

Comparison of patient evaluations of health care quality in relation to WHO measures of achievement in 12 European countries

Jan J. Kerssens,¹ Peter P. Groenewegen,² Herman J. Sixma,¹ Wienke G.W. Boerma,¹ & Ingrid van der Eijk³

Objectives To gain insight into similarities and differences in patient evaluations of quality of primary care across 12 European countries and to correlate patient evaluations with WHO health system performance measures (for example, responsiveness) of these countries.

Methods Patient evaluations were derived from a series of Quote (Quality of care Through patients' Eyes) instruments designed to measure the quality of primary care. Various research groups provided a total sample of 5133 patients from 12 countries: Belarus, Denmark, Finland, Greece, Ireland, Israel, Italy, the Netherlands, Norway, Portugal, United Kingdom, and Ukraine. Intra-class correlations of 10 Quote items were calculated to measure differences between countries. *The world health report 2000 — Health systems: improving performance* performance measures in the same countries were correlated with mean Quote scores.

Findings Intra-class correlation coefficients ranged from low to very high, which indicated little variation between countries in some respects (for example, primary care providers have a good understanding of patients' problems in all countries) and large variation in other respects (for example, with respect to prescription of medication and communication between primary care providers). Most correlations between mean Quote scores per country and WHO performance measures were positive. The highest correlation (0.86) was between the primary care provider's understanding of patients' problems and responsiveness according to WHO.

Conclusions Patient evaluations of the quality of primary care showed large differences across countries and related positively to WHO's performance measures of health care systems.

Keywords Health care evaluation mechanisms; Patient participation; Primary health care/standards; Delivery of health care/standards; Quality of health care; World Health Organization; Comparative study; Europe (source: MeSH, NLM).

Mots clés Mécanismes évaluation soins; Participation malade; Délivrance soins/normes; Programme soins courants/normes; Qualité soins; Organisation mondiale de la Santé; Etude comparative; Europe (source: MeSH, INSERM).

Palabras clave Mecanismos de evaluación de la atención de salud; Participación del paciente; Atención primaria de salud/normas; Prestación de atención de salud/normas; Calidad de la atención de salud; Organización Mundial de la Salud; Estudio comparativo; Europa (fuente: DeCS, BIREME).

الكلمات المفتاحية: آليات تقييم الرعاية الصحية، مساهمة المرضى، الرعاية الصحية الأولية، معايير الرعاية الصحية الأولية، إيتاء الرعاية الصحية، معايير إيتاء الرعاية الصحية، جودة الرعاية الصحية، منظمة الصحة العالمية. (المصدر: رؤوس الموضوعات الطبية المكتب الإقليمي لشرق المتوسط).

Bulletin of the World Health Organization 2004;82:106-114.

Voir page 112 le résumé en français. En la página 113 figura un resumen en español.

يمكن الاطلاع على الملخص بالعربية في صفحة 113.

Introduction

In 2000, WHO reported an international comparison of health system performance (1). On the basis of five measures of health system achievement, 191 Member States were ranked (2). Improvements of the health status of the population and the equality of health status distribution across the population are two important goals that address the core business of health care systems. The third goal is to ensure fairness in the financing of health care, with expenditure reflecting a patient's ability to pay rather than their risk of illness (3, 4). Health care systems should also be responsive to the legitimate expectations of populations

for non-health enhancing aspects, so level and distribution of responsiveness are the fourth and fifth goals. Responsiveness includes respect for dignity, confidentiality, and autonomy of persons, as well as client orientation (prompt service, quality of facilities, access to social support, and choice of provider). The measurement of the concepts of level of responsiveness and distribution of responsiveness are independent of the measurement of the three other goals (5).

As responsiveness addresses expectation and client orientation, it clearly is in the domain of patient views on health care. Donabedian defines quality as the degree to which health services

¹ NIVEL, Netherlands Institute for Health Services Research, PO Box 1568, 3500 BN Utrecht, Netherlands. Correspondence should be sent to Dr Kerssens, Senior Research Fellow at this address (j.kerssens@nivel.nl).

² NIVEL, Netherlands Institute for Health Services Research, and Department of Sociology and Department of Human Geography, Utrecht University, Utrecht, Netherlands.

³ Department of Gastroenterology and Hepatology, University Hospital Maastricht, Maastricht, Netherlands.

Ref. No. 02-001156

(Submitted: 6 January 03 – Final revised version received: 6 August 03 – Accepted: 19 August 03)

meet the needs, expectations, and standards of care of the patients, their families, and other beneficiaries of care (6). Expectations are studied very often in health quality research (7–9). For example, according to a model proposed by Babakus & Mangold (10), patients' judgements about quality are equal to their perception of quality minus their expectations (11, 12), but the measurement of expectation is characterized by diversity in approach in terms of definition, content, and measurement (13). In practice, expectations can refer to ideal health care, anticipated health care, or desired health care, and sometimes people do not even have explicit expectations (14). Zastowny et al. and Sixma et al. took the desired health care approach by concentrating on normative expectations, importance scores attached to these normative expectations (importance dimension), and actual experiences (performance dimension) (15, 16). In this model, expectations are reflected in statements such as "Health care providers should not keep me waiting for more than 15 minutes". Performance relates to cognitive awareness of the actual experience of the use of health care services: for example, "At my last appointment, they kept me waiting for more than 15 minutes". Although performance refers to an actual situation, importance scores attached to the expectation component refer to the fact that some features of health services are more significant than others. Quality of care judgments (Q) of individual patients (i) can be calculated by multiplying performance scores (P) by importance scores (I) of different health care aspects (j). As a formula, this equates to $Q_{ij} = P_{ij} \times I_{ij}$. Quality of care scores reflect the patients' view of health care and how patients want to be treated by health care professionals on quality aspects that are particularly relevant to them, taking into account the multidimensionality of the concept. A great deal of overlap exists between WHO's definition of responsiveness of health care systems — meeting the needs, or legitimate expectations, of the population for non-health enhancing dimensions of their interactions with the health system — and the way quality of care from the patients' perspective is defined by Sixma et al. (16). The distinction between quality of care from the patient's perspective and from the perspective of other stakeholders — such as health care providers (for example, care according to professional standards or protocols) or managers (for example, care based on efficiency) — needs to be kept in mind.

