Netherlands
The Journal of Medicine

REVIEW ARTICLE

# Clinical indicators: development and applications

**H. Wollersheim[1*], R. Hermens[1], M. Hulscher[1], J. Braspenning[1], M. Ouwens[1], J. Schouten[2], H. Marres[3], R. Dijkstra[1], R. Grol[1]**

[1]Centre for Quality of Care Research (WOK 117), [2]Department of General Internal Medicine; [3]Ear, Nose and Throat Department, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, the Netherlands, [*]corresponding author: tel.: +31 (024)-361 88 19, fax: +31 (0)24-354 17 34

## ABSTRACT

**Clinical indicators give an indication of the quality of the patient care delivered. They must comply with high-quality standards and should be constructed in a careful and transparent manner. Indicators must be relevant to the important aspects of quality of care. There should be adequate research evidence that the recommendations from which they are derived are related to clinical effectiveness, safety and efficiency. They should measure the quality in a valid and reliable manner with little inter- and intra-observer variability so that they are suitable for comparisons between professionals, practices, and institutions. Indicators are selected from research data with consideration for optimal patient care (preferably an evidence-based guideline), supplemented by expert opinion. In the selection procedure, the feasibility, such as their measurability and improvability, is important beside validity and reliability. A clinical indicator should be defined exactly and expressed as a quotient. After a try-out, the measurements and reporting should follow. The report contains an in-depth analysis of causal and contributing factors associated with the measured results. A description of the clinical circumstances and a correction for case mix should be included to allow for a justified interpretation. The indicators must be part of an improvement strategy, for which comparison feedback is often used. We give examples of indicator development and applications in oncology, diabetes care, and the use of antibiotics for treating pneumonia. We explain how comparison with reference data can be used to construct improvement programmes.**

## KEYWORDS

Clinical indicators, quality improvement, implementation, guidelines.

## INTRODUCTION

There is a sense of discomfort among doctors about the increased legislation and control in Dutch health care. Reports and articles concerning suboptimal and unsafe care are making a stronger and stronger call for accounting for the quality of care.[1] A method of public justification, introduced by the Dutch Inspectorate of Health (IGZ) and others, includes the performance indicators for hospitals. At the end of 2003, the inspectorate (www.igz.nl) presented a set of indicators[2] to the Council of Dutch Hospitals for annual publication on their web site and in their annual report.

Methods to justify the level of care activities by quantification were first used two decades ago in the United States, followed by the United Kingdom and Denmark. It is striking that care providers in Dutch hospitals are on the sidelines when it comes to the development and application of indicators. We reply in this article by describing how professionals themselves can work together in devising indicators for the quality of their activities and how they can use these indicators for the purpose of improving the quality of care.

Insight into the quality of care is necessary because research shows time and again that the quality of patient care is not optimal in 30 to 40% of the cases.[3] To acquire insight into the quality of the care provided, one can take measurements with the 'indicators'. An indicator is a measurable aspect of care provided for which there is evidence that it represents quality on the grounds of scientific research or consensus among experts.[4] There are indicators that are more suitable for internal quality improvement (clinical indicators) and indicators that are especially appropriate for external appraisal (performance indicators).[5]

From the viewpoint of measurement, there are three types of indicators: outcome, process, and structure indicators.[6] This article focuses on the clinical indicators. These are mainly process indicators, aimed at measuring and improving clinical activities in practices and care institutions.

The first part of this article describes a carefully founded systematic procedure to achieve qualitatively good clinical indicators. We have considerable experience with this procedure.[7] Compared with other methods in the literature,[8] this approach offers an acceptable balance between robustness and feasibility. The second part of this article shows, with examples, how carefully developed clinical indicators can provide insight into the quality of patient care and can support and direct improvement activities.

## METHOD OF DEVELOPMENT

It is important that clinical indicators meet careful quality requirements such as relevance, validity, reliability, and applicability.[9]

*Table 1* summarises the most important quality requirements. The clinical relevance is stipulated by the degree of scientific proof that the indicator contributes to health benefit. The indicator must represent important dimensions of care (professional care, but also organisational and patient-specific care). Clinimetrical analyses are necessary to determine the validity and reliability. Besides clinimetric properties such as face, content and construct validity also its validity in the context of its actual use should be considered. The practical usability is determined by the acceptance, measurability, and improvability. Generally, the measurability from routine value files cannot be well estimated beforehand and can best be determined by a test measurement. The result of the measurement must be useful for quality improvement. This requires that the indicator is sensitive to change, has sufficient discriminating capacity for comparison, and is useful as a decision tool.

