the tests is a failure against the entire indicator. This kind of indicator is what the patient safety literature would usually call a ‘care bundle’.

A third example in the QOF is SMOKING 5:

The percentage of patients with any or any combination of the following conditions: CHD, PAD, stroke or TIA, hypertension, diabetes, COPD, CKD, asthma, schizophrenia, bipolar affective disorder or other psychoses whose notes record smoking status in the preceding 15 months

In this example of a composite indicator, identical indicators relevant to a numbers of disease areas have been ‘bundled’ together in to one indicator.

Other types of composite indicators have also been described in the literature [2].

Why consider ‘all-or-nothing’ indicators

All-or-nothing indicators have been recommended by developers where different indicators synergistically affect patient outcomes, as is the case for vascular risk. An advantage is that they can identify opportunities for improvement even when performance on individual indicators is generally high [1,3].

Specific ‘all or nothing’ or ‘care bundle’ indicators have been implemented in a range of secondary care settings such as paediatric and adult ICU, medical and surgical wards and Accident and Emergency departments in North America and the UK [4, 5]. Although higher compliance rates with bundles have been associated with improved outcomes [6], these may be difficult to sustain because of a combination of system and human factors which often results in rates below 50% [7,8,9].

One of the problems with ‘all-or-nothing’ measures is when achievement is presented as a single numerator and denominator it is not possible to determine, without further audit, where failure against the indicator may be occurring. However, this is not a difficult IT problem to overcome in the context of the QOF data extraction.

It is important to consider what is reasonable to include in a composite all-or-nothing indicator. For example, including measures with very different likelihoods of success may be problematic, because achievement of the overall indicator will be driven primarily by the constituent indicators that have the lowest likelihood of being achieved. In the diabetes context, the distinction

would be between ‘simple’ process measures that can be done within a review (e.g. smoking status, BMI, blood pressure, blood tests) and ‘complex’ process measures which require the person with diabetes to be in a particular place with a particular professional (e.g. foot and retinal examination).

The role of exception reporting in the QOF would need to be considered in the design of an ‘all or nothing’ indicator. Where QOF indicators have a number of elements that all must be achieved to be counted as a success it is not possible to exception report against one element of the indicator. This applies, for example, to indicators that require a battery of blood tests (DEM 4), a number of drugs prescribed (CHD14) or different pieces of advice (EPILEPSY 9).

Where a series of care process have been brought together in to one ‘all of nothing’ indicator, there may be a risk of people being exception reported against the entire indicator on the basis of declining one element, or being unsuitable for one element of the indicator. This has implications on which process measures should be included within one ‘all or nothing’ indicator.

Appendix B outlines further some of the implications for indicator design, including implications for exception reporting.

Other considerations

Some of the indicators in the QOF that relate to care processes have been retired, or recommended for retirement, for example the measurement of blood pressure has been retired from the QOF, and the measurement of BMI has been recommended for the QOF.

However, the rationale for a new ‘all-or-nothing’ process indicator is that this would likely identify room for improvement in the reliability of care, even if most of the underlying processes were being delivered to high levels. The retirement of individual indicators would therefore not preclude the introduction of an ‘all-or-nothing’ indicator.

A number of options as to what to include in a new ‘all-or-nothing’ indicator would need to be considered:

1. All nine care processes
2. Only subset of the care processes e.g. BMI, blood pressure, HBA1c measurement etc

3. An outcome bundle: careful consideration as to the appropriateness of this option would need to be considered and as to which outcome measures should be included in the bundle.

Appendix A Diabetes indicators in the 2012/13 QOF

DM32. The practice can produce a register of all patients aged 17 years and over with diabetes mellitus, which specifies the type of diabetes where a diagnosis has been confirmed

DM2.The percentage of patients with diabetes whose notes record BMI in the preceding 15 months

DM26. The percentage of patients with diabetes in whom the last IFCC-HbA1c is 59 mmol/mol or less in the preceding 15 months

DM27. The percentage of patients with diabetes in whom the last IFCC-HbA1c is 64 mmol/mol or less in the preceding 15 months

DM28. The percentage of patients with diabetes in whom the last IFCC-HbA1c is 75 mmol/mol or less in the preceding 15 months

DM21. The percentage of patients with diabetes who have a record of retinal screening in the preceding 15 months

DM29. The percentage of patients with diabetes with a record of a foot examination and risk classification: 1) low risk (normal sensation, palpable pulses), 2) increased risk (neuropathy or absent pulses), 3) high risk (neuropathy or absent pulses plus deformity or skin changes in previous ulcer) or 4) ulcerated foot within the preceding 15 months

DM10. The percentage of patients with diabetes with a record of neuropathy testing in the preceding 15 months

DM30. The percentage of patients with diabetes in whom the last blood pressure is 150/90 or less