In order to select relevant quality of care aspects, Sixma et al. followed a general and disease-specific approach that included the use of focus group discussion (16). In this procedure, a series of instruments was tailored to the needs of various patient groups (for example, patients with chronic obstructive pulmonary disease (COPD), patients with rheumatism, and patients with diabetes) or was aimed at specific providers, for example, general practitioners or occupational therapy services (17), first in the Netherlands and later in other countries. These instruments are termed Quote instruments (QUality Of care Through patients' Eyes). The main difference between Quote instruments and the usual measures of patient satisfaction is that the former concentrate on (more objective) "reports" rather than (subjective) ratings of satisfaction or excellence, which makes them more interpretable and actionable for quality improvement purposes; they also usually consist of generic and category-specific quality of care aspects. Despite a great deal of overlap with the concept of responsiveness, Quote instruments differ in that they are not strictly limited to non-medical aspects and, depending on the expectations of specific patient categories, not always are all seven elements of responsiveness (dignity, confidentiality, autonomy, prompt attention, social support, basic amenities, and choice of provider) represented in each instrument.

With various Quote instruments being applied in different countries, patient views of what is important in evaluating quality of health care and their actual experience in the same areas are now available. We aimed to compare the Quote performance scores across several European countries (12) to gain insight into the similarities and differences in patient evaluations of the quality of care (another study has focused on the importance dimension (Groenewegen, PP, Kerssens JJ, Sixma HJ, Van der Eijk I, Boerma WGW, personal communication, 2003)). If country differences in people's experience of health care do exist, we wanted to find out whether these differences correlated to other health system performance measures, notably the level and distribution of responsiveness and overall performance according to WHO's rankings (18, 19). Positive correlations between these WHO measures and the mean Quote performance scores for each country would give evidence for the convergent validity of the Quote instruments and WHO's measurements. We aimed to find out whether patients in different European countries assess the quality of various aspects of care differently and whether health system performance measures correlate with patient evaluation of quality of care.

Methods

Materials

We collected data from various studies to produce a database for our analysis. We used the first Dutch Quote instruments (for disabled people, people with COPD or rheumatism, and elderly people), which contained 16 general importance and performance indicators, as a starting point for our database (16, 20–22). In the Supporting Clinical Outcomes in Primary Care for the Elderly project, the Quote-elderly instrument was translated into Danish, English, Finnish and German, after a double forward-backward procedure, and some participating groups collected data (23). A large contribution for our database came from an international study of patients with inflammatory bowel disease, which was carried out in eight countries (24). In our study, we included data from these studies on 10 generic questions (out of an original 16) that related to both general practitioners and specialists. We obtained additional material from studies from the United Kingdom (Quote-disabled (17)), Belarus (25), and the Ukraine.

Table 1 gives the number of respondents in each patient group and in each country, as well as the way in which patients were selected. In Belarus and the Ukraine, patients were selected at the general practitioner's office (opportunity sample), and in Belarus, a randomized sample of 500 general practice patients was selected from more than 2000 patients. In all participating countries, patients with inflammatory bowel disease were selected randomly from hospital lists. All patients with inflammatory bowel disease evaluated primary care as well as specialist care, but for our study, we used only their evaluations of primary care. Elderly patients were selected in Finland from primary health care centres' files and in Ireland from a home care organization; again, we only used evaluations of primary care. In the Netherlands, all patients, with the exception of patients with inflammatory bowel disease, were chosen randomly from general practitioners' files. Finally, disabled patients from the United Kingdom were chosen from the files of the occupational therapy service.

Answer formats for performance items included were: no (1), not really (2), on the whole, yes (3), and yes (4). All items referred to the general practitioner or primary care provider. For the

Table 1. Number of respondents in patient groups and countries

| Country | Patient group | Patient selection | No. of patients | |
|----------------|---------------------------------------|------------------------------------|-----------------|-------------|
| | | | Per group | Per country |
| Belarus | General practice | General practice office | 500 | 500 |
| Denmark | Inflammatory bowel disease | Hospital files | 102 | 102 |
| Finland | Elderly | Primary health care centres' files | 143 | 143 |
| Greece | Inflammatory bowel disease | Hospital files | 96 | 96 |
| Ireland | Inflammatory bowel disease | Hospital files | 57 | 73 |
| | Elderly | Home care organization's files | 16 | |
| Israel | Inflammatory bowel disease | Hospital files | 46 | 46 |
| Italy | Inflammatory bowel disease | Hospital files | 201 | 201 |
| Netherlands | Migrants | General practice files | 152 | 2873 |
| | Inflammatory bowel disease | Hospital files | 192 | |
| | Elderly | General practice files | 338 | |
| | Disabled | General practice files | 334 | |
| | Diabetes | General practice files | 681 | |
| | Chronic obstructive pulmonary disease | General practice files | 604 | |
| | Rheumatism | General practice files | 572 | |
| Norway | Inflammatory bowel disease | Hospital files | 93 | 93 |
| Portugal | Inflammatory bowel disease | Hospital files | 36 | 36 |
| Ukraine | General practice | General practice office | 490 | 490 |
| United Kingdom | Disabled | General practice files | 480 | 480 |
| Total | | | 5133 | 5133 |