**Table 1.** *Quality characteristics of the indicators*

| | |
|---|---|
| Relevancy | Relevant to important aspects (effectiveness, safety, and efficiency) and dimensions (professional, organisational, and patient oriented) of quality of care |
| Validity | - Strong correlation with the current quality of care<br>- Valid on the basis of good scientific proof and experience |
| Reliability | - Low inter- and intra-observer variation<br>- Available and reliable date sources<br>- Statistically reliable, i.e. reported as an average or median with confidence intervals and valid for comparison, i.e. corrected for case mix and sociodemographic variables |
| Feasibility | - Easily available<br>- Applicable to quality improvement; i.e. easy to build in improvement initiatives<br>- Sensitive to improvement in time<br>- Useful to base decisions on (caregivers, patients, regulating agencies)<br>- Applying to those who should use them |

We can take clinical indicators from the existing literature or develop them ourselves. If we use data about optimal care (from scientific research such as systematic reviews, meta-analyses or from evidence-based guidelines) in the development process, they can be augmented with experts' clinical experience in a structured consensus procedure.[10] These experts select and prioritise the recommendations. The steps for the development and practical application of indicators (*table 2*) are as follows.

**Table 2.** *Steps in the development and application of clinical indicators*

**I. Selection of relevant patient group or care process.**
Criteria:
1. Experience with care problems (variation, suboptimal care, lack of safety, complaints, costs, long waiting and process times)
2. Important to the purpose of the department, care institution, or scientific association; or of political or moral importance
3. High volume
4. Enough evidence available

**II. Literature search for indicators already developed or data about optimal care available (preferably recent evidence-based guidelines)**

**III. Composition of a balanced consensus group and application of a structured development procedure**
1. Specification. Extraction of concrete recommendations from evidence-based guidelines
2. Prioritising. Selection by an expert panel on the basis of relevance for health benefit, efficiency, measurability, and improvability

**IV. Operationalisation. Processing to definition and proportion**

**V. Availability**
1. Data. Choice of database and unambiguous method of data collection by well-instructed data collectors
2. Practice test. Test of measurability and intra- and inter-reviewer reliability

**VI. Report**
1. Statistics, tabulations, and data presentation
2. Correction for case mix and sociodemographic variables
3. Clear explanations of the results

**VII. Application to the system of quality improvement**
1. Feedback with self, external, or standard comparisons
2. Analysis and discussion of clinical indicators with a low score
3. Analysis of obstructing and conducive factors for providing optimal care
4. Formulation of improvement and implementation strategy and carrying out of the project plan
5. Monitoring of indicators as measurements of effect and for maintenance of improvement
6. Process analysis (was the improvement process carried out as agreed?)

## THE SELECTION OF RELEVANT CARE PROCESSES

First, a choice must be made about which relevant care process one wants to develop indicators for. The most

important selection criterion is the problems experienced in providing optimal care. Problems are visible if there is an unexplained variation in the care between care providers or care institutions,[11] or if it appears that data about optimum care are not applied, or not applied correctly,[12] e.g. there is consequently a lack of safety (high morbidity, mortality, complications, or errors), dissatisfaction (complaints and dissatisfaction of patients and employees), or inefficiency (capacity problems and high costs). Because developing and measuring indicators is time consuming, it is judicious to select care processes with a considerable volume (many patients and high costs of staff and resources).

## LOOKING FOR SCIENTIFIC PROOF IN THE LITERATURE

To find out whether indicators for the selected care processes have already been developed, it is advisable first to consult databases and sources of international indicators such as http://www.rand.org, http://www.ahcpr.gov, http://www.newcastle.ac.uk/qip and http://nprdc.man.uk. Indicators developed elsewhere should be tested with the criteria of relevance, validity, and reliability. They must also be adjusted to the Dutch situation to be able to give answers to the local problems. Apart from a gain in efficiency, the fact that reference values are known is an advantage when indicators are adopted. However, it is often necessary to adjust them[13] and the involvement required in developing them oneself is missing when indicators are adopted.

