Published by Oxford University Press in association with The London School of Hygiene and Tropical Medicine *Health Policy and Planning* 2007;**22**:294–302
 Advance Access publication 27 July 2007 doi:10.1093/heapol/czm020

A cross-national comparison of the quality of clinical care using vignettes

John W Peabody[1]* and Anli Liu[2]

Accepted 2 April 2007

In studies comparing clinical practice to evidence-based standards, researchers have found that quality of care is inconsistently provided to different segments of the population in both developing and developed countries. To test the hypothesis that quality of care varies widely within different countries, we conducted a prospectively designed evaluation of quality for three common clinical conditions: diarrhoea, tuberculosis and prenatal care. Five countries participated in the study: China, the Philippines, Mexico, El Salvador and India. Within each country, physicians were randomly selected from tertiary care hospitals, district level hospitals, and public and private outpatient clinics. A total of 488 previously validated case vignettes were administered to 300 participating physicians. Vignettes were scored according to evidence and expert based quality criteria. We used a random effects model to estimate the associations between quality scores by case, physician characteristics, study site, and country. We found that average quality of care was low (61.0%), but there exists a wide variation in overall quality (30–93%). While there was little difference in average quality scores *between* countries (60.2 to 62.6%), variation *within* countries was broad. The wide variation was consistent across facility type, medical condition and domain of care. We also found that younger, female, tertiary care and specialist physicians performed better than their counterparts. We conclude that some physicians provide exceptional care even in the setting of limited resources. Furthermore, poor quality can be addressed by health policy planners by directing remediation toward the lowest performers.

Keywords Health policy, quality of clinical care, quality measurement, clinical vignettes, international health, tuberculosis, diarrhoea, prenatal care

[1] Professor and Deputy Director, Institute for Global Health, University of California, San Francisco, USA.

[2] School of Medicine, c/o Office of Curricular Affairs, Box 0410, University of California, San Francisco, CA 94143–0410, USA. Email: anl9058@nyp.org

* Corresponding author. Institute for Global Health, 50 Beale Street, Suite 1200, San Francisco, CA 94105, USA. Tel: +1 415 597-8202. Fax: +1 415 597-8299. E-mail: Peabody@psg.ucsf.edu

KEY MESSAGES

- Quality of clinical care is a critical measure of health system performance in both developed and developing countries. Measuring and reporting on quality has proven to be an impediment to bettering quality.
- We compared the quality of care in five countries using clinical vignettes, a validated and increasingly used method for benchmarking the quality of clinical practice.
- Overall, the average quality was poor and much the same in the five countries studied. Within countries, however, there was wide variation in quality scores. Across countries, physicians who were younger, women or who worked in tertiary facilities were more likely to provide better care. Quality was not associated with income.
- Overall care quality could be improved by directing remediation efforts towards the poorest providers in all countries and at all levels of care.

Introduction

In studies comparing clinical practice with evidence-based standards, researchers have found that quality of care is inconsistently provided to large segments of the population (McGlynn *et al*. 2003). While most of these comparisons have been made in developed countries, evidence is emerging that variation exists in developing countries as well (Loevinsohn *et al*. 1995; World Bank 2004; Peabody *et al*. 2006a). Although data are limited, there appears to be striking variation in quality within individual developing countries—across facilities, among providers, and between specialists and non-specialists (Walker *et al*. 1988; Beracochea *et al*. 1995; Peabody *et al*. 1998; Nolan *et al*. 2001; Weinberg *et al*. 2001; Dumont *et al*. 2002). This observation contradicts established notions that care is 'better' in some countries than in others. With so much variability *within* countries, variation *among* countries does not seem meaningful.

To test the hypothesis that quality of care varies widely within different countries for the same clinical conditions, we measured the quality of clinical practice for diarrhoea, prenatal care and tuberculosis among physicians in five countries with different systems of organizing and financing care. We looked at one element of quality, the process of care, or 'the set of activities that take place between and within practitioners and patients' (Donabedian 1980), which can be measured by the degree to which physicians comply with predefined indicators (Wennberg *et al*. 2002). Because process occurs every time a patient and a provider come together, it is easier to measure than adverse health outcomes. It is also relatively easy to observe in a clinical setting, and specific changes in clinical practice can be evaluated for changes in outcome (Peabody 1995).

Measuring quality

Measuring the level of quality of care in both rich and poor countries alike has been difficult. For cross-national comparisons, methods must be identified to measure process of care across a wide range of demographics and systems. In this study, we used clinical vignettes—written case scenarios administered to doctors—because we believe they offer the most advantages and the fewest drawbacks. By design, they are completely case-mix adjusted and thus can account for variations in individual health status that would otherwise confound a cross-national survey.

