Using Vignettes to Compare the Quality of Clinical Care Variation in Economically Divergent Countries

John W. Peabody, Fimka Tozija, Jorge A. Muñoz, Robert J. Nordyke, and Jeff Luck

Objective. To determine whether clinical vignettes can measure variations in the quality of clinical care in two economically divergent countries.

Data Source/Study Setting. Primary data collected between February 1997 and February 1998 at two Veterans Affairs facilities in the United States and four government-run outpatient facilities in Macedonia.

Study Design. Randomly selected, eligible Macedonian and U.S. physicians (>97 percent participation rate) completed vignettes for four common outpatient conditions. Responses were judged against a master list of explicit quality criteria and scored as percent correct.

Data Collection/ Extraction. An ANOVA model and two-tailed t-tests were used to compare overall scores by case, study site, and country.

Principal Findings. The mean score for U.S. physicians was 67 percent (+/− 11 percent) compared to 48 percent (+/− 11 percent) for Macedonian physicians. The quality of clinical practice, which emphasizes basic skills, varied greatly in both sites, but more so in Macedonia. However, the top Macedonian physicians in all sites approached or—in one case—exceeded the median score in the U.S. sites.

Conclusions. Vignettes are a useful method for making cross-national comparisons of the quality of care provided in very different settings. The vignette measurements revealed that some physicians in Macedonia performed at a standard comparable to that of their counterparts in the United States, despite the disparity of the two health systems. We infer that in poorer countries, policy that promotes improvements in the quality of clinical practice—not just structural inputs—could lead to rapid improvements in health.

Key Words. Benchmarking/methods, quality indicators, health care, quality of health care, cross-national comparative study, clinical vignettes

The variation in health status between countries is attributed to such commonly cited factors as national income, education of girls, and even political governance (World Bank 1993). Although these factors are helpful markers of population health, there is growing interest in the performance of national health systems both as a way to explain variation and as a means to improve

health (Roemer 1991; World Health Organization 2000). The shift in thinking is driven by clear evidence showing cross-national differences in health services exist despite very similar levels of socioeconomic attainment and medical technology (World Bank 1993; Schieber 1997; The Technological Change in Health Care [Tech] Research Network 2001).

This new evidence suggests that differences may be based on the *process of care* (defined as what a physician or others do when seeing a patient). The quality of *clinical practice*, which comprises a major element of the process of care, is of particular interest as a policy variable. It is sensitive to changes made in the present, rather than over a long period of time as is the case with socioeconomic and system-level factors (Schieber 1997).

Immediate improvements in the overall process of care, given the same level of inputs (such as staffing and equipment), appear to result both in rapid improvements in health outcomes and lower costs (Donabedian 1980; Beracochea et al. 1995; Haddad, Fournier, Machouf, and Yatara 1998; Jamison and Sandbu 2001). Indeed, a tenet of health sector governance is that a policy that produces better or more efficient process of care will produce better health status in the population (Musgrove 1996; Peabody et al. 1999). Missing from this debate, however, are specific and reliable measures of the quality of clinical practice and direct comparisons between physicians in different countries (Walker 1983; Haddad, Fournier, and Potvin 1998; Saidel et al. 1998; Jamison and Sandbu 2001).

We and others recognized that any study directly comparing the quality of clinical practice in different settings must overcome several methodological and conceptual impediments (Liu et al. 1992). First, how can measurements take into account variations in case mix among the underlying patient populations in different countries? If case mix is not accounted for, clinical severity, comorbidities, and core sociodemographic factors as well as the

---

Address correspondence to John W. Peabody, M.D., Ph.D., Institute for Global Health, University of California, San Francisco, 74 New Montgomery, Suite 508, San Francisco, CA 94105. Dr. Peabody is also with the Department of Epidemiology and Biostatistics, University of California, San Francisco, the San Francisco Veterans Affairs Medical Center, the Department of Health Services, University of California, Los Angeles School of Public Health, and RAND. F. Tozija, M.D., Ph.D., is with the Cathedra of Social Medicine Medical Faculty, University of Cyril and Methodius, Skopje, Macedonia. J. A. Muñoz, Ph.D., is with Escuela Superior de Economía y Negocios, San Salvador, El Salvador. R. J. Nordyke, Ph.D., is with Zynx and Department of Health Services, University of California, Los Angeles School of Public Health. J. Luck, M.B.A., Ph.D., is with the Department of Health Services, University of California, Los Angeles School of Public Health and Veterans Affairs Greater Los Angeles Health Care System.

utilization of care and the promptness of access to care will cofound any comparison of clinical practice. Second, and even more problematic, how should clinical practice in different countries, particularly developing or low-income countries, be measured? Historically, quality measurement has relied on medical record reviews, which are subject to biases that limit their ability to fully reflect actual practice (Luck et al. 2000). The problems associated with using medical records are compounded by differing record keeping practices, which vary not only from place to place within the same country but from country to country (Walker 1983; Fowles et al. 1995; Katz et al. 1996; Bogardus et al. 2001). Third, even though it is recognized that the quality of clinical practice is particularly critical in settings that lack resources, the emphasis in the developing world has been on improving structural elements of health care, such as staffing, insurance, medications, supplies, equipment, and infrastructure to expand coverage (Walker 1983; Forsberg, Barros, and Victora 1992; Reerink and Sauerborn 1996; Peabody, Gertler, and Leibowitz 1998). As a result, most comparisons of quality between countries or regions focus on comparing structural elements (Forsberg, Barros, and Victora 1992). It would be far better to directly compare clinical practice rather than structural elements, since—in developed countries, where it has been studied—better clinical practice alone has led to better outcomes (Donabedian 1980; Jans, Schellevis, and LeCoq 2001).