In order to develop clinical indicators, there must be a basis of recommendations with adequate scientific proof of their effectiveness, safety, and efficiency. A systematic search of the literature for the provision of optimal care is needed. Here we assume that evidence-based guidelines, such as those of the Dutch College of General Practitioners (NHG) and the Dutch Institute for Healthcare Improvement (CBO), act as ideal extraction sources.[14]

## THE CONSENSUS: GROUP COMPOSITION AND PROCEDURE

On the one hand, a group of experts is composed to prioritise the scientific data and on the other hand to complete the data with the knowledge of experience. National experts from a guideline committee are a logical choice. It is a good idea to augment this group of 'content experts' with some methodological experts. All the professions involved in the care process should be represented. To prevent the perspective from being limited, special attention must be paid to the participation of paramedics and specialised nurses. Besides them, managers, health economists, and patients are frequently lacking in such teams.[15] A review of studies that compared consensus methodologies shows that, starting with 12 participants, adding more participants seldom changes the result of the selection procedure in any important way.[16]

In the structured development procedure, the phases of specification and prioritisation can be distinguished:

**Specification:** At least two content experts select the core recommendations from an evidence-based guideline. Because many quality documents have a narrative design, the pretreaters must sometimes transform the consensus text into concrete recommendations.

**Prioritisation:** The second phase consists of systematic prioritising on the basis of a number of relevance requirements[8] such as the degree of evidence on which the indicator recommendation has been based[14] and the importance of the indicator for the outcome of patient care (effectiveness, safety, and costs).

The opinion of the experts can be obtained in discussions at meetings or from anonymous mail surveys. The latter is more efficient and reaches further, but it lacks the nuance of discussion and argumentation. Generally, we combine both in what is known as the Rand-modified Delphi methodology.[17] A panel of experts anonymously rates the core recommendations in writing on a point scale, for example, from 1 to 9. 'Relevance for health benefit and efficiency, measurability, and improvability' are much-used dimensions for assessing the items. The panel can make observations concerning the formulations chosen and add new recommendations. After calculating the average score of all the experts, each recommendation is accepted according to a previously determined weighting (e.g. average score above 60%), considered again, or rejected. The summarised results are discussed in the group. In a second round, all doubtful recommendations, all newly added recommendations, and all reformulated recommendations are rated again. This eventually produces a list with prioritised recommendations against which actual practice can be reviewed.

The result must balance the types (structure, process, and outcome) and the quantity of indicators well. Generally, there is some conflict between the many indicators selected (to get to the greatest possible insight into the care process), and the quantity of work which must be spent on recording. If one puts too much emphasis on minimising the recording efforts, and only a few indicators are selected, then only limited components of the care process can be judged. This can lead to neglect of important parameters that are not rated. Our experience prompts us to recommend selecting about 12 clinical indicators for a care process to achieve a good balance.

## OPERATIONALISATION

The experts operationalise the prioritised recommendations proportionally with exact descriptions of the nominator and denominator. The indicator is so defined that the larger the proportion, the better the care. The denominator describes the patient group in absolute numbers: for example for those with diabetes mellitus. The numerator reflects the actual result in the patient group. Thus, one reaches, for example, the process indicator of 'the percentage of patients for whom the $HbA_{1c}$ concentrations have been determined once a year' as a measure of the care provided for diabetes. The relevant outcome indicator here would be the percentage of diabetes patients with $HbA_{1c}$ below 7%.

## PRACTICAL TESTING

The measurements must produce reliable clinical indicators. Reliability means that there is little variation between data collectors and that the individual data collector is consistent. This requires rigid definitions and a consistent, complete, and swift manner of recording from reliable data sources. Data collectors must be trained with an eye to univocal interpretation, collection, and classification of data.

The data should preferably be collected in an automated fashion from existing files because this requires the least extra effort. Unfortunately these files often serve a purpose (e.g. financial registrations) other than indicator collection so that the definition and description of the indicator is often not exactly the same.[18]

Surveys of patients and caregivers or data from medical records serve as alternative data sources. In the case of record analysis, the subjective interpretation of notes, missing data, and the lack of considerations for making decisions reduce the reliability. The prospective collection of data to be interpreted unambiguously is, of course, the best approach, but in practical terms this often cannot be realised.