Although vignettes are not the same as actual clinical practice, they have been validated in two prospective studies as a measure of clinical practice, using the 'gold standard' of unannounced standardized patients. In these studies, standardized patients (SPs) who were experienced actors typically working in SP medical education programmes, were introduced unannounced into clinical practice. Immediately after each visit, the SPs completed a checklist, indicating what actions the doctor took. The checklist generated a SP score. The SP score was compared with the score of the corresponding clinical vignette, which used the identical scoring criteria, completed by the same physician. Each SP visit also generated a medical record that was retrieved. The retrieved abstracted chart was scored using the above-mentioned criteria, which generated a third score. The SP checklists, medical records from the SP visits, and corresponding vignettes completed by the same physicians were then compared to determine the validity of each measurement method. The results from the two studies consistently showed that vignettes produced scores closer to the gold standard of SPs than did the abstracted medical charts ($P < 0.05$). Analyses from the two separate studies confirmed this finding to be robust across sites, case, complexity and level of training ($P < 0.05$) (Peabody *et al*. 2000; Peabody *et al*. 2004).

Methods

We conducted a prospectively designed evaluation of quality of care among randomly selected physicians located in four health care settings: tertiary care hospitals, district level hospitals, and public and private outpatient clinics. Five countries participated in the study: China, India, the Philippines, El Salvador and Mexico. Data were collected between June and August of 2003. Because it was not practical or feasible to collect a complete or even a random sample of countries, the countries were purposively selected to represent a broad spectrum of health care in the developing world that exhibited a wide variation in basic statistics such as population size, overall health status and health care systems, including financing.

The three clinical conditions (diarrhoea, tuberculosis and prenatal care) were selected because (1) they represent

Table 1 Tertiary hospitals used as reference facilities

Country	Tertiary hospital
China	Peking Union Medical College
El Salvador	Instituto Nacional de la Seguridad Social
India	All India Institute of Medical Sciences
Mexico	Instituto Nacional de Ciencias Medicas y Nutricion
Philippines	Philippine General Hospital

common outpatient conditions in the developing world; (2) they have a high associated burden of disease; (3) higher-quality care for each has led to demonstrable improvements in health outcomes; and (4) appropriate treatment technology was available in all sites.

Eligibility criteria for physician participation were broad. Doctors were eligible if they had a license to practice medicine; had graduated from a nationally accredited medical school; provided relevant specialist or primary care; practiced predominantly in one of the four specified settings; and voluntarily agreed to participate. Non-medical doctors were excluded from the study because of the wide variation in licensing and practice that exists between countries.

Our stratified sample frame of facilities began with the selection of the public tertiary care hospital within each country with a reputation for providing the best medical care (see Table 1). These tertiary facilities all accepted referrals from other urban district-level hospitals. From each tertiary hospital, we obtained a complete list of specialists in obstetrics, paediatrics and internal medicine. Eight doctors from each specialty were randomly selected and asked to participate by completing the vignette related to their area of specialty.

To sample referring district hospitals, we created a roster of the referring district hospitals situated within a 100 km radius (60 miles) of the tertiary hospital. These non-randomly selected hospitals were included if they provided inpatient care, were within the defined radius, were publicly financed, and regularly referred specialty and advanced care cases to the identified tertiary facility.

Similarly, to sample public and private doctors, we generated rosters of public and private practitioners. Public and private clinics had to provide outpatient care, be located within 100 km of the referral and district hospital, and regularly refer their patients to the district or (in some cases) to the leading referral hospital. We asked regional health authorities, obtained official rosters of facilities, looked in the phone book, and contacted medical societies to generate these rosters.

The rosters, generated from the snowball technique, yielded a large number of providers eligible for the study from each type of facility. From each group's roster, we randomly selected eight doctors. If a doctor was not available when we set the time to complete the case or visited the clinics, we substituted the next name on the roster. If a doctor did not provide obstetrical care, we randomly selected an obstetrician from the rosters of the district level facilities, public clinics and private clinics. The eight generalists were asked to complete vignettes for all three cases (if doctors did not handle obstetrical cases, they were evaluated for two cases and an obstetrician was evaluated for the prenatal case).

Data collection

Doctors were solicited in person and by telephone contact. Participation was voluntary and completely confidential. No names were linked with responses. Of those asked to participate, only 5–10% refused, depending on the country. In each country, we collaborated with a local site/project coordinator to: obtain approval from a local Institutional Review Board (IRB) and/or local authorities; develop a roster of doctors at the four sites; begin recruitment at the four sites where doctors would complete clinical vignettes; and establish a schedule for data collection. The coordinators also had the responsibility of identifying and obtaining ethics approval from the institution in each country that had jurisdiction over research and research structures. The overall study was approved by the Institutional Review Board (IRB) at the University of California, San Francisco.