We hypothesized that we could measure the variations in the quality of clinical practice in economically divergent countries by using a valid and reliable method we have developed called *clinical vignettes.* Vignettes are useful for comparisons because they overcome the three problems discussed above: case-mix adjustment, disparate medical record keeping, and emphasis on inputs or structural elements of care (Dresselhaus et al. 2000). In results published elsewhere, we have reported that clinical vignettes accurately measure *actual* clinical practice for a variety of clinical conditions (Peabody et al. 2000). In two prospective validation studies, vignettes captured differences in the quality of clinical practice between sites and health care systems when compared to a gold standard measurement (discussed further in the Methods section) (Dresselhaus et al. 2000; Peabody 2001). Related validation studies showed that the construct validity of vignettes exceeds that of quality measurements that rely on clinical records (Dresselhaus, Luck, and Peabody 2002; Luck and Peabody 2002).

The purpose of this study is to determine if vignettes are a useful method for making explicit cross-national comparisons of the quality of clinical practice, even in economically divergent countries, where heretofore, a

measurement tool has been lacking. We directly compared the quality variation of outpatient clinical practice from one area of the United States to the Republic of Macedonia, a middle-income country. Macedonia is an ideal setting for this study because of its long tradition of clinical care, the availability of all diagnostic and therapeutic interventions needed to test the conditions we studied, and a deeply held belief by providers of the value of giving high-quality care to the population.

METHODS

Site and Participants

This was a prospectively designed study conducted at two sites in the United States and four sites in Macedonia between February 1997 and February 1998. The U.S. sites were outpatient clinics that form part of the government-run Veterans Affairs health care system; they are located in the western part of the country. The Macedonia sites were also outpatient facilities of the government-run health care system and were located in the south and central areas.

   The Macedonian Health System. Like most countries in Easter Europe, primary health care in Macedonia is delivered through a system of large health centers and smaller clinics. Health centers, located in larger towns, provide primary and secondary outpatient care, as well as limited inpatient care in some areas. Clinics are typically smaller and can be urban or rural. Clinics also include health stations (small, urban community primary care facilities) that are typically administered by the local health center. There are also a small number of private primary care clinics. Macedonians are free to pursue care at either centers or clinics.

   Health Status in Macedonia. Life expectancy in 1997 was 70.4 years for males and 74.9 for females, which is comparable to many middle- and higher-income countries (European Observatory on Health Care Systems 2000). Also, like middle- to high-income countries, the leading causes of death are cardiovascular disease and cancer (European Observatory on Health Care Systems 2000). Additional indicators show that Macedonia has fully transitioned to a middle-income country health profile. For example, in 1998, 97 percent of all births were attended by a health professional and the under-12-month immunization rate for measles was 96 percent (World Health Organization 1999). In addition, there were 20.4 physicians per 10,000 population in Macedonia, comparable to the 21.3 per 10,000 in the United States in 1995 (European Observatory on Health Care Systems 2000; National Center for

Health Statistics 2003). Thus, while the health systems in the United States and Macedonia differ substantially, primary care physicians in these two countries face many of the same case-mix issues: management of chronic, lifestyle-based conditions as well as common infectious diseases and obstetric concerns.

For the U.S. sites, all practicing primary care physicians including attendings and residents (but not interns) were eligible to participate. Ninety-eight of 101 eligible physicians (97 percent) agreed to be in the study and 40 (20 per site) were then randomly selected to complete vignettes. From Macedonia, all primary care physicians working in four administrative regions were identified. As in the United States, eligibility was based on having an active primary care practice and voluntarily consenting to be in the study. Three hundred seventeen out of 319 physicians (99 percent) agreed to be in the study. Of those, 200 physicians were randomly selected and completed vignettes identical to the ones given U.S. physicians. No physician participating in the study from either country had seen or completed these vignettes prior to the study.

Quality Measurement

In previous publications, we have described how vignettes are developed (Glassman et al. 2000). Vignettes present physicians with a written scenario involving a fictitious patient and ask how they would respond. They are given 12–20 minutes to complete the vignette or "see the patient."

The vignettes are organized into five sections, or domains, which, when completed in chronological order, recreate the normal sequence of events in an actual patient visit: taking the patient's history, performing the physical examination, ordering radiological or laboratory tests, making a diagnosis, and administering a treatment plan. Physicians proceed from one section to the next by reading the information presented in the vignette and indicating—in an open-ended format—what actions they would take. Physicians are asked to be specific and they are given a range of the number of explicit responses within each domain. For example, the vignette might ask, "What are the 7–10 most important elements of the physical examination that you would like to do on this patient?" After providers give their responses (in this case the elements of the exam), they are given the answers. Once they are given answers they are not allowed to go back and revise previous responses. This gives the vignettes a question-and-answer format that closely resembles an actual patient visit.

The vignettes used in this study simulated the following common clinical conditions: (1) coronary artery disease; (2) low back pain; (3) chronic

obstructive pulmonary disease; and (4) diabetes mellitus. These conditions were chosen for three reasons. First, previous vignette validation studies, which tested the accuracy of vignettes against standardized patients (the gold standard), employed these same four conditions. Therefore, we were assured that—for these clinical cases—vignettes would accurately capture clinical practice. Second, the four conditions have a high prevalence in both countries (and worldwide). We used common presentations to minimize cultural bias and all cases were cases that were typically found in a primary care setting. Third, these cases emphasized taking a history and doing an appropriate physical examination in a primary care setting rather than sophisticated technology or highly specialized care. And fourth, the cases used only diagnostic strategies plus affordable, effective treatments that were available in both nations.