The clearer the definition and the more complete and more useful the source material is for measurement, the more reliable the results will be. Because the measurability is frequently difficult to estimate in advance and often proves to be disappointing in practice, it is wise to perform a limited test measurement beforehand. Sometimes it appears that more than half of the suggested indicators cannot reliably be measured.[19] Furthermore, test measurements often lead to refining the definitions.

## REPORTING

After further selection and adjustment on the basis of the practical testing, the collection of the definite dataset, statistical processing (reproduction in averages or medians, with the standard deviation and the confidence intervals) and reporting and interpreting of the data follows. Then tables or figures can be reproduced. The reporting often requires corrections for confounding factors such as case mix and sociodemographic variables.[20]

When data are used for comparisons there is always much discussion concerning which risk factors are important, which ones influence the results, and which risk correction method is the most suitable. To correct for confounders patient groups are often taken from similar settings, and subpopulations are excluded from the denominator or categorised in low- and high-risk populations with separate scores. A more refined methodology consists of correction on the basis of co-variables in a multiple logistic regression model.[21] A disadvantage of a sophisticated correction method is that the resulting data are difficult to understand, even for experts.

The comparison between the results obtained and reference data must challenge professionals to make improvements. There are three forms of comparison: self, external, and standard. In the relative sense, one can make comparisons with one's own performances at the time (self-comparison) or with others (external comparison, such as with best practice). In the absolute sense, one can make a comparison to a predetermined standard (benchmark). An advantage of self-comparison is that no correction for confounders is necessary, assuming that the population and patient characteristics remain rather constant in time.

## BUILDING IN A SYSTEM OF QUALITY IMPROVEMENT

Registration of the clinical indicators is not a purpose in itself; it is the base for developing and evaluating improvement strategies. The improvement interventions themselves generally consist of two steps. First, the scores are reported to the care providers; this is the feedback. The literature shows that feedback is an effective improvement strategy that, on average, leads to an improvement of 10 to 15%.[22] Second, unsatisfactory scores must trigger quality improvement.

## IMPROVEMENT PROJECTS

The impact of feedback can be maximised by having experts link it to a well-founded form of quality improvement such as periodic audits. Of course, other strategies for improvement can also be used; for a complete overview of possible improvement and implementation strategies, see www.qualitytools.ahrq.gov/qualityreport and Grol and Wensing.[23]

Within the framework of local quality improvement, the indicator is used to identify bottlenecks. If a score is unsatisfactory, an in-depth analysis must take place: why is the care the way it has been observed to be? The problems within a care process with a poor score can be inventoried, e.g. by means of surveys in which possible and feasible solutions are asked about. An analysis of obstructing and conducive factors for optimum care is essential.[24] The improvement programme is converted into a concrete project proposal with a responsible project leader. A project goal with the intended gain in the indicator score in a given time is formulated. One must take into account factors such as the investment necessary and the expected participation of those involved, and integration with initiatives already planned. It is wise to systematically review the literature regarding the planned improvement efforts. Preferably, effective elements of intervention programmes, important to the relevant problems, should be incorporated into the improvement plan. For example, we first searched the literature to detect effective elements in improving care for the chronically ill before intervention activities were executed.[25]

The project proposal contains a description of the strategy of change, taking into account the obstructing and conducive factors. Both a process analysis (was the project carried out as agreed?) and an outcome analysis (did the indicator improve as intended?) should be included. Naturally it is important to monitor the improvement in the indicator score periodically after the project so that the impact does not fade away in time.

## EXAMPLES OF DEVELOPMENT AND APPLICATION OF CLINICAL INDICATORS

Here we discuss the practical development of some clinical indicators and the results of attempts to improve the patient care. In the first example, the emphasis is on the development of clinical indicators for oncology, i.e. patients with a head or neck tumour. In the second and third examples, the emphasis is on the practice tests, the resulting scores, and the improvement strategies that are based on them. In the second example we also discuss a chronic syndrome (diabetes mellitus) with an intervention specific to the patient and in the third example to prescribing medicines (antibiotic use for pneumonia) with an intervention specific to the caregiver.