Vignettes

We administered 488 paper-and-pencil vignettes to 300 participating doctors. To preserve data integrity, doctors were not told the nature of the condition or disease involved in the vignettes ahead of time. Eight vignettes were not completed and were dropped from the analysis, leaving a total of 480.

The vignettes were organized in five sections or skill domains designed to recreate the sequence of a typical patient visit: presentation of the patient and the patient's medical complaint, followed by (1) history taking, (2) physical examination, (3) radiological or laboratory tests ordered, (4) diagnosis, and (5) treatment plan. In each domain, doctors were asked open-ended questions about what information they would obtain from the patient. Once physicians completed a domain, they could not return to a previous domain to revise their answers or use the new information to change or improve their previous answers.

Scoring criteria

To create the scoring criteria, we conceptualized quality as the comprehensive provision of services for a given clinical case in a manner leading to better outcomes for individuals and populations (Peabody *et al*. 2004). We identified candidate quality criteria for a full range of provider activities that make up the process of outpatient primary care and have been shown in the evidence-based quality-of-care literature to lead to better outcomes. This involved describing the complete set of actions that would need to be undertaken by physicians taking care of patients. Rather than using single-point measures, such as screening for a co-morbidity or recommending the right antibiotic, the vignettes comprehensively assess all of the necessary items required to make a diagnosis and implement correct therapy. Both overall and domain scores can be generated to produce measures that more completely capture physician performance.

The quality criteria for each condition were derived from three sources: an evidence-based literature search on the clinical practices that lead to better health outcomes; United

States and international clinical guidelines; and local expert panels of academic and community physicians comprising both generalists and specialists. We used recommendations by the expert panels to finalize the master criteria list derived from the literature and guidelines (Glassman *et al*. 2000). Items experts considered most critical were assigned a weight of 1.0. Individual items deemed less important, such as multiple physical examination items related to a single construct, were grouped into categories, implicitly assigning them lower weights, typically 0.50 or 0.33.

Scoring was done by two trained abstractors blinded to physician identity. They reviewed each vignette answer sheet and indicated on a scoring form items the physician completed successfully. The raw item scores were aggregated into category scores. These weighted scores (an average of 41 categories per case) were totalled and divided by the total possible score, generating a percentage correct score for each vignette. For further sub-analyses, each scoring category was assigned to one of the five domains of care.

Vignette administration

Vignettes were first translated into Spanish and Chinese and then back-translated to check for consistency, using different pairs of bilingual physicians to ensure linguistic and technical accuracy, including local variations in medical terms and care. Before scoring, the responses were translated by the same bilingual physicians. Ten per cent of the translations were randomly retranslated to ensure accuracy and consistency. Another 10% were randomly audited and scored a second time to ensure scoring and accuracy. The error rate was 5%, well within usual standards observed from abstraction (Zadnik *et al*. 1998; Labelle and Swaine 2002).

Before completing the vignettes, each doctor was asked to complete a short survey concerning individual characteristics such as age, gender, level of education, salary, length of time in practice, and size of patient list. Vignette administration/data collection was carried out over approximately 1 week per country. A standardized implementation was used by each site coordinator to ensure consistency across sites. We administered vignettes on site at the hospital or clinic during scheduled appointments to groups of physicians. The vignettes took approximately 15–20 minutes each to complete.

Analysis

We used an ANOVA model to estimate the associations between overall scores by case, physician characteristics, study site, and country. To account for the different means between cases, the quality scores were normalized for the model. We developed a multivariate regression model to evaluate the impact of doctor characteristics on quality score, and a random effects model to account for possible clustering of scores. Although tests for clustering were not significant, we use the latter model in reporting our results. The independent variables influencing quality were attributed to either the doctor or the system of care and were modelled as follows:

$$Quality_i = \alpha + \beta age_i + \sum_{j=1}^{J} \gamma_j P_{ji} + \sum_{k=1}^{K} \omega_k F_{ki} + \sum_{l=1}^{L-1} \theta_l C_{li} + u_i$$

where $Quality_i$ is a quality measure for the ith physician, P_{ji} is a measure of the jth characteristic of the ith physician, F_{ki} measures the kth characteristic of the ith physician's facility, C_{li} is a dummy variable for the lth country, and u_i is a random disturbance term. No salary values were imputed. All statistical evaluation was done using Stata® 8.0.

Findings

The average quality of care for the 480 cases completed in the study was 61.0%. The scores ranged from 30% to 93% for all cases (Figure 1). Cross-country variation in average quality scores ranged from 60.2% to 62.6% (Figure 2). The range within each country, however, was broad (Figure 2). The difference between the bottom 5% and the top 5%, for example, ranged from 43% in the Philippines to 51% in China. This wide variation was consistent across facility type (Figure 3) and condition ($P < 0.05$, not shown). Analysis of domain of care again showed wide variation, especially for testing and treatment (Figure 4).