Each participant completed vignettes for four to eight cases. In both countries, the four conditions were divided into a simple and a slightly more complex case. The complex cases were distinguished by having one of two common comorbidities—hypertension or hypercholesterolemia. The cases were administered in a random order in one or two separate sittings depending on physician availability. To avoid a learning effect, no single sitting involved both the simple and the slightly more complex cases. To give the reader a sense of the conditions, Box 1 provides a brief summary of the simple and complex cases for coronary artery disease (CAD).

Prior to administration, the vignettes were extensively piloted in the two countries. Piloting revealed that physicians in both settings were familiar not only with being evaluated by means of vignettes but also with the detailed level of responses required. For example, for the patient with coronary artery disease, participants understood that it was not sufficient to report cardiac evaluation under physical examination; they needed to auscultate for a gallop rhythm or murmurs and measure the jugulovenous pressure. More than 60 physicians (roughly 30 per country) participated in the pilot testing and focus groups. To avoid contamination, the preliminary evaluations were done in locations removed from the study sites.

Before the vignettes were administered in Macedonia, they were translated and back translated into Macedonian by different pairs of bilingual physicians to ensure accuracy. Prior to scoring, the responses were translated by the same four bilingual physicians. Ten percent of the response translations were randomly retranslated to ensure accuracy and consistency. A single team consisting of one physician and two trained nurse abstractors completed the task of scoring to eliminate interrater variation between sites.

Box 1: Coronary Artery Disease Cases

CASE 1 (Simple)	CASE 2 (Complex)
A 65 year-old man, a new patient, comes to the clinic for follow-up of a heart attack that he had 3 months ago. In the history and physical the doctor needs to ascertain that the patient has been free of pain and has no difficulty performing routine activities since the heart attack, that he is overweight and continues to smoke, but has normal blood pressure. After the doctor records what he intends to do, this information is revealed to the doctor and then the doctor is asked what lab tests he or she would order (an EKG, cholesterol test), what the diagnosis is (uncomplicated heart attack or myocardial infarction) and how he or she would proceed with treatment.	A 62 year-old new patient presents with roughly the same story—a recent heart attack with similar risk factors—but in this history and physical, the doctor needs to learn that the patient has difficulty with routine exercises and easily becomes short of breath since he has run out of his medication. On examination the patient is found to have slight swelling of the ankles and slightly elevated blood pressure. When this information is revealed to the doctor, he or she is expected to order the same tests as in the first case plus a blood chemistry test and a chest X-ray. The EKG confirms that the patient has had a heart attack in the past.
The key element of case one is that the doctor recognizes that the heart attack is recent, associated with reversible risk factors, and the patient needs to be on aspirin and a beta-blocker. These last two interventions are affordable and have been demonstrated to prevent early death in population studies.	The key element in case two is that this is a heart attack complicated by mild heart failure. The doctor needs to evaluate for potential risk factors (again) and the patient needs to be placed on aspirin, a diuretic to remove the excess fluid associated with the heat failure, and be placed on a second drug (typically an ACE inhibitor). Like the first case the scientific evidence shows that this treatment prolongs survival in population of patients with heart failure.

Scoring Criteria

We conceptualized high-quality clinical practice as the *comprehensive* provision of services for a given clinical case that leads to better outcomes for individuals and populations. We determined what a physician would have to do during a patient visit to treat a clinical case in a manner consistent with standard practice recommendations. This involved describing a comprehensive set of actions that need to be undertaken by the physician. Scoring, therefore, did not rely on single-point measures such as determining if an antibiotic was prescribed or if the patient was screened in the history for a comorbidity. Instead, we used comprehensive measures that captured whether the physician: (1) determined the entire relevant history, (2) performed the relevant physical exam items, (3) ordered the necessary laboratory or imaging tests, (4) made the correct diagnosis including etiology, and (5) prescribed a complete treatment (management) plan.

We identified candidate criteria for each of the five domains of the vignette, first from the evidence-based literature on clinical care that lead to better outcomes and, second, from expert panels. The evidence-based criteria for each of the eight cases were initially identified from international clinical guidelines. In both countries, we then submitted all candidate criteria to local expert panels of academic or community physicians including both generalists and expert specialists in the four conditions. Based on their recommendations and group consensus, we finalized a master criteria list that was comparable across countries. (See Table 1 for an example of the criteria list for the coronary artery disease case.)

Abstractors (scorers), who were masked to physician identity, reviewed each vignette answer sheet and indicated on a scoring form those criteria the physician had successfully completed. A physician's score, expressed as a percentage correct, was calculated as the number of correctly completed criteria divided by the total number of criteria for that case. For further subanalyses, scores were calculated in a similar fashion for each of the five

Table 1:   Scoring Criteria for CAD Case 2

Domain	Criteria
History	• Date of myocardial infarction (MI)
	• Recent treatment/procedures
	• Angina and other symptoms
	• Selected risk factors/comorbidities
	• Prevention
	• Drug treatment
	• Risk factors
	• Symptoms of congestive heart failure (CHF)
Physical Exam	• Cardiac auscultation
	• Lung auscultation
	• Evaluate for peripheral vascular disease
Test Ordering	• Electrolytes, blood urea, and/or creatinine
	• Cholesterol
	• EKG
	• Echocardiograph (if available)
	• Exercise treadmill testing
Diagnosis	• Large anterior MI
	• Evidence of CHF
Treatment	• ACE inhibitor
	• Aspirin
	• Diuretic
	• Prevention-counseling
	• Follow-up visit

domains of the encounter (history taking, physical examination, test ordering, diagnosis, and treatment).

