Out-of-hours primary care: development of indicators for prescribing and referring

PAUL GIESEN¹, M. WILLEKENS¹, H. MOKKINK², JOZÉ BRASPPENNING¹, W. VAN DEN BOSCH² AND RICHARD GROL³

¹Centre for Quality of Care Research (WOK), ²Postgraduate Training of General Practitioners, and ³Department of General Practice, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

Abstract

Background. Dutch general practitioners have reorganized their out-of-hours primary health care to general practice cooperatives. Good insight into the quality of delivered medical care is important to make the accountability of health practitioners and managers transparent to society and to identify and minimize medical errors.

Objective. Development of a set of quality indicators for internal quality improvement in out-of-hours primary clinical care.

Method. A systematic approach combining the opinion of three different general practitioner expert panels, and an empirical test in daily practice. The indicators were based on clinical, evidence-based, national guidelines. We tested the validity, feasibility, reliability and opportunity for quality improvement.

Results. Of the 80 available national clinical guidelines, 29 were approved and selected by the first general practitioner expert panel. Out of these 29 guidelines, 73 indicators concerning prescribing and referring were selected by the second panel. In an empirical test on 36,254 patient contacts, 7,344 patient contacts (22.7%) were relevant for the assessment of these 73 indicators. Six indicators were excluded because they scored more than 15% missing values. In total, 38 indicators were excluded because the opportunity for quality improvement was limited (performance score ≥90%). In the final meeting, the third general practitioner expert panel excluded five indicators, leading to a final set of 24 indicators.

Conclusion. This study shows the importance of subjecting indicators to an empirical test in practice. The national clinical guidelines are only partially applicable in the assessment of out-of-hours primary care. They need to be expanded with topics that are related to general practitioner care in an out-of-hours setting and acute medical problems.

Keywords: Out-of-hours, quality indicators, primary care

Since the turn of the millennium, Dutch general practitioners have reorganized their out-of-hours primary health care to large-scale general practice cooperatives [1, 2]. Although the general practitioners [2, 3] and patients proved, in general, positive about this new organization [4, 5], the quality of medical care delivered by these general practice cooperatives is as yet unclear. Good insight into this quality of medical care is important to make the accountability of health practitioners and managers transparent to society and to identify and minimize medical errors [6].

Performance measurement of general practice cooperative clinical care is, for instance, based on well-developed, evidence-based, clinical guidelines. To make a valid and reliable assessment of the current practice, key recommendations from clinical guidelines can be translated into measurable elements—the so-called indicators [7]. These indicators need to be rigorously developed and need to be valid, reliable and usable in quality measurement [6].

In the Netherlands, the Dutch College of general practitioners have been developing evidence-based, national guidelines for primary care for over 15 years [8]. It is unclear whether these national guidelines are applicable to out-of-hours care. The problems presented in out of hours are different from daytime care due to the more urgent and more ad hoc character of the patients’ complaints [9]. Furthermore, the context differs from that of daytime care: triage nurses and general practitioners usually do not know the patients’ medical history, and general practitioners treat patients only for urgent complaints that cannot wait until daytime [9, 10]. In this article, we describe the development of a set of quality indicators for internal quality improvement in out-of-hours primary care based on clinical
guidelines and test of validity, feasibility, reliability, and opportunity [6, 11].

Methods

We systematically developed a set of quality indicators and tested it in three steps on the basis of criteria described elsewhere [6].

Deriving indicators from guidelines

We based our indicators on the available evidence-based, national, clinical guidelines for general practice [8]. These guidelines were developed in a rigorous procedure that took AGREE criteria into account and combined a systematic review of the literature with consensus meetings of general practitioners (AGREE: Appraisal of Guidelines Research & Evaluation: international instrument to provide a framework for assessing the quality of clinical practice guidelines) [12]. For this reason, we assumed the validity of the clinical guidelines to be sufficient.