We modelled physician characteristics to determine which ones were associated with higher vignette scores. We included age, gender, specialty, type of facility and the number of patients the physician followed. Overall, the characteristics of

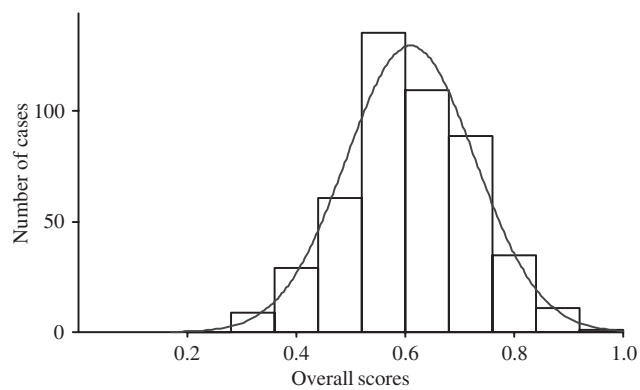

Figure 1 Distribution of overall score

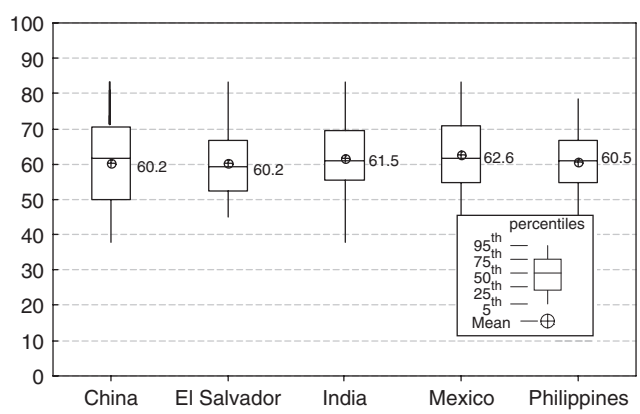

Figure 2 Comparison of overall scores across countries

the 300 doctors who completed the 480 vignettes were similar across countries (Table 2).

We found that country of practice did not predict differences in quality. However, younger doctors and female doctors had significantly higher quality scores. Tertiary facilities, followed by private clinics, then public clinics, also scored higher compared with district hospitals. Specialists had higher quality scores compared with primary care physicians. By contrast, neither the country of practice nor the patient load predicted differences in the quality of care (see Table 3). Finally, we ran the same model with physician income and found that it did not predict quality ($P < 0.05$, not shown).

Next, we modelled these same physician characteristics with monthly income as the dependent variable and found that quality did not predict income. However, younger physicians (<35 years) earned less per month than their older counterparts. Tertiary care physicians earned more than their colleagues in other facilities. The number of patients a physician followed, in contrast to the previous model on quality, showed that those who followed fewer than 1000 patients had lower incomes (Table 4).

When we compared skills across domains by level of care, we found that physicians in tertiary care hospitals typically took better histories, did more complete exams, made the diagnosis more accurately and prescribed the correct treatment. Except for testing and diagnosis, where there were relatively few scoring criteria, these differences were all statistically significant. As before, there was wide variation in the performance means by level of care (Figure 4 and Table 5).

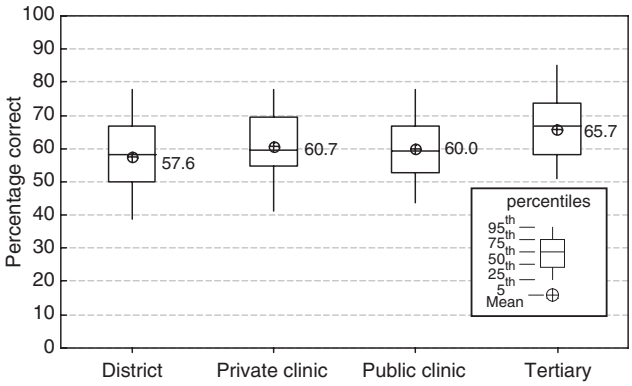

Figure 3 Variation in overall score by facility type

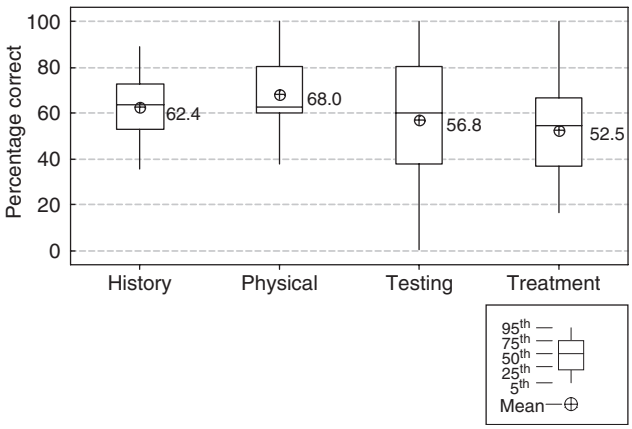

Figure 4 Variation by domain
Note: Diagnosis is omitted from this analysis because by construct there were only 2–3 items available for determining a distribution.