Analyses

The statistical analysis compared scores between countries—overall and disaggregated by disease, case, and domain of the encounter—as well as among sites. The statistical significance of the differences in scores between countries overall and for each of the four diseases, and the difference among sites, was evaluated by using ANOVA models that included factors for disease, country, study site, and physician. The disease and country variables were crossed, study site was nested within country, and physician was nested within site; the interaction between disease and country was not significant. The significance of differences in scores between countries for each of the eight cases and each of the five domains of the encounter were evaluated using a two-tailed t-test. Because of the very large differences in mean scores between countries, other comparisons were made on the basis of percentiles. Specifically, we determined the number of Macedonian physicians who scored above the 50th percentile of U.S. physicians, and subsequently the number who scored above the 25th percentile.

Earlier Studies Validating Vignettes as a Measure of Actual Clinical Practice

This study's open-ended vignettes had previously been validated against actual clinical practice (Peabody et al. 2000; Peabody 2001). In those studies, standardized patients (SPs)—actors rigorously trained to present into clinics as actual patients—served as the gold standard measurement of actual practice. The SPs were introduced unannounced into a doctor's outpatient practice (detection rate 3 percent in the first study) (Glassman et al. 2000). After an appointment with a physician, SPs recorded on a checklist the items performed by the physician. The accuracy of the SP checklists was also validated against audio recordings produced by concealed pocket pen recorders planted on SPs during a visit (Luck and Peabody 2002).

To do the validation calculations, the SP checklists, medical records from the SP visits, and corresponding vignettes completed by the same physicians were scored and compared using identical criteria. In an ANOVA model, the vignettes consistently produced scores closer to the gold standard of SPs than did the charts ($p < .05$) (Peabody et al. 2000). This finding was robust across sites, case, complexity, and level of training ($p < .05$). This

showed conclusively that vignettes accurately reflect what physicians actually do in the privacy of their own offices when seeing a patient.

RESULTS

The mean score for all vignette cases in the United States was 67 percent $(+/-11$ percent) compared with 48 percent $(+/-11$ percent) in Macedonia (see Figure 1). These differences persisted across the eight individual cases, each site within the country, and by case complexity (see Figure 1 and Table 2). The greatest absolute divergences in scores were for simple and complex low back pain (24 percent and 25 percent), the simple chronic obstructive pulmonary disease (COPD) case (22 percent), and the simple coronary artery disease case (21 percent).

Analysis of the variation amongst the *highest-scoring* U.S. and Macedonian physicians showed that there was overlap between the two countries. We compared the median U.S. score (67 percent) and the 25th U.S. percentile score (60 percent) to the percentage of Macedonian physicians that matched these U.S. performance standards. Overall, 3.5 percent of Macedonian

Figure 1:   Average Vignette Scores by Disease
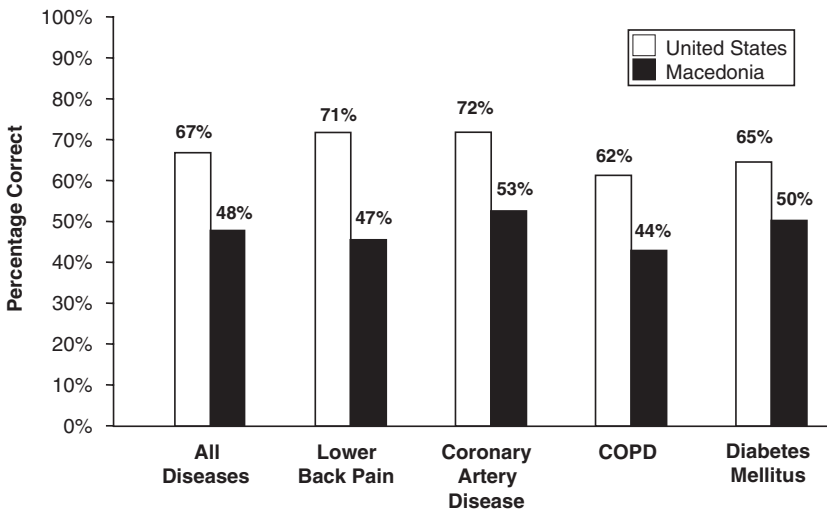
p < .0017 for all diseases combined
p < .0009 for Low Back Pain
p < .0025 for Coronary Artery Disease
p < .0063 for COPD (Chronic Obstructive Pulmonary Disease)
p < .0082 for Diabetes Mellitus

Table 2:   Comparison of Vignette Scores by Case between Countries

Case	United States (%)	Macedonia (%)	Difference (%)
LBP* 2	71	46	25
LBP 1	71	47	24
COPD* 1	66	44	22
CAD* 1	73	53	21
DM* 1	69	51	17
CAD 2	70	53	17
COPD 2	57	43	14
DM 2	62	50	12

*LBP = low back pain; COPD = chronic obstructive pulmonary disease; CAD = coronary artery disease; DM = diabetes mellitus; 1 = simple case; and 2 = complex case.
$p < .0001$ for all cases.

physicians matched the median U.S. score and 14.7 percent matched or exceeded the 25th percentile of U.S. physicians.

The variation between clinical skill sets was greater in Macedonia than in the United States For example, by domain, U.S. physicians obtained their highest average score for physical examination skills (79 percent) and lowest for treatment (53 percent), for a variation range of 26 percent. Meanwhile, Macedonia physicians scored highest on history taking (61 percent) and lowest on treatment (27 percent), for a variation range of 34 percent (See Table 3).

When we looked at the within-site variation, we observed a wide range in performance in both countries. Figure 2 plots the interquartile range of scores (25th to 75th percentile) as a box and the 5th to the 95th percentile as lines. In addition to the broad range of performance within a specific site (shown in Figure 2), it is apparent that the highest-scoring Macedonian physicians (the top 5 percent) from the best-performing Macedonian site (labeled

Table 3:   Comparison of U.S. and Macedonian Physicians' Vignette Scores Broken Down by Domains of Clinical Care

	United States (%)	Macedonia (%)	Difference (%)
History	74	61	13
Physical exam	79	45	34
Diagnosis	59	39	20
Testing	66	55	11
Treatment	53	27	26
All Domains	67	48	19

$p < .0001$ for all domains and all domains combined.

