

Medical Decision Making

<http://mdm.sagepub.com>

Diagnostic Certainty as a Source of Medical Practice Variation in Coronary Heart Disease: Results from a Cross-National Experiment of Clinical Decision Making

Karen E. Lutfey, Carol L. Link, Lisa D. Marceau, Richard W. Grant, Ann Adams, Sara Arber, Johannes Siegrist, Markus Bönnte, Olaf von dem Knesebeck and John B. McKinlay
Med Decis Making 2009; 29; 606 originally published online May 21, 2009;
DOI: 10.1177/0272989X09331811

The online version of this article can be found at:
<http://mdm.sagepub.com/cgi/content/abstract/29/5/606>

Published by:



<http://www.sagepublications.com>

On behalf of:



[Society for Medical Decision Making](#)

Additional services and information for *Medical Decision Making* can be found at:

Email Alerts: <http://mdm.sagepub.com/cgi/alerts>

Subscriptions: <http://mdm.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations <http://mdm.sagepub.com/cgi/content/refs/29/5/606>

Diagnostic Certainty as a Source of Medical Practice Variation in Coronary Heart Disease: Results from a Cross-National Experiment of Clinical Decision Making

Karen E. Lutfey, PhD, Carol L. Link, PhD, Lisa D. Marceau, MPH,
Richard W. Grant, MD, MPH, Ann Adams, PhD, Sara Arber, PhD, Johannes Siegrist, PhD,
Markus Bönnte, MA, Olaf von dem Knesebeck, PhD, John B. McKinlay, PhD

The authors examined physician diagnostic certainty as one reason for cross-national medical practice variation. Data are from a factorial experiment conducted in the United States, the United Kingdom, and Germany, estimating 384 generalist physicians' diagnostic and treatment decisions for videotaped vignettes of actor patients depicting a presentation consistent with coronary heart disease (CHD). Despite identical vignette presentations, the authors observed significant differences across health care systems, with US physicians being the most certain and German

*physicians the least certain ($P < 0.0001$). Physicians were least certain of a CHD diagnoses when patients were younger and female ($P < 0.0086$), and there was additional variation by health care system (as represented by country) depending on patient age ($P < 0.0100$) and race ($P < 0.0021$). Certainty was positively correlated with several clinical actions, including test ordering, prescriptions, referrals to specialists, and time to follow-up. **Key words:** clinical decision making; medical practice variation; health disparities. (*Med Decis Making* 2009;29:606–618)*

Variations in the provision of health care and associated health disparities are a topic of long-standing concern for policy makers, clinical providers, and health researchers alike. Although differences in clinical decision making (CDM) have been observed for many conditions,¹ coronary heart disease has received particular attention and has been shown repeatedly to result in differential

diagnostic and treatment decisions by physicians.^{2,3} Differences have been observed in the use of coronary revascularization services,⁴ hospitalization for hypertension,⁵ history taking,⁶ and gender differences in attributions of cardiac-related symptoms.⁷

These variations in medical practice occur as a function of patient characteristics such as race,^{4,5} age,^{5,7,8} socioeconomic status,^{9–11} gender,^{12–14} and comorbidity status,¹⁵ as well as provider and system attributes such as gender,¹⁶ attitudes toward aging,¹⁷ perceptions of pressure from patients,¹⁸ and practice culture.¹⁹ Studies of the United States, including especially the RAND Health Services Utilization Study²⁰ and the Dartmouth Atlas of Health Care^{21,22} project, have consistently documented geographic variations in health care and have worked over the past 2 decades to understand why it occurs. Similar differences have been observed cross-nationally among the United States, France, and England²³; between the United States and Canada²⁴ and the United States and United Kingdom²⁵; among Eastern European countries²⁶; and in the SYMPHONY trial, a study of 37 countries.²⁷

Received 30 November 2007 from New England Research Institutes, Watertown, Massachusetts (KEL, CLL, LDM, JBM); Massachusetts General Hospital and Harvard Medical School, Boston (RWG); University of Warwick, Coventry, UK (AA); University of Surrey, Guildford, UK (SA); Dusseldorf University, Dusseldorf, Germany (JS); and University of Hamburg, Hamburg, Germany (MB, OK). Financial support for this study was provided entirely by a grant from National Institutes of Health, National Institute on Aging (grant #AG16747). The study sponsors had no involvement in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Address correspondence to Karen E. Lutfey, PhD, New England Research Institutes, 9 Galen Street, Watertown, MA 02472; e-mail: klutfey@neriscience.com.

DOI: 10.1177/0272989X09331811

How physicians process complex and varied sources of information has been a topic of interest for decades. Classic studies underscore the importance of threshold models for conceptualizing the triggers that prompt physicians to do further testing and, if warranted, provide treatment.^{28,29} More recently, related literature in social psychology and economics examines how physicians' cognitive processing—particularly prejudice, stereotyping, discrimination, and uncertainty—may bias their assessments of patients and decisions about their treatment.^{30–32} Low health literacy also contributes to difficulties in chronic illness management, and such challenges to doctor-patient communication may exacerbate physician uncertainty.^{33,34} This literature, which is quite focused on providing micro-level explanations for observed variation, is often not equipped to also provide information about cross-national differences.

We build on these 2 literatures by examining how physicians' diagnostic certainty functions in clinical decision making for coronary heart disease (CHD). Using cross-national experiment data, we are able to simultaneously 1) address the role of diagnostic certainty in observed variations and 2) estimate relative contributions of patient, provider, and health care system influences on clinical decision making.