### 1. Clinical indicators for head-neck tumours

Approximately 440 new patients with malignant head or neck tumours are seen in the Radboud University Nijmegen Medical Centre every year. Problems with the coordination of care and long waiting times for treatment were the reason to start improvement activities. The availability of recent evidence-based guidelines for treating carcinomas of the larynx, the cavity of the mouth, and the oropharynx[26,27] and an active, multiprofessional, tumour working group for improvement activities satisfied a number of preconditions for a good start.

Three reviewers extracted 30 concrete recommendations from the text of the guidelines which contained 85 recommendations. There was a high degree of scientific proof that the 30 recommendations represented good care. These were presented to the members of the tumour working group in a written round. To ensure a broad perspective, the tumour working party, which consisted mainly of clinicians (nose, ear and throat doctors, mouth and jaw surgeons, radiotherapists, a medical oncologist, a radiologist, a nuclear therapist, and a pathologist), was augmented with paramedics (a logopaedist and a dietician) and a specialised nurse. In the written round, the 15 experts were asked to rate the recommendations on a scale of 1 to 9. The criterion for the score was the expected relevance for health benefit when the recommendation was put into practice. They were also asked to prioritise the recommendations (in the form of a top five), to refine the formulation, and to add any new recommendations they wished. The criteria for judging the recommendation were set beforehand as cancel (score: 1–3), doubtful (score: 4–6) and definitely include (score: 7–9 or appearing more than once in the top five priorities).[28]

In addition to the 14 clinical indicators obtained from the guidelines, four additional recommendations for good organisation of care were formulated from the literature.[25] These recommendations were connected with the fields of coordination and continuity of care. A random sample of 30 patients were also asked to rate the recommendations the same way as the professionals did. This resulted in five more indicators, which were patient specific.

The total number of the recommendations was 23. Two were not measurable at all in the practice test. The measurability of the other 21 indicators was between 35 and 97%, with an average of 57%.[29]

On the basis of low scores for baseline measurements (*table 3*), a number of improvement projects are currently being carried out. These projects include the content and fine tuning of the information supply, the logistics of the care process (a planned clinical path) and improvement of support for the patient who is making lifestyle changes (stop smoking, reduce alcohol use, and change diet) and voice rehabilitation.

### 2. Diabetes mellitus

In a way analogous to the methodology already described in the section above, 58 internists and an expert panel developed 18 indicators (12 process indicators and 6 outcome) for good diabetes care with the aid of the CBO

**Table 3.** *Overview of scores of selected* clinical indicators (percentage of patients for whom the recommendation was carried out)*

| | Scores * |
|---|---|
| For patients with head or neck tumour ($n$ = 189)[28] | |
| 1. Provision of information (12 items) | 44 |
| 2. Psychosocial support | 21 |
| 3. Swallowing and voice rehabilitation | 20 |
| 4. Lifestyle support | |
| Alcohol consumption | 25 |
| Smoking | 30 |
| Diet | 0 |
| 5. Admission time (<24 hours) | 24 |
| 6. Time to treatment (<30 days) | 29 |
| For patients with diabetes mellitus ($n$ = 1465)[29] | |
| 1. Annual foot inspection carried out | 40 |
| 2. Exercise advice given | 29 |
| 3. Smoking pattern discussed | 27 |
| 4. Weighted annually | 12 |
| 5. Achieved an $HbA_{1c}$ of <7% | 23 |
| For patients with pneumonia ($n$ = 489)[30] | |
| 1. Antibiotic recommended by guideline is prescribed | 45 |
| 2. Sputum sample taken before start of antibiotic | 54 |
| 3. Blood sample taken before start of antibiotic | 57 |
| 4. Antibiotic stopped after 3 fever-free days | 11 |

* **Selected on the basis of a low score (ie<60%)**

guidelines. An expert panel approved these indicators. Then the indicators were measured in 13 hospitals in 1460 patients with diabetes mellitus. The medical records, questionnaires, and existing data files were used for collecting data.

The average adherence to the indicators was 64%. *Table 3* shows a number of low scores. Multilevel logistic regression analysis showed which factors were responsible for the low scores. The main factors were a lack of diabetes nurses in the practice concerned and a low educational level of the patients.[30] To involve the patients more actively in the care provided, the Dutch Diabetes Federation devised a 'diabetes passport'. In addition to informative material, the passport contained a check list in everyday terms for the activities which the care provider had to carry out according to the guideline. Thus, the patient was able to take part in obtaining insight into the activities that should be carried out according to the guideline.