Figure 2:   Range of Vignette Scores by Site
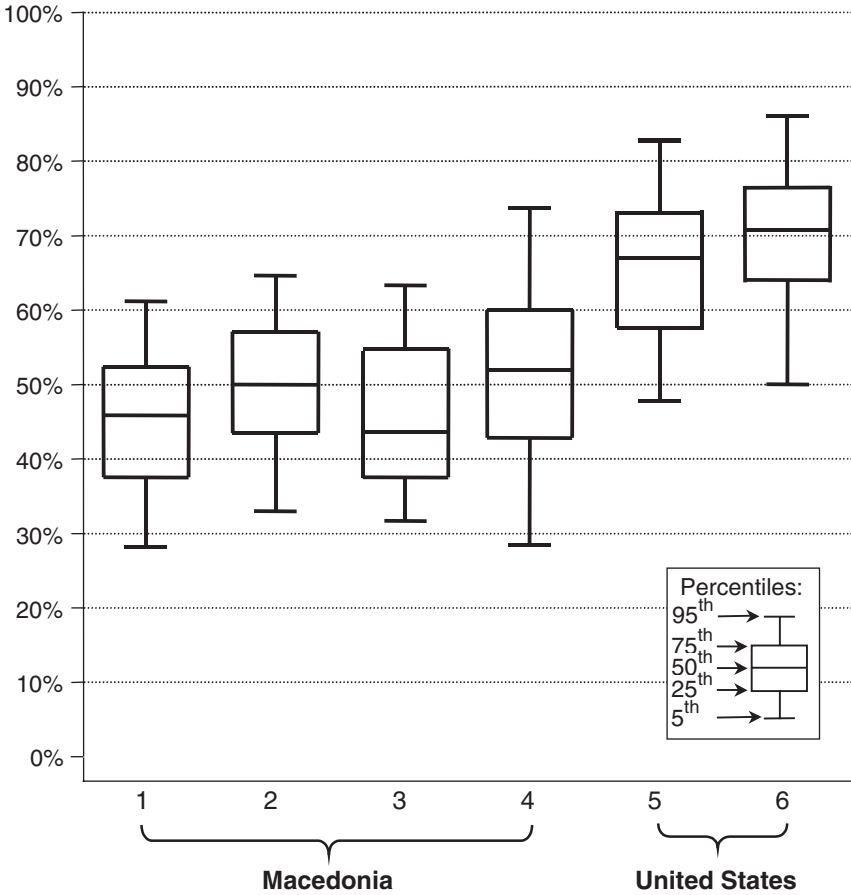
$p < .0001$ for difference between sites.

no. 4) exceeded the top quartile of the highest-scoring U.S. physicians at one U.S. site (labeled no. 5). In addition, the scores of the top 5 percent of Macedonian physicians in all Macedonian sites approached or—in one case—exceeded the median score of both U.S. sites.

DISCUSSION

Direct cross-national comparisons of the quality of clinical practice have been hampered by the limited availability of a suitable measurement method

(Walker 1983; Haddad, Fournier, and Potvin 1998; Saidel et al. 1998; Jamison and Sandbu 2001). In this prospective study, we demonstrated that clinical vignettes, previously validated against standardized patients (SPs), can be used to directly compare the quality of clinical practice in two economically divergent countries.

We found that scores measuring the quality of clinical practice for four common outpatient conditions were significantly different among randomly selected physicians in nonrandomly selected areas of the United States and Macedonia. These differences persisted across eight different cases and the five domains of clinical care such as history taking and diagnosis.

The most striking finding, however, was that the variation in the quality of clinical practice, as measured by the vignettes, was very large in both countries although more so in Macedonia than in the United States. This was particularly striking across the different domains and among physicians. When we looked at the highest vignette scores at all Macedonian sites, 14.7 percent of doctors matched or exceeded the score representing the 25th percentile of all the U.S. doctors. At one Macedonian site, the score representing the top two to three doctors (5 percent) exceeded the score representing the top five (25 percent) doctors at one site in the United States.

Many would argue that this snapshot of quality variation does not take into account the system-level effects that exist in both countries. Clearly, physicians practice within complex health systems. The organizational, financial, and political effects of these systems can impact the overall level of quality in both positive and negative ways. However, broad assessments of the quality of care in health systems in the past have obscured the role of clinical practice, a critical determinant of overall quality. Moreover, many elements of clinical practice, such as physician knowledge and skills, are independent of system-level effects. By isolating physician practice patterns from these system-level effects, vignettes may be able to provide a more accurate and unbiased assessment of the quality of clinical practice across disparate health systems. Measurements of medication compliance by patients, for example, can be combined with vignette measurements of clinical treatment to obtain a more comprehensive picture of the process of prescribing behaviors.

The widespread interest in having a more detailed look at clinical practice is based on the expectation that interventions, which change clinical practice and are introduced at the system-level, will produce better clinical outcomes. Since the groundbreaking and controversial 2000 World Health Report, *Health Systems: Improving Performance*, we and others have been prospectively examining the provision of care for specific diseases and trying to

measure the range of clinical practice among and within divergent heath care systems (Peabody et al. 1994; Tunstall-Pedoe et al. 2000; World Health Organization 2000; Mcclellan and Kessler 2002). These newer studies are in contrast to many previous studies, that only measured practice implicitly (Rees et al. 1978; Malone 1980; Nolan et al. 2001; Technological Change in Health Care [Tech] Research Network 2001) or only compared quality in developing countries by examining structural measures (e.g., staffing, equipment and supplies, drug usage, and triage capabilities) (Peabody et al. 1994; Nouira et al. 1998; Peabody et al. 1998; Laing, Hogerzeil, and Ross-Degnan 2001; Nolan et al. 2001; Stenson et al. 2001).