sake of readability, the term "GP" was used in the phrasing of the Quote items. We excluded items that did not refer to primary care (for example, those that related to specialists) from our analyses. Box 1 shows the items included in our analysis.

Statistical analysis

All 5133 patients gave importance and performance ratings for each of a maximum of 10 items. We used these ratings as dependent variables in a series of statistical analyses, with patients nested hierarchically by country. We used variance analysis to divide patient ratings into between-country variance (σ^2_c), which indicated the variation among the 12 countries, and pooled-within variance (σ^2_p), which related to the variation between patients within the countries. The intra-class coefficient, ρ , is defined as $(\sigma^2_c / (\sigma^2_c + \sigma^2_p))$. When σ^2_c is close to zero, ρ is also close to zero. In that case, no variance exists between countries and all variation is between patients. When σ^2_p is close to zero, ρ is close to unity. In that case, little variation among patients is seen, which indicates relatively large differences among countries. An intra-class coefficient of 0.15 is considered quite high (26). In contrast with traditional forms of analysis of variance, in which factors have "fixed" effects, countries are considered to have "random" effects. Such a variance component model is preferred to traditional analysis if the number of categories exceeds 10 (27, 28). In this study, we included 12 countries. We analysed ten variance component models, one for each performance item. As is the case in the analyses of variance with fixed effect, covariates can be included in the analysis to correct for confounding variables. We corrected the intra-class correlation coefficient for age and sex, because these variables sometimes are associated with performance

scores (16, 29–31). Although we planned to correct for different patient groups as well, this turned out to be inappropriate because of the small number of countries for some patient groups.

WHO measures of achievement

WHO measured responsiveness in a key informant survey, which comprised 1791 interviews in 35 countries and yielded scores (from 0 to 10) on each of seven elements of responsiveness, as well as overall scores (5). Responsiveness was a weighted sum of a number of elements (Box 2). The elements of responsiveness are not all of equal importance, so the key informers ranked the seven elements on the basis of their importance (32). Their rankings

Box 1. Quote instrument items used in 12 countries

1. My GP has a good understanding of my problems
2. My GP allows me to contribute to decisions on treatment or help I receive
3. My GP always takes me seriously
4. My GP is always on time for appointments
6. My GP tells me about medicines prescribed in language that I can understand
5. My GP does not keep me waiting for more than 15 minutes
7. My GP prescribes medicines which are fully covered by the national health system or social services
8. My GP is always easy to reach by telephone
9. My GP makes sure that I can see a specialist within two weeks after being referred
10. My GP always tells other health and social care providers about the services I require

Box 2. Elements used to measure responsiveness in WHO survey (5)**Respect for persons, including:**

- Respect for dignity of person
- Confidentiality, or person's right to determine who has access to their personal health information
- Autonomy with respect to a person's participation in choices about their own health. This includes helping to select the treatment to be received or declining it

Client orientation including:

- Prompt attention: immediate attention in emergencies and reasonable waiting times for non-emergencies
- Amenities of adequate quality, including cleanliness, space, and hospital food
- Access to social support networks (family and friends) for people receiving care
- Choice of provider or freedom to select the individual or organization to deliver care

in decreasing order of importance were: prompt attention, dignity, autonomy, confidentiality, social support networks, basic amenities, and choice of provider. The differences in the relative weights given to the elements were not large (33).

WHO makes a distinction between the level of responsiveness and its distribution. When a system responds well, on average, to what people expect of it, the level of responsiveness is high. When it responds equally well to everyone, without discrimination or differences, the distribution of responsiveness is high (1).

Overall performance is a function of five specific achievements: level of health, distribution of health, financial fairness, level of responsiveness, and distribution of responsiveness.

Table 2 gives the ranking of the 12 countries we included in our study for the three WHO indicators. Denmark had the most responsive health care system of the 12 countries (rank 4 on level of responsiveness). Norway was second and was tied with one other country (not included in our study) at rank 7–8. The least responsive health care system was in the Ukraine, which was ranked 96. Distribution of responsiveness varied little between the countries. Nine of the countries in our study were tied at rank 3–38. Overall performance was highest in Italy, which was ranked second. We correlated these rankings with the mean Quote performance scores for each item in each country in our study.

Results

Table 3 shows descriptive statistics of the 10 Quote performance items we included in this study, by mean value.

“My GP always takes me seriously” had the highest mean, which was halfway between “on the whole, yes” and “yes” on the four-point scale. Less than 2% of patients gave this item the most negative scale point (data not shown). This item had the smallest variance at the country level.

“My GP does not keep me waiting for more than 15 minutes” had the lowest mean, which was halfway between “no, not really” and “on the whole, yes”. More than 25% of patients rated this item as “no” (data not shown). This item had the highest variance at the patient level.