**3. Use of antibiotics for pneumonia**

On the basis of national and international guidelines, indicators for antibiotic use for pneumonia contracted at home were formulated in a systematic consensus procedure similar to those in the preceding examples.[31] Four of the 20 indicators were rejected because they did not fulfil the requirements of reliability and availability during the practice tests. Then the 16 remaining indicators were measured during a six-month period in the departments of internal medicine and lung diseases in eight medium-sized hospitals in a total of 1000 patients. The data were checked to see if they had to be corrected for case mix. To describe the case mix, demographic data, comorbidity, and seriousness of the disease were registered. Indeed, it appeared that the taking of blood samples for cultures was negatively influenced by the age of the patient. Sputum samples were more frequently cultured for exacerbations of chronic obstructive pulmonary disease (COPD) caused by airway infections in patients with a low score for the forced expiratory volume in 1 second ($FeV_1$).

*Table 3* shows the relatively low scores obtained for the measurements in pneumonia treatment. Especially the culturing of blood and sputum samples and the right choice and timely administration of the preferred antibiotic scored only moderately. The lowest score was for the indicator 'percentage of patients who stopped taking antibiotics three days after they were free of fever' (11%).

There were also high scores. Switching from a broad-spectrum to a narrow-spectrum antibiotic (adapted to the culture results) or from an intravenous to an oral antibiotic was performed according to the guideline in 80% of the patients. The dose or dose frequency was correctly adjusted to the kidney function in 77% of the patients.

The large variation between hospitals was striking. There was one hospital where a sputum sample was taken for every patient, while in another hospital this was only done in 24% of the patients.

The intervention programme for improving antibiotic use was aimed at the low scores and based on the interviews about factors obstructing optimal care provision. For example, the first-choice antibiotic became available at the emergency departments. It was agreed with the nurses that standard sputum and blood samples would be taken before the first administration of antibiotics.

## DISCUSSION

This article shows a manner of development and examples of application of clinical indicators. It is intended as the beginning of a discussion about how to reach a scientifically justified development and application.

During the development process the use of criteria such as relevance, validity, reliability, measurability and applicability of the indicators is essential.

The professional group for the Department of Social Medicine at the Academic Medical Centre (AMC) in Amsterdam is currently working on a development and testing instrument named AIRE (Appraisal of Indicators through Research and Evaluation). On the basis of

20 questions, the aim, relevance, setting, involvement of interested parties, the degree of scientific proof and practical use are inventoried and rated. The instrument is not yet definite and its value in practice must be still examined.

The two most important quality requirements for the indicators are that they must be based on recommendations of the highest level of scientific proof and that the data collection is reliable.

Clinical indicators are selected from qualitatively good scientific research into optimal care. They should be developed by a panel representing the occupational groups. To guarantee a high level of scientific evidence the recommendations should be extracted from an evidence-based guideline. It is still unclear which method for the production of indicators out of guidelines is the best. The composition of the panel (number of participants, representing professions, coherence in the group, dominance of individuals), the manner of prioritising (selection criteria, opportunities for correction and additions, the scale used) and the consensus procedure (rating system, research by mail or discussion meetings) each determine the outcome to a certain degree. The reliability of the consensus procedure is moderate: between 0.51 and 0.83 when expressed as kappa value.[32] The reproducibility can be improved if a high cut-off value is used, for example, above 8 on a scale of 9.

The practice test in the three investigations presented showed that between 10 and 20% of the indicators were not measurable. It is known from measurements of clinical indicators in Dutch general practices (well equipped with ICT) that an empirical test done in advance can be very worthwhile.[33] A set of 139 selected clinical indicators was examined to see if empirical data could be extracted. The available database came from a nationally representative group of general practices, the Dutch National Information Network of General Practitioners (LINH). After the empirical test, 79 of the 139 indicators were rejected. The reasons for rejection were too little validity,[18] insufficiently reliable,[25] and unsuitable data sources.[34]

Correction of sociodemographic variables and case mix is very important for a reliable interpretation. Our study shows this for antibiotic use: correction of these factors was necessary for older patient populations and for patients with more serious syndromes.