Discussion

Our large cross-sectional study of the quality of care in five countries shows that quality of care for diarrhoea, prenatal care and tuberculosis is low in all countries. There was only limited difference in average quality between countries. The variation in the quality of care within countries, however, was quite large, with vignette scores ranging from 30% to 93%. This variation persisted across facility type regardless of the clinical condition measured. Moreover, country identity did not predict differences in the quality of care in our model. Certain physician characteristics, by contrast, predicted better quality: younger physicians, women, specialists and non-district hospitals were found to provide better quality care.

An analysis by domain showed the history and physical domains were generally better performed and exhibited less variation than the testing and treatment domains. Overall, the highest standard of care by domain was in tertiary care hospitals, where doctors were more likely to perform the history taking and physical examination correctly and provide higher-quality treatment.

After controlling for country-level variables, it is interesting that income was not a proxy for quality. We controlled for the possibility that higher income might be due to higher volume but this was not the case. If, as in other service markets, patients are willing to pay more for higher quality, there was no evidence in our study that greater reimbursement was related to higher quality of services. We did find that doctors working in public clinics and non-specialists were reimbursed less. While older doctors did not provide as high a quality of care as their younger counterparts, they earned greater incomes. Gender did not predict income at the $P < 0.05$ significance level, although our data clearly suggest that female physicians are paid less than their male counterparts.

Our finding of extraordinary within-country variation in quality has two important implications. First, some physicians in developing countries perform exceptionally well, indicating that insufficient resources are not the sole or even an important predictor of physician capability. Secondly, poor quality can be addressed by directing remediation towards the poor performers. Strategies targeting poor performers would improve markedly the average quality provided for a given population. While some of these strategies can be directed by the type of facility or even the age of the physician, the overwhelming implication of this study is that quality of care must be measured before physicians can be targeted.

This study has several limitations. First, the sample frame was not intended to be representative. Instead, we began with the most well-known tertiary care hospital and then looked at referral hospitals and nearby facilities. However, by designing a sample frame that likely increases sample uniformity, we were likely conservative and underestimated the true variability if such a nationally representative sample could

Table 2 Physician characteristics by country

	Country					
Characteristics of 300 physicians	China	El Salvador	India	Mexico	Philippines	**Total number**
Level of care in sample						
Tertiary	24	25	24	24	26	123
District	8	24	9	24	8	73
Public clinic	24	14	9	8	8	63
Private clinic	8	9	8	8	8	41
Total	64	72	50	64	50	300
Age group						
Under 35	32	38	24	15	21	130
35–44	15	18	13	24	19	89
45–54	13	13	12	18	7	63
Over 55	4	3	1	7	3	18
Total	64	72	50	64	50	300
Gender						
Male	15	38	30	41	19	143
Female	49	34	20	22	31	155
Total	64	72	50	64	50	300
Specialty						
Primary care	16	25	26	23	28	118
Internal medical	17	18	7	17	7	66
Ob/gyn	23	15	9	14	8	69
Paediatrics	8	14	8	10	7	47
Total	64	72	50	64	50	300
Percentage of physicians under 35						
Tertiary care facilities	67	48	71	17	35	47
District care facilities	25	83	56	25	63	52
Public clinic care facilities	58	29	0	13	13	32
Private clinic care facilities	0	22	25	63	75	37
Total percentage of physicians under 35	50	53	48	25	42	44
Monthly income (US dollars in 2003)						
Mean	261	155	535	1380	858	
Standard deviation	134	146	221	772	1296	
Mean of top 25%	437	336	828	2477	2218	
Years practicing medicine						
Mean	15	9	13	15	12	
Standard deviation	11.1	7.5	8.8	10.9	9.5	
Years in the same location						
Mean	8	7	7	11	8	
Standard deviation	6.5	7.1	7.5	8.2	6.6	
Percentage of physicians following:						
0–500 listed patients	78	44	44	11	56	
501–1000 listed patients	8	14	10	5	14	
1001–1500 listed patients	11	6	2	5	4	
1501–2500 listed patients	0	8	8	5	4	
Over 2500 listed patients	0	24	24	31	22	
Not answered	3	4	12	44	0	
Total number of physicians = 300						

Table 3 Model of quality by country and other characteristics[a]