“My GP prescribes medicines which are fully covered by the national health system or social services” had the largest variance at the country level.

The uncorrected intra-class correlation coefficients varied from low (0.027 for “My GP does not keep me waiting for more than 15 minutes”) to high (0.456 for “My GP prescribes medicines which are fully covered by the national health system or social services”): the intra-class correlation coefficient of 0.456 meant that the performance ratings for a pair of patients within a country correlated at 0.456. The sex- and age-adjusted intra-class correlation coefficients were lower on average (by 13%), but still ranged from low to high.

Table 4 shows mean performance scores for all items to give an impression of the variation between the countries. For example, “My GP has a good understanding of my problems” had the highest mean score in Greece and the lowest mean score in the Ukraine. “My GP allows me to contribute to the decisions on the treatment or help I receive” also had the highest mean in Greece, but the lowest mean was in Belarus. The difference between Greece and Belarus with respect to this item was more than one point on the four-point Likert scale. The differences between Israel and the United Kingdom with respect to “My GP prescribes medicines which are fully covered by the national health system or social services”) was 1.31.

To look at the consistency of patient evaluations across the different countries, we ranked the 10 performance items according to their mean value within each country. Table 5 gives the ranks for the nine countries in which all 10 items were available. Some differences among countries in the rankings emerged. For instance, in Denmark, “My GP tells me about the medication prescribed in language that I can understand” was ranked first, but in Italy, the same item was ranked at position six. In Portugal, “My GP has a good understanding of my problems” was ranked first, but in other countries, this particular item was somewhere between positions two and five. A general pattern was seen, however: except for Greece, “My GP always takes me seriously” was ranked first or second and “My GP does not keep me waiting for more than 15 minutes” was ranked low in every country.

Quote performance scores in relation to WHO indicators

Table 6 shows the correlation between the mean of the Quote performance items for each country and WHO's ranking of the

Table 2. WHO ranking of 12 countries according to level and distribution of responsiveness and overall performance (5)

| Country | Responsiveness rank | | Overall performance |
|----------------|---------------------|--------------|---------------------|
| | Level | Distribution | |
| Italy | 22–23 | 3–38 | 2 |
| Norway | 7–8 | 3–38 | 11 |
| Portugal | 38 | 55 | 12 |
| Greece | 36 | 3–38 | 14 |
| Netherlands | 9 | 3–38 | 17 |
| United Kingdom | 26–27 | 3–38 | 18 |
| Ireland | 25 | 3–38 | 19 |
| Israel | 20–21 | 3–38 | 28 |
| Finland | 19 | 3–38 | 31 |
| Denmark | 4 | 3–38 | 34 |
| Belarus | 76–79 | 46 | 72 |
| Ukraine | 96 | 63–64 | 79 |

Table 3. Descriptive statistics for performance items, ranked by mean score

| Item | Variance in score | | | Intra-class correlation coefficient | |
|---|-------------------|-------------------|------------------|-------------------------------------|------------------------|
| | Mean score | Between countries | Between patients | Uncorrected ^a | Corrected ^b |
| 3. My GP is always takes me seriously | 3.54 | 0.021 | 0.460 | 0.044 | 0.017 |
| 4. My GP always on time for appointments | 3.37 | 0.034 | 0.451 | 0.070 | 0.087 |
| 6. My GP tells me about medicines prescribed in language that I can understand | 3.34 | 0.023 | 0.758 | 0.029 | 0.036 |
| 1. My GP has a good understanding of my problems | 3.27 | 0.073 | 0.684 | 0.096 | 0.010 |
| 2. My GP allows me to contribute to decisions on treatment or help I receive | 3.16 | 0.065 | 0.774 | 0.077 | 0.011 |
| 8. My GP is always easy to reach by telephone | 3.09 | 0.115 | 0.868 | 0.117 | 0.099 |
| 9. My GP makes sure that I can see a specialist within two weeks of being referred | 2.86 | 0.162 | 0.804 | 0.168 | 0.134 |
| 10. My GP always tells other health and social care providers about the services I require | 2.84 | 0.185 | 0.819 | 0.184 | 0.212 |
| 7. My GP prescribes medicines that are fully covered by the national health system or social services | 2.82 | 0.726 | 0.865 | 0.456 | 0.478 |
| 5. My GP does not keep me waiting for more than 15 minutes | 2.50 | 0.036 | 1.278 | 0.027 | 0.023 |

^a Uncorrected for age and sex.

^b Corrected for age and sex.

level of responsiveness and overall performance. The distribution of responsiveness (as shown in Table 2) was not analysed, because too many ties were present in the rankings. The small number of countries meant that only very high correlations were statistically significant. Most correlations were positive. The highest correlation was found between “My GP has a good understanding of my problems” and responsiveness. The correlations between the different items and level of responsiveness were about the same as the correlations between the different items and the overall performance, because responsiveness and overall performance correlated highly themselves (0.94).

Discussion

Our main objective was to compare the Quote scores in different countries to gain knowledge about the similarities and differences in patient evaluations of the quality of primary care. Performance scores of the Quote instruments were used as indicators of patient evaluation of the quality of primary care. Intra-class correlation coefficients were calculated to measure differences between countries and ranged from low to very high. Sex- and age-adjusted intra-class correlation coefficients were only slightly lower. Little variation exists in some respects (for example, health care providers have a good understanding of patients' problems in all countries) and large variation in other respects (for example, with respect to the prescription of fully covered medication and communication between health care providers).