In order to develop indicators from guidelines, we first investigated whether the available guidelines were applicable to the out-of-hours setting. A panel of six general practitioners were asked to judge the suitability of all 80 of the available national clinical guidelines for evaluation of clinical care at the general practice cooperative on the basis of ‘clinical relevance’. We used ‘contact frequency’ and ‘urgency’ as the criteria for this judgement. Urgency was used as a criterion because urgent complaints generally have substantial medical consequences for the patient, while their incidence may be limited [10]. A guideline was selected when at least five of the six panel members judged a specific guideline to be relevant for out-of-hours care.

Next, three members of the research team derived all recommendations concerning prescribing and referring from the selected national guidelines. We focused on recommendations for prescribing and referring to hospital specialists because they are best registered in patient records [13].

The selected recommendations were next presented to a second panel of seven general practitioner experts who were asked to judge these recommendations on their relevance and utility for evaluation of clinical care at a general practice cooperative. In case of a positive score, defined as at least six out of seven panel members, the recommendation was directly accepted; if only four or fewer panel members were in favour, the recommendation was immediately eliminated. A positive score of five members allowed further consideration of acceptance or elimination on the basis of consensus discussions by two general practitioner guideline experts.

The accepted recommendations were finally operationalized into indicators by defining them as numerators and denominators.

Empirical testing

We used routinely collected data from a cross-sectional study, in which computer-registered data about patient contacts from one large out-of-hours general practice cooperative were classified with diagnosis codes [14] and a code for urgency of complaints [10]. The study material consisted of records of all 36,254 patients who contacted a general practice cooperative in an urban–rural area in the east of the Netherlands, between July 2001 and June 2002. Records without medical content (e.g. messages from the hospital) were excluded. This particular general practice cooperative has features in common with other general practice cooperatives in the Netherlands.

We analysed the extent to which general practitioners followed the recommendations in the selected national guidelines by investigating computerized medical records of patients. Diagnose codes were assigned to each indicator, and each patient contact was scored as to what extent the related clinical guideline was followed.

Feasibility was defined as the percentage of ‘missing values’ [15]. A contact was considered a ‘missing value’ when a patient contact was incompletely registered or unclear for judgement. Doubtful cases were scored by consensus of two observers, and if no consensus was reached, we added them to the category of ‘missing values’. An indicator that scored a ‘missing’ percentage of 15 or more was excluded.

For inter-rater reliability, we performed a blinded random check of 330 decisions spread out on 37 indicators by three raters. The inter-rater reliability was calculated for two dimensions. The first one tested the agreement in scoring an item on whether a recommendation was followed (kappa: 0.82). The second one tested the extent of agreement on assessing an item as ‘not possible to judge’ (kappa: 0.86).

Opportunity for quality improvement was defined as a performance score of <90% because these indicators still have enough room for improvement [15]. We excluded every indicator with a performance score of 90% or more.

To get a more objective view on the representative of our set of indicators, we classed the complaints presented at this general practice cooperative as urgent or not urgent and compared the results with the data for the test population.

Final evaluation

For testing the validity of this set of indicators, we arranged a final meeting with four general practitioner experts on guideline development and the research team. We asked this panel to determine, on the basis of all the results and the experience gained from our tests, whether this set of indicators represents a good measure of the quality of care provided by general practitioners out of hours. Table 1 gives an overview of the three steps just described.

Results

Deriving indicators from guidelines

Selection of guidelines. Of the 80 available national clinical guidelines, 29 were approved and selected by the entire panel
of six general practitioners on the basis of clinical relevance (Table 2). The kappa of the judgements averaged 0.64.

Selection of key recommendations. The research team derived 138 recommendations concerning prescribing and referring from the 29 selected guidelines. The second general practitioner expert panel immediately accepted 54 recommendations, added 23 of 29 doubtful cases after consensus discussions, and immediately excluded 55 recommendations. Of the 77 remaining recommendations, 8 were combined, resulting in a set of 73 indicators.

Empirical testing

A total of 7344 of 36 254 patient contacts (22.7%) were relevant for the assessment of these 73 indicators. The 7344 contacts included 12 071 decisions that could be related to the clinical guideline recommendations.