Independent variables	Coefficient	Standard error	*P*-value
Country variables			
China	−0.05	0.17	0.77
El Salvador	−0.15	0.16	0.35
India	0.15	0.16	0.35
Mexico	0.13	0.18	0.47
Philippines	(dropped)		
Age and gender			
Age <35	0.44	0.21	0.04
Age 35–44	0.26	0.22	0.24
Age 45–54	0.20	0.22	0.36
Age ≥55	(dropped)		
Gender (female)	0.27	0.1	0.01
Level of care			
Tertiary	0.53	0.14	0.00
Public clinic	0.29	0.15	0.05
Private clinic	0.36	0.16	0.02
District	(dropped)		
Specialty			
Internal medical	0.33	0.16	0.03
Ob/gyn	0.36	0.16	0.03
Paediatrics	0.36	0.18	0.05
Primary care physician	(dropped)		
Variables of listed patients			
0 to 500 listed patients	−0.08	0.18	0.65
501 to 1000 listed patients	−0.27	0.22	0.24
1001 to 1500 listed patients	0.06	0.26	0.81
1501 to 2500 listed patients	0.24	0.27	0.37
Over 2500 listed patients	0.08	0.18	0.67
Not answered	(dropped)		
Constant	−0.84	0.31	0.01
Number of observations	481		
sigma_u	0.54	0.06	0.00
sigma_e	0.74	0.04	0.00
rho	0.34	0.07	

[a] Dependent variable: quality of care (Z-value of overall quality score).

Note: These results are from a random effects model estimating the associations between the independent variables of country, age, gender, level of care, specialty, number of patients and quality of care score. For each variable listed above, the lowest performing group (for example, doctor's age over 55 in the age category) was dropped in the regression model. The exception was for the country variable, where the Philippines, as the median-performing group, was dropped in our analysis. Using this regression model, we see that age less than 35, female gender, level of care above district hospital, and specialty were positively associated with better quality of care.

Table 4 Predictors of income[a]

Independent variables	Coefficient	Standard error	*P*-value
Overall Z-score value	0.07	0.04	0.09
Country variables			
China	−0.75	0.14	0.00
El Salvador	−1.51	0.13	0.00
India	−0.16	0.13	0.24
Philippines	0.36	0.15	0.02
Mexico	(dropped)		
Age			
Age <35	−0.55	0.17	0.00
Age 35–44	−0.20	0.17	0.26
Age 45–54	0.05	0.17	0.76
Age ≥55	(dropped)		
Gender (woman)	−0.15	0.08	0.08
Level of care			
Tertiary	0.60	0.12	0.00
District	0.39	0.12	0.00
Private clinic	0.35	0.14	0.01
Public clinic	(dropped)		
Specialty			
Internal medical	−0.32	0.12	0.01
Paediatrics	−0.45	0.13	0.00
Primary care physician	−0.37	0.12	0.00
Ob/gyn	(dropped)		
Variables of listed patients			
0 to 500 listed patients	−0.33	0.15	0.03
501 to 1000 listed patients	−0.35	0.18	0.05
1001 to 1500 listed patients	−0.06	0.21	0.77
1501 to 2500 listed patients	−0.03	0.21	0.90
Over 2500 listed patients	−0.10	0.15	0.49
Not answered	(dropped)		
Constant	6.82	0.25	0.00
Number of physicians	300[b]		
F statistic	27.83		
R-square	0.71		

[a] Dependent variable: income reported (log 2003 US dollars).

[b] 55 of 300 physicians did not provide income information.

Note: The model estimates the associations between the independent variables of country, age, gender, level of care, specialty, number of patients and income reported. For each variable listed above, the highest earning group was dropped in the categories of age, gender and specialty; the lowest earning group was dropped in the category of level of care. Using this regression model, we see that younger age, female gender, public clinic and smaller patient loads were negatively correlated with income.

have been collected. Thus, the true variation of quality between countries is unknown and our results report less variation than if a national sample had been possible. Second, although we did not see important differences among countries, our relatively small sample size of countries would benefit from being increased in future studies (Peabody *et al*. 2006b). Third, while 82% (245/300) of participants reported their income, reporting may be biased by cultural or other economic factors. Fourth, we evaluated the quality of care for only three clinical conditions. Finally, this study is confined to physicians and

Table 5 Mean and standard deviation of domain score by level of care (shown as percentages)

	History		Physical Exam		Testing		Diagnosis		Treatment	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Tertiary	**66%**	16%	**75%**	19%	56%	33%	**81%**	24%	**59%**	23%
District	58%	16%	63%	21%	**61%**	32%	77%	27%	49%	23%
Public clinic	63%	14%	66%	22%	55%	31%	75%	27%	51%	20%
Private clinic	63%	14%	68%	21%	55%	35%	77%	25%	52%	22%
Facility type with highest quality in each domain										
Highest score	**Tertiary**		**Tertiary**		**District**		**Tertiary**		**Tertiary**	
***P*-value**	<0.01		<0.01		0.13		0.11		<0.01	

does not evaluate other practitioners who may follow guidelines more consistently (Paredes *et al*. 1996; Rowe *et al*. 2003). Characterizing the performance of nurses and medical assistants would be another valuable contribution to the field of health policy.