Sex and age were used as demographic variables to control for a different patient mix in the countries concerned. Sex and age together explained about 13% of the variation among the countries. After we controlled for these characteristics, a considerable amount of variation among the countries remained.

Most of the correlations between mean Quote scores per country and WHO performance measures were positive.

The answering formats of Quote performance items were: no (1), not really (2), on the whole, yes (3), and yes (4), so 1 and 2 are on the negative side of quality and 3 and 4 on the positive side. A country's mean score of 2.5 meant that about half of the patients scored ≤ 2 and half of the patients scored ≥ 3 . For some items, some countries scored ≤ 2.5 . “My GP prescribes medicines which are fully covered by the national health system or social services” scored low in Denmark, Portugal, and especially the United Kingdom. In terms of quality improvement, much is to be gained in these countries by the prescription of fully covered medication. “My GP always communicates with other health and social care providers about the services I require” was another quality aspect with relatively low scores: in Denmark, Italy, Norway, and Portugal, scores were even lower than 2.5. “My GP always takes me seriously” was the item with the highest mean scores. In none of the countries was this performance score < 3.70 , which implies that almost all patients score 4 on the scale from 1 to 4. This means that this quality aspect is handled well in the countries compared.

Table 4. Mean scores of ten performance items by country

| Item ^a | Belarus | Denmark | Finland | Greece | Ireland | Israel | Italy | Netherlands | Norway | Portugal | Ukraine | United Kingdom |
|-------------------|---------|---------|----------------|--------|---------|--------|-------|-------------|--------|----------|---------|----------------|
| 1. | 3.06 | 3.67 | — ^b | 4.00 | 3.31 | 3.83 | 3.70 | 3.62 | 3.63 | 3.50 | 3.02 | 3.74 |
| 2. | 2.74 | 3.56 | 3.69 | 3.75 | 3.19 | 3.53 | 3.40 | 3.56 | 3.50 | 2.75 | 2.82 | 3.39 |
| 3. | 3.71 | 3.78 | 3.95 | 3.75 | 3.75 | 4.00 | 3.75 | 3.76 | 4.00 | 3.75 | 3.70 | 3.80 |
| 4. | 3.67 | 3.56 | 3.82 | 3.75 | 3.69 | 3.83 | 3.56 | 3.85 | 3.75 | 2.75 | 3.71 | 3.54 |
| 5. | 2.66 | 2.67 | — | 3.00 | 2.88 | 2.83 | 2.09 | 2.57 | 2.75 | 2.25 | 2.74 | 2.04 |
| 6. | 3.43 | 3.89 | 3.5 | 4.00 | 3.34 | 3.67 | 3.36 | 3.52 | 4.00 | 3.00 | 3.50 | 3.64 |
| 7. | — | 2.50 | — | 3.50 | 2.97 | 3.50 | 3.46 | 3.52 | 3.63 | 1.75 | — | 1.19 |
| 8. | 3.47 | 3.19 | 3.53 | 3.75 | 3.68 | 3.67 | 3.24 | 3.44 | 2.75 | 2.25 | 2.85 | 3.56 |
| 9. | — | 3.33 | — | 3.45 | 2.88 | 3.50 | 2.75 | 3.60 | 3.35 | 2.00 | — | 2.89 |
| 10. | 3.09 | 2.17 | 3.76 | 3.75 | 2.89 | 2.76 | 2.49 | 3.62 | 2.43 | 2.00 | 2.99 | 3.2 |

^a See Box 1 for item labels.

^b Item missing.

Table 5. Ranking of items^a by mean performance within each country for which all 10 items available

| Country | Rank | | | | | | | | | |
|----------------|------|---|---|----|---|----|----|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Denmark | 6 | 3 | 1 | 2 | 4 | 9 | 8 | 5 | 7 | 10 |
| Greece | 4 | 6 | 8 | 1 | 3 | 2 | 7 | 10 | 5 | 9 |
| Ireland | 3 | 4 | 8 | 6 | 1 | 2 | 10 | 7 | 9 | 5 |
| Israel | 3 | 6 | 8 | 7 | 1 | 4 | 9 | 2 | 5 | 10 |
| Italy | 3 | 7 | 1 | 4 | 2 | 6 | 8 | 9 | 10 | 5 |
| Netherlands | 4 | 3 | 1 | 10 | 6 | 9 | 2 | 7 | 8 | 5 |
| Norway | 3 | 6 | 7 | 1 | 4 | 2 | 9 | 8 | 5 | 10 |
| Portugal | 1 | 3 | 2 | 6 | 4 | 10 | 5 | 8 | 9 | 7 |
| United Kingdom | 3 | 1 | 6 | 8 | 2 | 4 | 10 | 9 | 5 | 7 |

^a Items described in Box 1.

The ranking of items by mean performance in each country (Table 5) showed some differences between the countries. The Quote instruments have a performance dimension and an importance dimension, and Groenewegen et al. reported on the ranking of the corresponding set of Quote importance items within the same countries (Groenewegen, PP, Keressens JJ, Sixma HJ, Van der Eijk I, Boerma WGW, personal communication, 2003). The performance items showed much more diversity among the countries than the importance items. For instance, “My GP should not keep me waiting for more than 15 minutes” was ranked last in all countries, while “My GP should always take me seriously” was ranked high in all countries. This seems to indicate that patient views about what is important in the quality of primary care are much more consistent among different countries compared with patient experience in the same areas.