The feasibility was high. Six indicators were excluded because they scored more than 15% missing values.

In total, 38 indicators were excluded because the opportunity for quality improvement was limited (performance score ≥90%).

The remaining 29 indicators were classed as not urgent or urgent. As described elsewhere, the entire general practice cooperative has a distribution of 16.3% urgent and 83.7% non-urgent complaints [22]. As related to the indicator set, the distribution of complaints was 35.1% urgent and 64.9% non-urgent complaints.

Final evaluation

In the final meeting, the third general practitioner expert panel and the research team discussed the results and experience gained. The panel determined whether this set of indicators represents a good measure of the quality of care provided by general practitioners in out of hours. On the basis of consensus, this general practitioner expert panel decided to exclude four indicators because they had a limited contact frequency at the indicator level [(1) cornea erosion without a foreign body: local antibiotics; (2) febrile convulsion: diazepam; (3) febrile convulsion: referral; (4) bronchiolitis: referral] and one indicator because of its limited evidence (sinusitis: use of decongestive nose drops). This led to a final set of 24 indicators (Table 3).

Discussion

As far as we know, this is the first attempt to develop indicators for testing the quality of primary clinical care in out of hours. With a systematic approach that combined expert opinion and testing in daily practice, we developed a set of 24 valid indicators. This study shows us the importance of subjecting indicators to an empirical test in practice, because
Table 3 Final set of 24 indicators for prescribing and referring in out-of-hours primary care

<table>
<thead>
<tr>
<th>Guideline and indicator</th>
<th>Patient contact (n)</th>
<th>Performance score (n)</th>
<th>Missing values (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acute coronary syndrome</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Give 240 mg acetylsalicylic acid (including the patients who use coumarine derivatives)</td>
<td>122</td>
<td>39.3</td>
<td>7.6</td>
</tr>
<tr>
<td><strong>Atrial fibrillation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First choice beta-blocker (excluding patients with heart failure)</td>
<td>47</td>
<td>45.5</td>
<td>6.4</td>
</tr>
<tr>
<td>Referral: atrial fibrillation and signs of haemodynamical instability (such as, chest pain, acute heart failure)</td>
<td>88.4</td>
<td>8.5</td>
<td></td>
</tr>
<tr>
<td><strong>Heart failure</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Give a quick-acting nitrate sublingual</td>
<td>133</td>
<td>27.6</td>
<td>4.5</td>
</tr>
<tr>
<td>Give intravenous administration of a loop diuretic</td>
<td>46.8</td>
<td>6.8</td>
<td></td>
</tr>
<tr>
<td>Referral: insufficient result of acute treatment, inadequate possibilities for home care taking and suspected myocardial infarct as the cause for the acute heart failure</td>
<td>88.2</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td><strong>Transient ischaemic attack</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients with a TIA without cardiac arrhythmias or valve disorders: give acetylsalicylic acid.</td>
<td>51</td>
<td>54.2</td>
<td>5.9</td>
</tr>
<tr>
<td><strong>Asthma in adults</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severe dyspnoea: give a beta-2-sympathicomimeticum</td>
<td>152</td>
<td>80.7</td>
<td>1.3</td>
</tr>
<tr>
<td>When not improving sufficiently add ipratropiumbromide</td>
<td>88.0</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td><strong>COPD treatment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severe exacerbation: give beta-2-sympathicomimeticum</td>
<td>218</td>
<td>44.7</td>
<td>1.4</td>
</tr>
<tr>
<td>When improved: give tablets with 25 mg prednisolon daily</td>
<td>70.3</td>
<td>11.9</td>
<td></td>
</tr>
<tr>
<td><strong>Pneumonia</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children: amoxicillin; second choice: azitromycine</td>
<td>75</td>
<td>80.6</td>
<td>10.7</td>
</tr>
<tr>
<td>Adults: doxycyclin; second choice: amoxicillin; third choice: erytromycin</td>
<td>225</td>
<td>53.6</td>
<td>14.7</td>
</tr>
<tr>
<td><strong>Bacterial skin infections</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When presenting with multiple lesions: local fuscidin crème</td>
<td>53</td>
<td>88.7</td>
<td>0</td>
</tr>
<tr>
<td>When reduced resistance, fever and other general signs or worsening of the symptoms despite local treatment: oral flucloxacillin</td>
<td>83.0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Otitis externa</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Otitis externa with fever and other general signs: oral flucloxacillin</td>
<td>80</td>
<td>89.9</td>
<td>1.3</td>
</tr>
<tr>
<td><strong>Otitis media acute</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antibiotics</td>
<td>370</td>
<td>81.6</td>
<td>3</td>
</tr>
<tr>
<td>Increase in general signs as decreased alertness or decrease in fluid consumption</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk factors for complications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No improvement after 3 days</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child: &lt;2 years with otitis media at two sites</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued)
testing our indicators in daily practice proved to reduce our set due to lack of measurability or variability [6, 15].