Selected clinical indicators can also be used to publicly account to society (patients, press, and government) if they are presented with clear explanations.[20] If used for external purposes, case mix correction is especially important. There is evidence that if comparison takes place and the case mix is not corrected properly beforehand, there is a risk that (well indicated) high-risk interventions will not take place, or that high-risk patient groups such as those with multiple or complicated disease will be avoided by care providers.[34] Public reporting without coercion, with the anonymity of the individual care providers, and with the necessary distinctions can reduce the threat of unjust judgement. It may prevent data manipulation and also strategic behaviour that can influence the quality of care negatively. There are indications that if feedback is given to caregivers about indicators which they themselves have devised, these undesirable effects do not occur, or occur to a lesser degree.[34] For this reason, the effect on care improvement may be greater than that of public reporting of externally developed and imposed performance indicators.

Besides the search for proper case mix correction and the most optimal use of clinical indicators, there are still many unsolved problems which require closer investigation. Important questions are:

1) What is the optimal and most unambiguous method of development?
2) How can patients and managers best be involved in development so that patient orientation and the organisation of care can also be measured?
3) How to transform the results of measuring clinical indicators into effective and efficient improvement strategies?

In this article, we contribute to this discussion.

## CONCLUSION

The development and use of clinical indicators are important steps on the way to optimising patient care. To continue on this route successfully, in the near future, investment must be made in studies to further improve the development of clinical indicators and to maximise their application. In the long run, it is desirable to link these indicators to a form of practice accreditation and a reward system. This is a course towards integration of clinical indicators into quality improvement systems that should be followed.

## REFERENCES

1. Meijerink Y, Gras M, Eland A, Kallewaard M, Haeck J, Berg M, et al. Werken aan verbetering; ziekenhuizen gaan prestaties openbaar maken. Med Contact 2003;40:1531-4.
2. Berg M, Gras M, Meijerink Y, Eland A, Kallewaard M, Haeck J, et al. Leren van cijfers; ziekenhuizen gaan prestaties openbaar maken. Med Contact 2003;40:1535-8.
3. McGlynn EA, Asch SM, Adams J. The quality of health care delivered to adults in the United States. N Engl J Med 2003;348:2635-45.
4. Campbell SM, Braspenning J, Hutchinson A, Marshall M. Research methods used in developing and applying quality indicators in primary care. Qual Saf Health Care 2002;11:358-64.
5. Berg M, Schellekens WMLCM. Paradigma's van kwaliteit. De verschillen tussen externe en interne kwaliteitsindicatoren. Med Contact 2002;57:1203-5.

6. Donabedian A. Evaluating the quality of health care. Milbank Memorial Fund Q 1966;44:166-206.

7. Campbell SM, Braspenning J, Hutchinson A, Marshall M. Research methods used in developing and applying indicators in primary care. BMJ 2003;326:816-9.

8. Braspenning J, Campbell S, Grol R. Measuring changes in patient care: development and use of indicators. In: Improving patient care. The implementation of change in clinical practice. Grol R, Wensing M, Eccles M [editors]. Elsevier. Edinburgh, 2005; p 222-34.

9. Arling G, Kane RL, Lewis T, Mueller C. Future development of nursing home quality indicators. The Gerontologist 2005:45:147-56.

10. Hearnshaw HM, Harker RM, Cheater FM, Baker RH, Grimshaw GM. Expert consensus on the desirable characteristics of review criteria for improvement of health quality. Qual Health Care 2001;10:173-8.

11. Haas S, Acker D Donahue C, Katz ME. Variation in hysterectomy rates across small geographic areas of Massachusetts. Am J Obstet Gynecol 1993;169:150-4.

12. Erchason J, Petz L, Keeler E, et al. The cost effectiveness of preoperative autologous blood donations. N Engl J Med 1995;332:719-24.

13. Kazandjian VA, Matthes N, Wicker KG. Are performance indicators generic? The international experience of the Quality Indicator Project. J Eval Clin Pract 2003;9:265-76.

14. Wollersheim H, Burgers J, Grol R. Clinical guidelines to improve patient care. Neth J Med 2005;63:188-92.

15. Ayanian JZ, Landrum MB, Normand SLT, McNeil BJ. Rating the appropriateness of coronary angiography. Do practising physicians agree with an expert panel and with each other? N Engl J Med 1998;338:1896-904.