Study limitations

Both the external and internal validity of this study have limitations.

External validity

The Quote scores are taken as an indicator for health care quality in 12 countries; however, in most of these countries only one patient group existed. Differences of disease characteristics could not be tested because of the small number of patient groups for all countries except the Netherlands. Furthermore, these patient evaluations relate to the general practitioners or primary care providers. Our quality indicators, therefore, are not nationwide indicators. A further step needed to enhance external validity is inclusion of not just general practitioner services but also other health care providers and institutions.

Internal validity

The small number of countries meant that the power to detect relations between the patient evaluations of health care quality and care system-related variables was very low. Only large correlation coefficients (≥ 0.60) are statistically significant. Table 6 showed some high correlations between two measures of health system achievements, responsiveness, and overall performance

Table 6. Correlations between mean performance for each country and 12 WHO indicators

| Item ^a | Responsiveness level | Overall performance |
|---|----------------------|---------------------|
| 1. My GP has a good understanding of my problems | 0.86 ^a | 0.85 ^a |
| 2. My GP allows me to contribute to decisions on treatment or help I receive | 0.82 ^a | 0.65 ^b |
| 3. My GP always takes me seriously | 0.72 ^a | 0.43 |
| 4. My GP always on time for appointments | 0.07 | -0.08 |
| 5. My GP does not keep me waiting for more than 15 minutes | -0.10 | -0.19 |
| 6. My GP tells me about medicines prescribed in language that I can understand | 0.53 | 0.26 |
| 7. My GP prescribes medicines that are fully covered by the national health system or social services | 0.28 | 0.19 |
| 8. My GP always easy to reach by telephone | 0.19 | 0.06 |
| 9. My GP makes sure that I can see a specialist within two weeks of being referred | 0.58 | -0.31 |
| 10. My GP always tells other health and social care providers about the services I require | -0.11 | -0.12 |
| Mean | 0.33 | 0.11 |

^a $P < 0.05$.

^b $P < 0.01$.

and the mean Quote performance scores per country, particularly with respect to “My GP has a good understanding of my problems” and “My GP allows me to contribute to the decisions on the treatment or help I receive”. This latter item relates to the aspect of autonomy, which is also an important feature of WHO responsiveness (33, 34).

Conclusion

The world health report 2000 has been criticized on its assumption of a universal value base for all health care systems, because concepts such as responsiveness may be valued differently in different countries (35). If this line of reasoning is followed, responsiveness could be supposed to be more in the domain of values than performance. The Quote instruments also have an importance dimension, so we analysed the correlation between

responsiveness and the mean Quote importance scores per country. Except for “My GP is always on time for appointments”, all the correlations between responsiveness and mean Quote importance items were lower than the correlations between responsiveness and the mean Quote performance items.

Our study supports the conclusion that responsiveness is more in the domain of health care quality than in the domain of patients’ or WHO key informants’ views and values of the health care system. ■

Funding: This study was financially supported by ZonMw “Netherlands Organization for Health Research and Development” project number 240-20-205.

Conflicts of interest: none declared.

Résumé

Comparaison des évaluations de la qualité des soins de santé par les patients et mise en corrélation avec les mesures OMS d’accomplissement dans 12 pays européens

Objectif Dégager les similitudes et les différences entre les évaluations de la qualité des soins primaires par les patients dans 12 pays européens et mettre ces évaluations en corrélation avec les mesures OMS de la performance des systèmes de santé (la réactivité par exemple) de ces mêmes pays.

Méthodes Les évaluations par les patients provenaient d’une série d’instruments Quote servant à mesurer la qualité des soins primaires. On a pris en compte différents groupes de recherche représentant au total un échantillon de 5133 patients dans 12 pays : le Bélarus, le Danemark, la Finlande, la Grèce, l’Irlande, Israël, l’Italie, la Norvège, les Pays-Bas, le Portugal, le Royaume-Uni et l’Ukraine. On a calculé les corrélations intraclasse de 10 éléments Quote pour mesurer les différences entre pays. Les mesures de la performance du *Rapport sur la santé dans le monde, 2000* faites dans les mêmes pays ont été corrélées avec les notes moyennes obtenues selon la méthode Quote.

Résultats Les coefficients de corrélation intraclasse sont faibles ou élevés selon que la variation est faible entre les pays pour certains aspects (par exemple les dispensateurs de soins primaires comprennent bien les problèmes des patients dans tous les pays) ou forte pour d’autres (la prescription de médicaments et la communication entre les dispensateurs de soins primaires, par exemple). La plupart des corrélations entre la note moyenne obtenue par les pays avec les instruments Quote et les mesures OMS de la performance étaient positives. La corrélation la plus forte (0,86) était celle entre la compréhension des problèmes des patients par les dispensateurs de soins primaires et la réactivité selon les critères OMS.

Conclusion Les évaluations de la qualité des soins primaires par les patients font ressortir d’importantes différences entre les pays et sont positivement corrélées aux mesures OMS de la performance des systèmes de santé.

Resumen

Comparación de las evaluaciones de la calidad de la atención sanitaria realizadas por los pacientes en relación con los índices OMS del desempeño en 12 países europeos

Objetivo Conocer con más detalle las similitudes y diferencias de las evaluaciones de la calidad de la atención primaria realizadas por los pacientes en 12 países europeos, y correlacionar dichas evaluaciones con los índices OMS (por ejemplo la responsividad) del desempeño de los sistemas de salud de esos países.