DM31. The percentage of patients with diabetes in whom the last blood pressure is 140/80 or less

DM13. The percentage of patients with diabetes who have a record of micro-albuminuria testing in the preceding 15 months (exception reporting for patients with proteinuria)

DM22. The percentage of patients with diabetes who have a record of estimated glomerular filtration rate (eGFR) or serum creatinine testing in the preceding 15 months

DM15. The percentage of patients with diabetes with a diagnosis of proteinuria or micro-albuminuria who are treated with ACE inhibitors (or A2 antagonists)

DM17. The percentage of patients with diabetes whose last measured total cholesterol within the preceding 15 months is 5mmol/l or less

DM18. The percentage of patients with diabetes who have had influenza immunisation in the preceding 1 September to 31 March

Appendix B Designing Bundling Indicators: Clinical informatics view

Summary

The purpose of this paper is to present the clinical informatics implications of bundling indicators used for measuring quality and outcomes in General Practice. The design of indicators that are bundled together may have a harmful effective on the accuracy and efficacy of the bundled indicator in incentivising patient care. There are three principles that will mitigate the risk. Each indicator in the bundle should have the same or very similar denominator specification, the same or very similar numerator threshold for success and the same or very similar exception code rate.

The effect of variation on dis-incentivising care of some patients can be reduced by continuing to incentivise individual care processes separately through the individual indicators. Rewarding individual indicators in the bundle separately from the bundle reward will continue to support practices that are striving to provide the best care to as many patients as possible. No way has been identified to successfully mitigate the effect of variation on the accuracy of the reported bundle achievement.

Background

Current indicators vary in complexity but most look for the completion of an individual care process. The simplest indicators report on questions such as how many patients in cohort *x* (the denominator) have a record of single clinical concept *y* which indicates the successful completion of a care process (the numerator).

Denominators and numerators

The definition of the cohort x may be complex but it is generally based on the presence of one of a defined cluster of codes in the patients' electronic patient record (EPR). The codes in the cluster are usually chosen because they represent a specific disease or clinical domain. However some patients in this cluster may be excluded because the indicator is not relevant to them, on grounds of age, gender or treatment received. The reason for exclusion is usually that the care process does not apply to them. Because of this flexibility, the denominators of indicators within one clinical domain may not all be the same. Achievement of an indicator is measured by looking in the records of patients in the denominator for codes that give evidence of successful completion of the care process. These codes define the numerator code clusters. There may be one or more code clusters in the

numerator. Further logical rules may also be applied to the specification of the indicator, such the application of specific date ranges to code clusters.

Effectively anonymised data

The practice is rewarded if the percentage of patients with evidence of successful completion of the indicator requirement exceeds a given threshold. This means that only aggregate data which is effectively anonymised needs to be disclosed by the practice so patient consent to the disclosure is not required.

Exceptions

Some patients fall within the specification of the denominator but for some good reason out of their control, the practice was unable to provide the care process to them. The practice can record an exception code in their record. The indicator queries are designed to ensure that, for each indicator, a patient with an exception code for the indicator is removed from the numerator and the denominator for that indicator. Thus the practice is not normally penalised for not completing that care process for that patient.

An excepted patient is not removed from the domain disease register and so the practice prevalence payments are not affected but if a proportionately large number of patients are “excepted” from the indicator the percentage of the remaining patients in the indicator that have to successfully complete the care process inevitably increases. This is a by-product of the exception process. At the margins practice payments may sometimes be effected but the effect is smaller than if exceptions were not allowed.

Specificity and sensitivity of the meaning and use of codes

The clinical concepts that are used as evidence in constructing the denominator and numerator populations are represented in the patients' records by coded data. The presence of many of the concepts in the record can be accurately determined by looking for one of a group or cluster of unambiguous codes but it is worth stating that this is not always the case. There are concepts where the code clusters are either less specific or less sensitive and sometimes both. Such clusters offer only an approximate measure of the incidence or prevalence of a concept in a patient's record. The explanation for this variation and its impact are not directly relevant to this paper and will not be discussed further.

An example of a simple QOF indicator
DM2: The percentage of patients with diabetes whose notes record BMI [body mass index] in the preceding 15 months. *Denominator based on diabetes mellitus register* Include patients — with a record of BMI in the preceding 15 months Exclude patients — registered with the practice for less than 3 months with an exception code in the preceding 15 months with diabetes mellitus diagnosed in the preceding 3 months *Numerator* Patients in the denominator with a record of BMI in the preceding 15 months

Clinical domains

Clinical Quality and Outcomes Framework (QOF) indicators are grouped into clinical domains. Most domains relate to one disease or important discrete clinical scenarios such as heart failure or palliative care. The majority of indicators in each domain have the same or similar denominator specifications and, especially for long term conditions, most of the care processes are provided by the practice at the same six or twelve monthly appointments. Examples include diabetes, hypertension or cardiovascular clinics. Alternatively all the care processes may be provided in one appointment focussed on the review of one condition, often using a call/recall system to make sure that patients attend regularly.