As we suspected, the national clinical guidelines are only partially applicable in the assessment of out-of-hours primary care due to the other context and the more urgent and ad hoc character of patients’ complaints [9, 10]. Around 80% of all the complaints presented were not covered by the selected guidelines and indicators.

The practice test performance was very good on an average. We had to exclude about half the indicators because of their limited ‘opportunity for quality improvement’. This selection criterion is particularly important for internal quality improvement. However, if the indicator set is to be used for external accreditation purposes, this selection criterion might not be desirable: we then need indicators that highlight both excellent and minimal performance [11].

**Strengths and limitations of this study**

We used a rigorous procedure with several different general practitioner panels, a combination of practice and guideline expertise, and testing of indicators on large numbers of patients and decisions.

The advantage of using indicators from clinical guidelines and studying every medical record is that it provides an opportunity to detect exactly what the specific problems are in daily practice and what the specific limitations of the national guidelines are [11, 15]. A disadvantage of studying every medical record is that it is very time-consuming and expensive [16].

The reliability was excellent, but it was tested only on the entire set and not on each indicator. The feasibility was excellent as well, but the data we used were already coded. To overcome feasibility problems in the future, a system for
extracting the relevant patient contacts from the database of a general practice cooperative has to be developed [16, 18].

By testing whether the indicator set covers the spectrum of problems presented, we found a small shift to more urgent problems. We appreciated this shift because of the possibly severe consequences of very urgent problems.

A limitation of our study is that the indicators used for prescribing and referring give no indication about the quality of the diagnostic process or the advice given, because this information proved to be insufficiently available in medical records [19]. The diagnostic process and patient education are crucial elements of the out-of-hours care and need to be assessed as well.

**Recommendations for practice, research and guideline development**

This study describes the development of a monitoring instrument for out-of-hours primary care. The findings highlight the difficulty of constructing rigorous and useful indicators in such a complex situation as an out-of-hours consultation. Our study has a narrow focus but has to be seen as a first step in the development of useful indicators for out-of-hours and urgent care. How to expand other important aspects of quality?

First, the contact registration on general practice cooperatives is of poor quality, and general practitioners have to be trained for better registration. Particularly, general practitioners have to justify why they differ from the guidelines and they had to describe this in the contact registration. Second, the existing national clinical guidelines are limited usable and need to be expanded with topics that are related to general practitioner care in out-of-hours and acute medical problems. We also recommend research into frequent urgent complaints because the evidence regarding them is limited. We also recommend research into urgent complaints because the evidence regarding them is limited [20–24]. As a spin-off of our research, the Dutch College of general practitioners has now taken the initiative to make special guidelines for out-of-hours and acute medical problems. At last, the scope could be expanded with communicative competencies, skills in taking patients’ history and physical examination. For this we need another design such as participated observation of analysis or video registration.

In summary, we have described how we systematically developed a valid set of quality indicators that can be used to assess the quality of medical care out of hours. Our first impression is that this general practice cooperative has, in general, a high performance score. More research is needed to evaluate the performance of general practitioners in the out-of-hours setting.

**References**


Accepted for publication 29 May 2007