Métodos Las evaluaciones de los pacientes se obtuvieron mediante una serie de instrumentos de Quote diseñados para medir la calidad de la atención primaria. Diversos grupos de investigación proporcionaron una muestra total de 5133 pacientes de 12 países: Belarús, Dinamarca, Finlandia, Grecia, Irlanda, Israel, Italia, Noruega, los Países Bajos, Portugal, el Reino Unido y Ucrania. Se calcularon las correlaciones intraclase de 10 elementos de Quote para medir las diferencias entre países. Las puntuaciones medias de Quote se correlacionaron con los índices del desempeño publicados en el *Informe sobre la salud en el mundo 2000* para esos mismos países.

Resultados Los coeficientes de correlación intraclase fueron

entre bajos y muy altos, lo que indica que hay muy pocas diferencias entre los países en algunos aspectos (por ejemplo, los proveedores de atención primaria tienen una buena comprensión de los problemas de los pacientes en todos los países) y grandes diferencias en otros (por ejemplo en lo que atañe a la prescripción de medicamentos y la comunicación entre dispensadores de atención primaria). La mayoría de las correlaciones entre las puntuaciones medias de Quote por países y los índices OMS del desempeño fueron positivas. La correlación más alta (0,86) fue la observada entre la comprensión de los problemas de los pacientes por parte de los dispensadores de atención primaria y el grado de responsividad según la OMS.

Conclusión Las evaluaciones de la calidad de la atención primaria realizadas por los pacientes difirieron considerablemente entre los países y resultaron estar positivamente relacionadas con los índices OMS de desempeño de los sistemas de atención de salud.

ملخص

مقارنة بين تقييم المرضى لجودة الرعاية الصحية وعلاقتها بمقاييس منظمة الصحة العالمية للأداء في ١٢ بلداً أوروبياً

الشديدة الارتفاع، مما يشير إلى اختلاف طفيف بين البلدان في بعض المجالات (فعلى سبيل المثال، لدى القائمين على إيتاء الرعاية الصحية الأولية تفهم جيد لمشاكل المرضى في جميع البلدان) مع اختلافات كبيرة في بعض الجوانب (مثل ما يتعلق بوصف الأدوية والتواصل مع القائمين على إيتاء الرعاية الصحية الأولية). لقد كانت جميع الارتباطات باحراز (كوت) الوسطية في كل بلد مع مقاييس منظمة الصحة العالمية للأداء إيجابية. وقد كان أعلى ارتباط (٠,٨٦) بين فهم القائمين على إيتاء الرعاية الصحية لمشكلات المرضى والتصدى لها وفقاً لمعايير منظمة الصحة العالمية.

النتيجة: لقد أظهر تقييم المرض لجودة الرعاية الصحية الأولية اختلافاً كبيراً فيما بين البلدان وهو يتعلق بشكل إيجابي بمقاييس منظمة الصحة العالمية لأداء نظم الرعاية الصحية.

الهدف: اكتساب الخبرة حول أوجه التشابه والاختلاف في تقييم المرضى لجودة الرعاية الصحية في ١٢ بلداً أوروبياً وعلاقة تقييم المرضى مع مقاييس منظمة الصحة العالمية لأداء النظم الصحية (مثل الاستجابة) في هذه البلدان.

الطريقة: لقد تم اشتقاق تقييم المرضى من سلسلة من أدوات (كوت) صممت لقياس جودة الرعاية الصحية. وقد درست مجموعات بحث مختلفة عينات بلغ إجمالي تعدادها ٥١٣٣ مريضاً ينتمون إلى ١٢ بلداً هي: بيلاروس، الدنمارك، فنلندا، اليونان، إيرلندا، فلسطين، إيطاليا، هولندا، النرويج، البرتغال، المملكة المتحدة وأوكرانيا. وقد تم حساب عشرة من عناصر (كوت) وارتباطها بالفئات لقياس الاختلاف بين البلدان. إن قياسات الأداء في التقرير الخاص بالصحة في العالم لعام ٢٠٠٠ قد ربطت في نفس البلدان مع الأحراز الوسطية ل(كوت). **الموجودات:** لقد تراوحت قيم معامل الارتباط بين الفئات بين المنخفضة إلى