Bundled indicators

Patient care is sub-optimal if some of the care processes or outcomes are not achieved and practice organisation should be focussed on ensuring that every patient receives all the care processes that would benefit them. Optimal care may be measured and incentivised by checking that all the individual care processes/outcomes are achieved for each patient. This can be achieved by combining individual indicators in a domain into a single bundled indicator.

Currently only aggregate data is extracted from GP systems to report on achievement of individual indicators. It is not possible to calculate from this data how many care processes an individual has received. That requires patient level data to be extracted and that would probably require explicit patient consent for the data to be disclosed by practices because of the risk of re-identification of the data. So the only way to bundle indicators is to process the data in the GP system before it is extracted.

The bundle logic is essentially include any patient from the denominator in the numerator if:

Indicator 1 is achieved
AND
Indicator 2 is achieved
AND
Indicator 3 is achieved
AND ...

There are examples of how this might be done in QOF for 2012-13 (CHD14, DEM3, SMOK5 and SMOK6). Consideration has also been given to bundling cohorts of patients with different co-morbidities into one indicator cohort, e.g. all patients with hypertension and end-organ damage or diabetes where the target blood pressure is lower than the rest of the population.

However there are circumstances where simply bundling existing indicators together can give misleading results. To bundle clinical indicators together successfully certain conditions must be met.

Recommendations

1. The denominator specification for each indicator in the bundle must be the same or very similar

The measurement of success in a bundled indicator must report on the absolute number of patients that have achieved all care processes. This is simple if all the patients are eligible for all the processes. However if some are not then the calculations must be more complicated and the results are less predictable.

Examples, where the bundle population is assumed to be the smallest or largest denominator:

	Indicator 1	**Indicator 2**	**Indicator 3**	**Bundle 1**	**Bundle 2**
Denominator	60	80	100	60	100
Numerator	55				
Achievement	92%	69%	55%	92%	55%

Assuming 55 patients in the bundle complete all the care process:

1. If the lowest indicator denominator is taken as the bundle denominator (Bundle 1) the reported success is 92%.

2. If the highest individual indicator is taken as the bundle denominator (Bundle 2) the reported success of the bundled indicator is 55%.

It is hard to say that either result is representative of the achievement of the practice. The only straightforward way to avoid this problem is to only bundle together individual indicators that have the same or very similar denominator specification.

2. The numerator threshold (payment stage) for each indicator in the bundle must be the same or very similar

Each indicator in the QOF has two individual thresholds for the indicator which relate to the difficulty of achieving the indicator target. Success thresholds for some indicators are much lower than the average for the domain (e.g. the top success threshold for diabetes indicators varies from 50% to 90%). If achievement thresholds for individual indicators in a bundle are significantly different, setting a reasonable overall threshold for the bundle becomes problematic.

1. If it is set at the lowest individual threshold, the incentive to achieve higher success rates on the easier indicators is reduced.

2. If it is set to the highest individual threshold then the difficulty in hitting raised thresholds for the more difficult indicators will become a dis-incentive to trying to succeed on the bundle.

Generally indicators such as a BMI recording are easier to achieve than indicators of an outcome such as hitting a blood pressure threshold. It takes one patient appointment to record height and weight. It takes several visits to lower raised blood pressure to a satisfactory level. Some indicators are carried out outside the practice and are not directly under control of the practice, and so a fair threshold may be lower than a simple indicator.

Examples, where the numerator thresholds for individual indicators are different:

	Indicator 1	**Indicator 2**	**Indicator 3**	**Bundle 1**	**Bundle 2**
Denominator	100	100	100	100	100
Numerator threshold	50%	65%	90%	55	90
Achievement for every indicator to pass bundle threshold				55%	90%

This suggests that bundles will only function well if the numerator thresholds for each individual indicator in the bundle can be reasonably set at the same or a very similar percentage.

3. The exception code rate for each indicator in the bundle must be the same or very similar

Exception rates in single clinical domains vary between indicators. The effect of an exception code in a patient's record is to remove the patient from the denominator and the numerator. The remaining patients who have completed the care process then count against a smaller denominator and the number of patients that have to complete the care process to hit the indicator threshold is reduced.

Example of the effect of exception rates on the number of patients needed to hit a 90% threshold with an original denominator without exceptions of 100

	Indicator 1	**Indicator 2**	**Indicator 3**	**Bundle 1**	**Bundle 2**
Denominator	100	100	100	100	100
Exception rate	2%	5%	7%	2%	7%
Amended denom	98	95	93		
Threshold	90%				
Numerator required	88	85	84	88	84

There are three ways that exceptions can be treated in bundled indicators

1. Exceptions codes could be ignored in bundled indicators. Calculation of the bundle numerator will be simpler but when a practice was unable to complete a care process for a good reason despite their best endeavours they may fail to hit the total indicator which would be a disincentive for completing the rest of the care processes and it would be contrary to how all other indicators work.