Major strengths of this study include: the use of vignettes, which provide a validated, case-mix-adjusted method for measuring the quality of care cross-nationally (Peabody *et al*. 2004); the study's large sample size and the concurrent collection of the data; its focus on the process of care, which is the proximal determinant of health care outcome; and its use of evidence-based criteria that were developed according to international standards of clinical practice.

Despite their widespread use to measure the quality of clinical practice, there remains some debate about whether vignettes only measure knowledge and not actual behaviour. Certainly, there is a well-documented gap between knowledge and practice (Igun 1994; Goel *et al*. 1996; Ofori-Adjei and Arhinful 1996; Paredes *et al*. 1996; Rowe *et al*. 2003). Influences or factors such as reimbursement opportunities, social expectations and lack of oversight also affect clinical practice in developing countries. However, in our validation studies we found that vignettes do capture actual clinical practice (Peabody *et al*. 2000; Peabody *et al*. 2004). The validation studies compared open-ended vignettes with the medical record and standardized patients. In this setting, vignettes accurately reflected actual clinical practice. We attribute the ability of our vignettes to measure clinical practice to their unique open-ended format that requires the provider to take the history, perform the physical examination and order tests while building a base of knowledge, which is very similar to an encounter with an actual patient. We have also found that vignettes capture variation when given simultaneously in a developed and a developing country (Peabody *et al*. 2004). However, it is possible that different factors, such as the generation of income from prescriptions for unnecessary medications or working with less clinical oversight, would result in higher quality scores on vignettes than would be seen in actual practice. If this were the case, our estimation of quality would be underestimated and scores would be lower. Because the validation studies have only been done in a developed setting, it would be ideal if future studies further validated the ability of vignettes to measure actual practice in both public and private settings in developing countries.

Conclusions

Implicit in health care planning is that the delivery of high quality medical care will improve the health of the populations. However, this belief has been extraordinarily difficult to measure. One of the most important technological limitations has been measuring the actual quality of care received by a population (Peabody *et al*. 1999). This paper elaborates on the use of vignettes as a way to make cross-national comparisons of quality among different population of patients simultaneously. These comparisons did not show that quality was substantially different among countries. However, the enormous within-country variation in quality seems to be a global problem that has been documented in developed and developing countries (Walker *et al*. 1988; Beracochea *et al*. 1995; Peabody *et al*. 1998; Nolan *et al*. 2001; Weinberg *et al*. 2001; Dumont *et al*. 2002). If health policy makers intend to improve the health status of the general population, the quality of care must be measured. Only then can effective measures be taken to target providers in need of improvement.

Acknowledgements

This paper is the product of an international collaboration. We are indebted to the following individuals for their contributions. We would like to thank Lily Alisse for helping to plan the study design, collecting data in China and the Philippines, and for scoring the vignette responses; Dan Bertenthal for conducting the preliminary analysis and error checking, and to Lin Zhang for completing the analysis. In China, we would like to thank Dr Jane Yunjing Ren, Dr Zhang Kong Lai; in El Salvador, Dr Jorge Alberto Munoz, Dr Jaime Avila; in India, Dr Naveet Wig, Dr Narendra Arora, Dr Neerja Bhatla, Dr Rakesh Lodha, Yogesh Singh, Alok Kumar Gupta, Chander Mohan Mittal and Saurabh Jha; in Mexico, Dr Carlos Carillo, Dr Stefano Bertozzi, Dr Ofelia Poblano Verastegui and Juan Pablo Gutierrez; in the Philippines, Dr Jesus Sarol, Dr Mario Taguiwalo, Dr Mike Alfiler and Oliver Asis. These individuals helped us to navigate the respective health systems of each country, obtained the ethical clearance and collected the data. We would also like to thank Sarah Weston and Cindy Muzio for their assistance preparing the manuscript.

Note to readers

Readers interested in conducting their own evaluations of clinical care quality are invited to contact the corresponding author for copies of vignettes and scoring sheets used in this study.

References

Beracochea E, Dickson R, Freeman P, Thomason J. 1995. Case management quality assessment in rural areas of Papua New Guinea. *Tropical Doctor* **25**: 69–74.

Donabedian A. 1980. *The definition of quality and approaches to its assessment*. Ann Arbor, MI: Health Administration Press.