16. Murphy MK, Black NA, Lamping DL, et al. Consensus development methods and their use in clinical guideline development. Health Technol Assess 1998;2:1-88.

17. Normand SLT, McNeil BJ, Peterson LE, et al. Eliciting expert opinion using the Delphi technique: identifying performance indicators for cardiovascular disease. Int J Qual Health Care 1998;10:247-60.

18. Powell A, Davies H, Thomson R. Using routine comparitive data to assess the quality of health care: understanding and avoiding common pitfalls. Qual Saf Health Care 2003;12:122-8.

19. Braspenning JJC, Schellevis FG, Grol RPTM. Tweede Nationale Studie naar ziekten en verrichtingen in de huisartsenpraktijk. Kwaliteit huisartsenzorg belicht. Nijmegen/Utrecht: WOK/NIVEL, 2004.

20. Grol R, Wollersheim H. Naar een correct meetsysteem. UMC's en zorgverleners moeten prestatie indicatoren verder ontwikkelen. Med Contact 2005;60:1348-51.

21. Berg K, Mor V, Morris J, Murphy KM, Moore T, Harris Y. Identification and evaluation of existing nursing homes quality indicators. Health Care Financing Review 2002;23:19-36.

22. Van der Weijden T, Grol R. Feedback and reminders. In: Improving patient care; the implementation of change in clinical practice. Grol R, Wensing M, Eccles M [editors]. Elsevier. Edinburgh, 2005: p158-72.

23. Implementatie. Effectieve veranderingen in de patiëntenzorg. Eds Grol R, Wensing M [editors]. Elsevier, Maarsen 2001.

24. Grol R, Wensing M. What drives change? Barriers to and incentives for achieving an evidence based practice. MJA 2004;180;S57-60.

25. Ouwens M, Wollersheim H, Hermens R, Hulscher M, Grol R. Integrated care programmes for chronically ill patients: a review of systematic reviews. Int J Qual Health Care 2005;17:141-6.

26. Kwaliteitsinstituut voor de Gezondheidszorg (CBO), Nederlandse Werkgroep Hoofd-Halstumoren (NWHHT), Vereniging van Integrale Kankercentra. Richtlijn larynxcarcinoom. Van Zuiden Communications BV. Alphen aan de Rijn; 2000.

27. Nederlandse Werkgroep Hoofd-Halstumoren (NWHHT), Nederlandse Vereniging voor Keel-Neus-Oorheelkunde en Heelkunde van het Hoofd-Halsgebied, Kwaliteitsinstituut voor de Gezondheidszorg (CBO). Richtlijn mondholte- en orofarynxcarcinoom. Van Zuiden Communications BV. Alphen aan de Rijn; 2004.

28. Campbell SM, Hann M, Hacker J, Roland MO. Quality assessment for three common conditions in primary care: validity and reliability of review criteria developed by expert panels for angina, asthma and type 2 diabetes. Qual Saf Health Care 2002;11:125-30.

29. Ouwens MMTJ, Marres HAM, Hermens RRP, et al. Quality of integrated care for patients with head and neck cancer: development and measurement of indicators. Head and Neck 2006; in press.

30. Dijkstra RF, Braspenning JCC, Huijsmans Z, et al. Patients and nurses determine variation in adherence to guidelines at Dutch hospitals more than internists or settings. Diab Med 2004;21:586-91.

31. Schouten JA, Hulscher ME, Wollersheim H, et al. Quality of antibiotic use for lower respiratory tract infections at hospitals: (how) can we measure it? Clin Infect Dis 2005;41:450-60.

32. Shelleke PG, Kahan JP, Bernstein SJ, Leape LL, Kamberg CJ, Park RE. The reproducibility of a method to identify the overuse and underuse of procedures. N Engl J Med 1998;338:1888-95.

33. Braspenning JJC, Pijnenborg L, Grol RPTM. Ontwikkelen van kwaliteits-indicatoren. In: Werken aan kwaliteit in de huisartsenpraktijk. Indicatoren gebaseerd op de NHG-Standaarden. Braspenning JJC, Pijnenborg L, in t Veld CJ, Grol RPTM [editors]. Bohn Stafleu van Loghum. Houten 2005; p10-20.

34. Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. JAMA 2005;293:1239-44.