Studies of the quality of clinical practice in developing countries in the past have also been hampered by often being observational (Amonoo-Lartson, Alpaugh-Ojermark, and Neumann 1985; World Health Organization 1990; Bryce et al. 1992; Gilson, Kitange, and Teuscher 1993; Beracochea et al. 1995; McClellan and Kessler 2002), retrospective (Walker, Ashley, and Hayes 1988), or descriptive (Madden et al. 1997) and they are most commonly limited to studies of perinatal care practice (Graham et al. 2000). Recently, to overcome measurement difficulties, other researchers have also begun using vignettes in prospective evaluation to measure quality of clinical care using a prospective, random sample of providers (World Health Organization 1990; Montagu 2002).

Like this study, the few existing reports that attempted to measure the quality of clinical practice also found that the (average) level of provider knowledge and skills were wanting. In one observational study in Papua New Guinea, for example, only 19–39 percent of patients had their history adequately taken (depending on the type of provider) (Beracochea et al. 1995). In another observational study done in Pakistan, only 56 percent of providers reached an acceptable minimal standard for diagnosis and only 35 percent met the acceptable standard for treatment (Thaver et al. 1998). A health facility survey administered in Bangladesh revealed that only 39 percent of doctors interviewed were able to select correct treatment for a child showing signs of dehydration (World Health Organization 1990). These studies, like ours, evaluated common clinical care for conditions for which affordable and effective treatments exist regardless of country. It is also interesting to note that, as we found here, the skills were the highest for history taking and physical examination but decreased in the areas of testing and diagnostic accuracy and reached a nadir with treatment.

Advances in evidence-based clinical practice, as well as the limited association between structural quality measures and health outcomes, highlight

the importance of improving what physicians do in clinical practice. We believe it is crucial to measure whether clinical practice for common conditions in developing countries meets international standards. Measurement must address standards of clinical practice that are linked to better outcomes, lead to performance improvement interventions that are feasible with local resources, and be able to measure changes in clinical practice over time. We believe that vignettes can fulfill all of these requirements.

The implication of our findings, if replicated in other studies, are important: We found that there is both large variation in the quality of clinical practice and that some physicians in a lower-income country do as well or better than their counterparts in a wealthier country. This supports the hypothesis that quality of clinical practice could be improved under existing economic circumstances. Improving the clinical practice of low-end performers would raise the average and lead to improved health outcomes at a lower cost and in a much shorter time than other typical health reform measures that invest in buildings, equipment, or other material goods.

This study also showed that even the simple things like history taking and the physical examination are done inadequately and, although it is more of a problem in Macedonia, it is a problem in both countries. Moreover, these problems were robust and found across conditions, domains, and sites. An often overlooked goal of public policy is to create conditions and incentives for all physicians to meet high standards (Institute of Medicine 2001). This contrasts with policies that invest in structural elements or—even more distally—rely on long-term economic growth, to improve population health. If policies and other interventions that target specific skills, such as history taking, were successfully introduced, this study demonstrates that some doctors operating even in settings where resources are severely constrained could still provide high-quality care. Thus, being able to measure clinical practice in divergent settings with a tool such as vignettes makes it possible to identify practice disparities and suggest interventions that could improve clinical practice.

This study has four main limitations. First, the samples were not nationally representative and may not reflect all of the geographic variations in care within a country. However, not only was it not the intent of this study to define the level of quality for two countries, the finding that the between-site variation is greatly exceeded by the within-site variation in both countries makes any national level comparison irrelevant. Second, validation of the vignettes, although rigorous, was done only in the United States. This limitation may be difficult to overcome because validation in a developing country would require training standardized patients and placing them

unannounced in the country's clinical care facilities, as we did in our original validation studies. Third, this study only looks at two sites or countries and the sample size in subanalyses of doctors was small. We also confined our study to primary care physicians and to common outpatient conditions. To correct this, more cross-national comparisons involving generalists and specialists are needed to see if the variability we found in this study is robust in other sites. Fourth, we did not measure the patients' health outcomes. Although our scoring criteria are largely evidence-based and known to lead to better health, we do not know if the differences in quality found here are linked to differences in health outcomes. One way to address this problem would be to make a cross-national comparison of vignettes, as we have done here, and then to simultaneously measure the health status of patients with the four common conditions. These limitations are combined with the strengths of the study, which include its prospective design, random sampling of doctors, validated and case-mix-adjusted measure of quality, and use of explicit criteria.

Direct cross-national comparisons of clinical practice provide insight into the quality performance of national health care systems. Previous research, although limited, supports the intuition that quality of clinical care is poor in many countries, and few would disagree that improving the quality of clinical care using existing resources is an international health priority. With direct comparisons using tools such as clinical vignettes, it is possible to identify sites and basic clinical skills that could be improved. We believe that research on the quality of clinical care and related interventions, guided by the growing body of knowledge that shows how quality can be improved using feedback, guidelines, management techniques (Loevinsohn, Guerrero, and Gregorio 1995; Institute of Medicine 2001), and financial and nonfinancial incentives (Kumaranayake et al. 2000), could help reduce the disparities in health status between countries.

REFERENCES

Amonoo-Lartson, A., M. Alpaugh-Ojermark, and A. Neumann. 1985. "An Approach to Evaluating the Quality of Primary Health Care in Rural Clinics in Ghana." *Journal of Tropical Pediatrics* 31 (5): 282–5.