Dumont A, De Bernis L, Bouillin D *et al.* 2002. Maternal morbidity and qualification of health-care workers: comparison between two different populations in Senegal. *Journal de gynécologie, obstétrique et biologie de la reproduction* **31**: 70–9.

Glassman PA, Luck J, O'Gara EM, Peabody JW. 2000. Using standardized patients to measure quality: evidence from the literature and a prospective study. *The Joint Commission Journal on Quality Improvement* **26**: 644–53.

Goel PK, Ross-Degnan D, McLaughlin TJ, Soumeraj SB. 1996. Influence of location and staff knowledge on quality of retail pharmacy prescribing for childhood diarrhea in Kenya. *International Journal for Quality in Health Care* **8**: 519–26.

Igun UA. 1994. The knowledge-practice gap: an empirical example from prescription for diarrhoea in Nigeria. *Journal of Diarrhoeal Disease Research* **12**: 65–9.

Labelle J, Swaine BR. 2002. Reliability associated with the abstraction of data from medical records for inclusion in an information system for persons with a traumatic brain injury. *Brain Injury* **16**: 713–27.

Loevinsohn BP, Guerrero ET, Gregorio SP. 1995. Improving primary health care through systematic supervision: a controlled field trial. *Health Policy and Planning* **10**: 144–53.

McGlynn EA, Asch SM, Adams J *et al.* 2003. The quality of health care delivered to adults in the United States. *New England Journal of Medicine* **348**: 2635–45.

Nolan T, Angos P, Cunha AJ *et al.* 2001. Quality of hospital care for seriously ill children in less-developed countries. *The Lancet* **357**: 106–10.

Ofori-Adjei D, Arhinful DK. 1996. Effect of training on the clinical management of malaria by medical assistants in Ghana. *Social Science and Medicine* **42**: 1169–76.

Paredes P, Pena M, Flores-Guerra E, Diaz J, Trostle J. 1996. Factors influencing physicians prescribing behavior in the treatment of childhood diarrhea: knowledge may not be the clue. *Social Science and Medicine* **42**: 1141–53.

Peabody JW. 1995. *Will measuring the quality of antenatal care tell us how it works?* RGSD-112. Santa Monica, CA: RAND.

Peabody JW, Gertler PJ, Leibowitz A. 1998. The policy implications of better structure and process on birth outcomes in Jamaica. *Health Policy* **43**: 1–13.

Peabody JW, Rahman MO, Gertler PJ *et al.* (eds). 1999. *Policy and health: implications for development in Asia*. Cambridge: Cambridge University Press.

Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. 2000. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *Journal of the American Medical Association* **283**: 1715–22.

Peabody JW, Luck J, Glassman P *et al.* 2004. Measuring the quality of physician practice by using clinical vignettes: a prospective validation study. *Annals of Internal Medicine* **141**: 771–80.

Peabody JW, Tozija F, Munoz JA, Nordyke RJ, Luck J. 2004. Using vignettes to compare the quality of clinical care variation in economically divergent countries. *Health Services Research* **39**: 1951–70.

Peabody JW, Nordyke RJ, Luck J *et al.* 2006a. Quality of care and its impact on population health: a cross-sectional study from Macedonia. *Social Science and Medicine* **62**: 2216–24.

Peabody JW, Taguiwalo MM, Robalino DA, Frenk J. 2006b. Improving the quality of care in developing countries. In: Jamison DT, Breman JG, Measham AR *et al.* (eds). *Disease control priorities in developing countries* Chapter 70. Oxford: Oxford University Press.

Rowe AK, Lama M, Onikpo F, Demin MS. 2003. Risk and protective factors for two types of error in the treatment of children with fever at outpatient health facilities in Benin. *International Journal of Epidemiology* **32**: 296–303.

Walker GJ, Ashley DE, Hayes RJ. 1988. The quality of care is related to death rates: hospital inpatient management of infants with acute gastroenteritis in Jamaica. *American Journal of Public Health* **78**: 149–52.

Weinberg M, Fuentes JM, Ruiz AI *et al.* 2001. Reducing infections among women undergoing cesarean section in Colombia by means of continuous quality improvement methods. *Archives of Internal Medicine* **161**: 2357–65.

Wennberg JE, Fisher ES, Skinner JS. 2002. Geography and the debate over Medicare reform. *Health Affairs (Millwood)* Suppl Web Exclusives: W96-114.

World Bank. 2004. *The World Development Report 2004: Making services work for poor people*. New York: Oxford University Press.

Zadnik K, Mannis MJ, Kim HS, Miller M, Marquez M. 1998. Interclinician agreement on clinical data abstracted from patients' medical charts. *Optometry and Vision Science* **75**: 813–6.