2. Single exception codes may be counted against the whole bundle so that every patient who has a single exception is removed from the bundle denominator and numerator. Except in the case where an exception applies to all indicators the total exception rate will increase as the sum of exception codes increases. Where six indicators are

bundled and the average indicator exception rate is 5%, the total number of patients excepted could rise to 30%. There would be no incentive to complete all the care processes for the 30% of patients who have a single exception to one indicator. This is perhaps an extreme case but even lower total exception rates would be a significant disincentive to care.

3. If the exception codes are applied just to the individual indicator the number required to hit the threshold for each indicator will vary unless the exception rate is the same for all the indicators in the bundle (see the table above). So variation in exclusion rates causes a problem of setting a threshold for the bundle. However, if the exception reasons and frequency of coding for bundled indicators is very similar or the same then an accurate threshold that does not de-incentivise care can be used.

Example of the use of these recommendations:

Sometimes local retinopathy screening programmes have failed to provide a service to all patients in the locality because the service has been withdrawn unexpectedly. Patients who it was not possible to screen would be excepted from the QOF. This might have a huge impact, reducing the denominator and increasing the exception rate for the bundle by 10s of percent. This makes the interpretation of the data reported for the bundled indicator unreliable. It would undermine the Outcomes Framework process.

The three recommendations predict that retinopathy screening could be a poor choice for a bundled indicator on diabetes.

4. The impact of variation in the denominator specification, the success numerator threshold or the exception code rate will be reduced if the achievement of individual indicators is still incentivised through the QOF.

One impact of bundling indictors with different denominator specification, success numerator threshold or exception code rate is to dis-incentivise individual care processes for patients who are not within the success criteria for the bundle. This can be minimised by reducing the impact on practices' overall QOF achievement if individual care process completion is still incentivised though individual indicators.

The other effect is to reduce the accuracy and therefore the credibility of the achievement reported by QOF for the bundle of care processes.

Conclusion

There is evidence that the best way to provide effective care for patients with long term conditions is to organise provision of care so that all the components of care can be delivered at a single visit to the practice, at regular intervals usually six or twelve monthly. A bundled indicator of all the relevant care processes might measure the thoroughness of care provided by the practice and incentivise this form of organisation of care provision.

However if the denominator specification, the success numerator threshold or the exception code rate of the bundled indicators is not the same or very similar, there is a tendency to dis-incentivise the best standard of care for some patients. This risk can be mitigated by continuing to incentivise individual care processes through individual indicators. However this will not mitigate the impact on the accuracy of the reported achievement and therefore the credibility of QOF may be harmed.

These recommendations should be taken into account when designing bundled indicators.

Ralph Sullivan
National Clinical Lead for Primary Care
Health and Social Care Information Centre
13 November 2012

References

1 de Wet C, McKay J, Bowie B (2012) Combining QOF data with the care bundle approach may provide a more meaningful measure of quality in general practice. BMC Health Services Research 12:351

2 Reeves D, Campbell SM, Adams J et al. (2007) Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. Medical Care 45: 489–596.

3. Guthrie B, Emslie-Smith A, Morris AD. (2009) Which people with Type 2 diabetes achieve good control of intermediate outcomes? Population database study in a UK region. Diabetic Medicine 26(12) : 1464-5491

4.Crunden E, Boyce C, Woodman H et al. (2005): An evaluation of the impact of the ventilator care bundle. Nursing in Critical Care 10(5):242–246.

5. Carter C (2007) Implementing the severe sepsis care bundles outside the ICU by outreach. Nursing in Critical Care 12(5):225–230.

6. Guerin K, Wagner J, Rains K, et al (2010): Reduction in central line-associated bloodstream infections by implementation of a post insertion care bundle. American Journal of Infection Control, 38(6):430–433.

7. Gao F, Melody T, Daniels DF et al (2005): The impact of compliance with 6-hour and 24-hour sepsis bundles on hospital mortality in patients with severe sepsis: a prospective observational study. Critical Care 9(6):764–770.

8. Baldwin LN, Smith SA, Fender V, et al (2008) An audit of compliance with the sepsis resuscitation care bundle in patients admitted to A&E with severe sepsis or septic shock International Emergency Nursing 16(4):250–256.

9. McNeill G, Dixon M, Jenkins P (2008) Can acute medicine units in the UK comply with the Surviving Sepsis Campaign ‘six-hour care bundle’? Clinical Medicine 8(2):163–165.