Beracochea, E., R. Dickson, P. Freeman, and J. Thomason. 1995. "Case Management Quality Assessment in Rural Areas of Papua New Guinea." *Tropical Doctor* 25 (2): 69–74.

Bogardus, S. T., V. Towle, C. S. Williams, M. M. Desai, and S. K. Inouye. 2001. "What Does the Medical Record Reveal about Functional Status? A Comparison of

Medical Record and Interview Data." *Journal of General Internal Medicine* 16 (11): 728–36.

Bryce, J., M. J. Toole, R. J. Waldman, and A. Voigt. 1992. "Assessing the Quality of Facility-Based Child Survival Services." *Health Policy and Planning* 7 (2): 155–63.

Donabedian, A. 1980. *The Definition of Quality and Approaches to Its Assessment.* Ann Arbor, MI: Health Administration Press.

Dresselhaus, T. R., J. Luck, and J. W. Peabody. 2002. "The Ethical Problem of False Positives: A Comparison of Standardized Patients and the Medical Record." *Journal of Medical Ethics* 28: 291–4.

Dresselhaus, T. R., J. W. Peabody, M. Lee, M. M. Wang, and J. Luck. 2000. "Measuring Compliance with Preventive Care Guidelines: Standardized Patients, Clinical Vignettes, and the Medical Record." *Journal of General Internal Medicine* 15 (11): 782–8.

European Observatory on Health Care Systems. 2000. *Health Care Systems in Transition: The Former Yugoslav Republic of Macedonia.* Copenhagen: European Observatory on Health Care Systems.

Forsberg, B. C., F. C. Barros, and C. G. Victora. 1992. "Developing Countries Need More Quality Assurance: How Health Facility Surveys Can Contribute." *Health Policy and Planning* 7 (2): 193–6.

Fowles, J. B., A. G. Lawthers, J. P. Weiner, D. W. Garnick, D. S. Petrie, and R. H. Palmer. 1995. "Agreement between Physicians' Office Records and Medicare Part B Claims Data." *Health Care Financing Review* 16 (4): 189–99.

Gilson, L., H. Kitange, and T. Teuscher. 1993. "Assessment of Process Quality in Tanzanian Primary Care." *Health Policy* 26 (2): 119–39.

Glassman, P. A., J. Luck, E. M. O'Gara, and J. W. Peabody. 2000. "Using Standardized Patients to Measure Quality: Evidence from the Literature and a Prospective Study." *The Joint Commission Journal on Quality Improvement* 26 (11): 644–53.

Graham, W., P. Wagaarachchi, G. Penney, A. McCaw-Binns, K. Y. Antwi, and M. H. Hall. 2000. "Criteria for Clinical Audit of the Quality of Hospital-Based Obstetric Care in Developing Countries." *Bulletin of the World Health Organization* 78 (5): 14–20.

Haddad, S., P. Fournier, N. Machouf, and F. Yatara. 1998. "What Does Quality Mean to Lay People? Community Perceptions of Primary Health Care Services in Guinea." *Social Science Medicine* 47 (3): 381–94.

Haddad, S., P. Fournier, and L. Potvin. 1998. "Measuring Lay People's Perceptions of the Quality of Primary Health Care Services in Developing Countries. Validation of a 20-Item Scale." *International Journal of Quality in Health Care* 10 (2): 93–104.

Institute of Medicine. 2001. *Crossing the Quality Chasm.* Washington, DC: National Academy Press.

Jamison, D. T., and M. E. Sandbu. 2001. "Global Health: WHO Ranking of Health System Performance." *Science* 293 (5535): 1595–6.

Jans, M. P., F. G. Schellevis, and E. M. LeCoq. 2001. "Health Outcomes of Asthma and COPD Patients: The Evaluation of a Project to Implement Guidelines in General Practice." *International Journal for Quality in Health Care* 13 (1): 17–25.

Katz, J. N., L. C. Chang, O. Sangha, A. H. Fossel, and D. W. Bates. 1996. "Can Comorbidity Be Measured by Questionnaire Rather Than Medical Record Review?" *Medical Care* 34 (1): 73–84.

Kumaranayake, L., P. Mujinja, C. Hongoro, and R. Mpembeni. 2000. "How Do Countries Regulate the Health Sector? Evidence from Tanzania and Zimbabwe." *Health Policy and Planning* 15 (4): 357–67.

Laing, R., H. Hogerzeil, and D. Ross-Degnan. 2001. "Ten Recommendations to Improve Use of Medicines in Developing Countries." *Health Policy and Planning* 16 (1): 13–20.

Liu, K., M. Moon, M. Sulvetta, and J. Chawla. 1992. "International Infant Mortality Rankings: A Look Behind the Numbers." *Health Care Financing Review* 13 (4): 105–18.

Loevinsohn, B. P., E. T. Guerrero, and S. P. Gregorio. 1995. "Improving Primary Health Care through Systematic Supervision: A Controlled Field Trial." *Health Policy and Planning* 10 (2): 144–53.

Luck, J., J. Peabody, T. Dresselhaus, L. Martin, and P. Glassman. 2000. "How Well Does Chart Abstraction Measure Quality? A Prospective Comparison of Standardized Patients with the Medical Record?" *American Journal of Medicine* 108 (8): 642–9.

Luck, J., and J. W. Peabody. 2002. "Using Standardised Patients to Measure Physicians' Practice: Validation Study Using Audio Recordings." *British Medical Journal* 325 (7366): 679.

Madden, J. M., J. D. Quick, D. Ross-Degnan, and K. K. Kafle. 1997. "Undercover Careseekers: Simulated Clients in the Study of Health Provider Behavior in Developing Countries." *Social Science Medicine* 45 (10): 1465–82.