References

1. World Health Organization. *The world health report 2000 — Health systems: improving performance*. Geneva: World Health Organization; 2000.
2. Taipale V. There is a need for assessment and research in health policy. In: Häkkinen U, Ollila E, editors. *Themes from Finland. The world health report 2000. What does it tell us about health systems? Analyses by Finnish experts*. Helsinki: National Research and Development Centre for Welfare and Health (Stakes); 2000. p. 1-2.
3. McKee M. Measuring the efficiency of health systems. The world health report sets the agenda, but there's still a long way to go. *BMJ* 2001;323:295-6.
4. Murray CJL, Knaul F, Musgrove P, Xu K, Kawabata K. *Defining and measuring fairness in financial contribution to the health systems. GPE discussion paper series: no. 24*. Geneva: World Health Organization. WHO document EIP/GPE/FAR.
5. Valentine NB, Silva A de, Murray CJL. *Estimating responsiveness level and distribution for 191 countries: methods and results. GPE discussion paper series: no. 22*. Geneva: World Health Organization. WHO document EIP/GPE/FAR.
6. Donabedian A. Twenty years of research on quality of medical care, 1965-1984. *Evaluation and the Health Professions* 1985;8:243-65.
7. Pascoe GC. Patient satisfaction in primary health care: a literature review and analysis. *Evaluation and Program Planning* 1983;6:185-210.
8. Strasser S, Aharony L, Greenberger D. The patient satisfaction process: moving toward a comprehensive model. *Medical Care Review* 1993;50:219-48.
9. Van Campen C, Sixma HJ, Friele RD, Kerssens JJ, Peters L. Quality of care and patient satisfaction: a review of measuring instruments. *Medical Care Research and Review* 1995;52:109-33.
10. Babakus E, Mangold WG. Adapting the SERVQUAL scale to hospital services: an empirical investigation. *Health Services Research* 1992;26:767-86.
11. Parasuraman A, Zeithaml VA, Berry LL. A conceptual model of service quality and its implications for future research. *Journal of Marketing* 1985;49:41-50.
12. Parasuraman A, Zeithaml VA, Berry LL. SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing* 1988;64:12-40.
13. Staniszewska S, Ahmed L. The concepts of expectation and satisfaction: do they capture the way patients evaluate their care? *Journal of Advanced Nursing* 1999;29:364-72.

14. Thompson AGH, Suñol R. Expectations as determinants of patient satisfaction: concepts, theories and evidence. *International Journal of Quality of Health Care* 1995;7:127-41.
15. Zastowny TR, Stratmann WC, Adams EH, Fox ML. Patient satisfaction and experience with health services and quality of care. *Quality Management in Health Care* 1995;3:50-61.
16. Sixma HJ, Kerssens JJ, van Campen C, Peters L. Quality of care from the patients' perspective: from theoretical concept to a new measuring instrument. *Health Expectations* 1998;1:82-95.
17. Calnan S, Sixma HJ, Calnan MW, Groenewegen PP. Quality of local authority occupational therapy services: developing an instrument to measure the user's perspective. *British Journal of Occupational Therapy* 2000;63:155-62.
18. Tandon A, Murray CJL, Lauer JA, Evans DB. *Measuring overall health systems performance for 191 countries. GPE discussion paper series: no. 30.* Geneva: World Health Organization. WHO document EIP/GPE/EQC.
19. Evans DB, Tandon A, Murray CJL, Lauer JA. Comparative efficiency of national health systems: cross national econometric analysis. *BMJ* 2001;323:307-10.
20. Sixma HJ, van Campen C, Kerssens JJ, Peters L. Quality of care from the perspective of elderly people: the QUOTE-elderly instrument. *Age and Ageing* 2000;29:173-8.
21. Van Campen C, Sixma H, Kerssens JJ, Peters L. Assessing non-institutionalized asthma and COPD patients' priorities and perceptions of quality of health care: the development of the QUOTE-CNSLD instrument. *Journal of Asthma* 1997;34:531-8.
22. Van Campen C, Sixma HJ, Kerssens JJ, Peters L, Rasker JJ. Assessing patients' priorities and perceptions of the quality of health care: the development of the QUOTE-rheumatic-patients instrument. *British Journal of Rheumatology* 1998;37:362-8.
23. Supporting Clinical Outcomes in Primary Care for the Elderly. *Common evaluation protocol. Introducing and evaluating the use of health outcome measures in primary care for elderly people.* Brussels: Biomedical & Health programme of the European Communities; 1998.
24. Van der Eijk I, Sixma H, Smeets T, Veloso FT, Odes S, Montague S, et al. Quality of health care in inflammatory bowel disease: development of a reliable questionnaire (QUOTE-IBD) and first results. *American Journal of Gastroenterology* 2001;96:3329-36.
25. Boerma WGW, Schellevis FG, Rousovitch V. *Going ahead with primary care and general practice in Belarus.* Utrecht: Netherlands Institute for Health Services Research; 2002.
26. Goldstein H. *Multilevel statistical models.* New York: Halsted Press; 1995.
27. Searle SR, Casella G, McGullogh CE. *Variance components.* New York: Wiley; 1992.
28. Snijders TAB, Bosker RJ. *Multilevel analysis: an introduction to basic and advanced multilevel modelling.* London: Sage; 1999.
29. Fox JG, Storms DM. A different approach to sociodemographic predictors of satisfaction with health care. *Social Science and Medicine* 1981;15A:557-64.
30. Weiss GL. Patient satisfaction with primary medical care: evaluation of sociodemographic and predispositional factors. *Medical Care* 1988;26:383-92.
31. Ovreteit J. *Health service quality: an introduction to quality methods for health services.* London: Blackwell Scientific; 1992.
32. Gakidou E, Murray CJL, Frenk J. *Measuring preferences on health system performance assessment. GPE discussion paper series: no. 20.* Geneva: World Health Organization. WHO document EIP/GPE.
33. Darby C, Valentine N, Murray CJL, de Silva A. *World Health Organization (WHO): strategy on measuring responsiveness. GPE discussion paper series: no. 23.* Geneva: World Health Organization. WHO document EIP/GPE/FAR.
34. De Silva A. *A framework for measuring responsiveness. GPE discussion paper series: no. 32.* Geneva: World Health Organization. WHO document EIP/GPE/EBD.
35. Mooney G, Wiseman V. World Health Report 2000. Challenging a world view. *Journal of Health Services Research and Policy* 2000;5:198-9.