Malone, M. I. 1980. "The Performance of Enrolled Community Nurses in the Management of Child Morbidity at an Integrated Maternal Child Health Clinic." *East African Medical Journal* 57 (1): 12–23.

McClellan, M. B., and D. P. Kessler. 2002. *A Global Analysis of Technological Change in Health Care: Heart Attacks.* Ann Arbor, MI: University of Michigan Press.

Montagu, D. 2002. "Franchising of Health Services in Low-Income Countries." *Health Policy and Planning* 17 (2): 121–30.

Musgrove, P. 1996. *Public and Private Roles in Health: Theory and Financing Patterns.* Washington, DC: World Bank.

National Center for Health Statistics. 2003. *Health, United States, 2003: Chartbook on Trends in the Health of Americans.* Atlanta: Centers for Disease Control and Prevention.

Nolan, T., P. Angos, A. J. Cunha, L. Muhe, S. Qazi, E. A. Simoes, G. Tamburlini, M. Weber, and N. F. Pierce. 2001. "Quality of Hospital Care for Seriously Ill Children in Less-Developed Countries." *Lancet* 357 (9250): 106–10.

Nouira, S., E. Roupie, S. El Atrouss, I. Durand-Zaleski, C. Brun-Buisson, F. Lemaire, and F. Abroug. 1998. "Intensive Care Use in a Developing Country: A Comparison between a Tunisian and a French Unit." *Intensive Care Medicine* 24 (11): 1144–51.

Peabody, J., P. Gertler, and A. Leibowitz. 1998. "The Policy Implications of Better Structure and Process on Birth Outcomes in Jamaica." *Health Policy* 43 (1): 1–13.

Peabody, J., J. Luck, P. Glassman, T. Dresselhaus, and M. Lee. 2000. "Comparison of Vignettes, Standardized Patients, and Chart Abstraction: A Prospective Validation Study of Three Methods for Measuring Quality." *Journal of the American Medical Association* 283 (13): 1715–22.

Peabody, J., M. Rahman, P. Gertler, J. Mann, D. Farley, J. Luck, D. Robalino, and G. Carter. 1999. *Policy and Health: Implications for Development in Asia.* Cambridge, UK: Cambridge University Press.

Peabody, J. W. 2001. *Developing Comparative Data to Improve Care: Using Standardized Patients and Vignettes to Compare Quality among VA and Non-VA Facilities.* Academy of Health Sciences Research (AHSR) Annual Meeting, Atlanta, GA.

Peabody, J. W., O. Rahman, K. Fox, and P. Gertler. 1994. "Quality of Care in Public and Private Primary Health Care Facilities: Structural Comparisons in Jamaica." *Bulletin of the Pan American Health Organization* 28 (2): 122–41.

Reerink, I. H., and R. Sauerborn. 1996. "Quality of Primary Health Care in Developing Countries: Recent Experiences and Future Directions." *International Journal for Quality in Health Care* 8 (2): 131–9.

Rees, P. H., L. R. Bagg, D. P. Hansen, and J. J. Thuku. 1978. "Medical Care in a Tropical National Reference and Teaching Hospital: Outline Study of Cost-Effectiveness." *British Medical Journal* 2 (6130): 102–4.

Roemer, M. I. 1991. *National Heath Systems of the World.* New York: Oxford University Press.

Saidel, T. J., B. Vuylsteke, R. Steen, N. S. Niang, F. Behets, H. Khattabi, L. Manhart, A. Brathwaite, I. F. Hoffman, and G. Dallabetta. 1998. "Indicators and the Measurement of STD Case Management in Developing Countries." *AIDS* 12 (2, supplement): S57–65.

Schieber, G., and A. A. Maeda. 1997. *A Curmudegeon's Guide to Financing Health Care in Developing Countries.* Innovations in Health Care Financing: Proceedings of a World Bank Conference. Washington, DC: World Bank.

Stenson, B., L. Syhakhang, C. S. Lundborg, B. Eriksson, and G. Tomson. 2001. "Private Pharmacy Practice and Regulation. A Randomized Trial in Lao P.D.R." *International Journal of Technology Assessment in Health Care* 17 (4): 579–89.

Thaver, I. H., T. Harpham, B. McPake, and P. Garner. 1998. "Private Practitioners in the Slums of Karachi: What Quality of Care Do They Offer?" *Social Science Medicine* 46 (11): 1441–9.

Technological Change in Health Care (TECH) Research Network. 2001. "Technological Change around the World: Evidence from Heart Attack Care." *Health Affairs* 20 (3): 25–42.

Tunstall-Pedoe, H., D. Vanuzzo, M. Hobbs, M. Mahonen, Z. Cepaitis, K. Kuulasmaa, and U. Keil. 2000. "Estimation of Contribution of Changes in Coronary Care to Improving Survival, Event Rates, and Coronary Heart Disease Mortality across the WHO Monica Project Populations." *Lancet* 355 (9205): 688–700.

Walker, G. J., D. E. Ashley, and R. J. Hayes. 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 (2): 149–52.

Walker, G. J. A. 1983. "Medical Care in Developing Countries: Assessment and Assurance Quality." *Evaluations and the Health Professions* 6 (4): 439–52.

World Bank. 1993. *Investing in Health.* Washington, DC: World Bank.

World Health Organization. 1990. *Programme for Control of Diarrhoeal Diseases: Interim Programme Report 1990.* Geneva: World Health Organization.

———. 1999. *Emergency and Humanitarian Action: Baseline Statistics.* Geneva: World Health Organization.

———. 2000. *World Health Report 2000 Health Systems: Improving Performance.* Geneva: World Health Organization.