From a Bayesian decision theory perspective, these variations in clinical decision making should be determined in part by the prevalence of the relevant condition in the larger population. In cross-national studies, for example, women have slightly higher prevalence of angina relative to men (despite higher mortality rates among men), a difference that should influence individual clinical decisions.^{35–37} Research on the socially constructed aspects of health statistics, however, suggests that biases in clinical decision making and medical treatment contribute independently to differences in some types of health statistics.³⁸ Additional work has shown women have poorer outcomes after acute myocardial infarction³⁹ and in cardiovascular and diabetes care,^{13,14} both after adjusting for covariates. From this perspective, differences in rates may not only reflect epidemiologic differences in underlying disease but also cumulative interactions between patients and physicians.

Using data from a video vignette experiment, we examine the magnitude and sources of variation in CDM for identical presentations of CHD in 3 different health care systems: 1) the free market medical system in the United States, 2) the government-based National Health Service of the United

Kingdom,⁴⁰ and 3) the nonprofit insurer-based system used in Germany.⁴¹ This work expands existing social psychological research on CDM to include cross-national comparisons, as well as providing specific evidence for how uncertainty operates in CDM. Building on earlier work showing physicians in different countries had comparable rates of correct diagnoses, we find that diagnostic certainty has independent and unique effects on clinical decision making.²⁵ We address the following research questions: 1) How certain were physicians of their diagnoses of CHD, and how did that vary by health care system? 2) Which types of physicians were the most certain? 3) Which types of patients elicited the highest certainty levels among physicians? 4) How did these patient and provider effects vary across 3 countries? 5) How was diagnostic certainty associated with subsequent patient management, such as information seeking, test ordering, prescribing, lifestyle recommendations, and referrals/follow-up?

DATA AND METHODS

We conducted a factorial experiment to simultaneously measure the effects of 1) patient attributes (age, gender, race, and socioeconomic status), 2) physician characteristics (gender and years of clinical experience), and 3) separate health care systems (the United States, United Kingdom, or Germany) on physician diagnostic certainty and subsequent medical decision making when providers are presented with identical signs and symptoms indicative of CHD. Experiments were conducted in the United States (Massachusetts), United Kingdom (the West Midlands, Southeast London, and Surrey), and Germany (Northern Rhine/Westfalia region).⁴² A full factorial of combinations of patient age (55 years v. 75 years), gender, race (white v. black in the United States, white v. Afro-Caribbean in the United Kingdom, and white only in Germany), and socioeconomic status (SES; lower v. higher social class, depicted by occupation as a cleaner/janitor v. a teacher) was used for the videotaped vignette scenarios ($2^4 = 16$ unique vignettes). The decision to omit the race factor from the German experiment was based on discussions with our German colleagues, who advised us that the physicians in our sampling area saw few black patients in their everyday practices. Audios for the vignettes were dubbed into German and backtranslated to ensure accuracy. One of the 16 combinations was shown to each physician.

CHD was selected for the vignettes because 1) it is among the most common and costly problems presented by older patients to primary care providers⁴³; 2) it represents a clinically well-defined medical condition; 3) it admits a range of diagnostic, therapeutic, and lifestyle actions; and 4) its reported prevalence differs among the United States, United Kingdom, and Germany. Scripts for the vignettes were developed from several tape-recorded role-playing sessions with experienced clinical advisers, and professional actors were trained (under experienced physician supervision) to realistically portray a patient presenting with the signs/symptoms of disease to a primary care provider. Patients in the vignette presented with signs and symptoms that were consistent with CHD, including chest pain worsening with exertion, pain in the back between the shoulder blades, stress, and elevated blood pressure. All patients included a nonverbal "Levine's fist," a well-known gesture indicative of cardiac pain. Because CHD is a spectrum condition, and live patients do not typically present as clear-cut textbook cases of specific conditions, the vignette also built in several red herring symptoms potentially indicative of a gastrointestinal diagnosis. To this end, the patient also complained of indigestion, feeling worse after a large or spicy meal, having pain similar to heartburn but unresponsive to antacids, and feeling full and "gassy." This was done not to specifically make the physicians' diagnostic task more difficult but to more accurately represent how actual patients present, based on advice from our clinical advisers. The vignette also incorporated references to the patient's mood, including the spouse's report that the patient has been difficult to be around and the patient's self-report of feeling irritated and having decreased energy (see the appendix for an illustrative excerpt from the vignette).

After viewing the videotaped vignette, physicians were asked, "What do you think is going on with this patient?" and for each possibility, they were asked for their level of certainty on a scale of 0 to 100 (where 0 = no certainty and 100 = complete certainty). Physicians were asked through an open-ended question from the interviewer to list their full set of differential diagnoses, and if CHD was present anywhere on that list, they were counted as having considered the condition, and the accompanying certainty level was included in the present analysis. Physicians were also asked how they would manage the patient in terms of asking for additional information, performing physical examinations, ordering tests, prescribing medications, giving lifestyle advice, and referring to other physicians.

On the basis of the theoretical approach outlined above and our concerns about the social construction of epidemiologic base rates, we do not appeal to population rates of disease to ascertain the "correctness" of a given diagnosis. Rather, we assume that although a physician's diagnostic priors (among many other sources of learning) may inform decision making, patient-specific clinical information should be used above and beyond those preexisting base rates. Therefore, the vignette for this study was purposely designed to present a set of signs and symptoms sufficient to trigger a CHD diagnosis, regardless of the epidemiologic prevalence in any subgroup of the population. In this sense, the vignette provides sufficient information to suspect the condition, regardless of the social characteristics of the patient.

The study used a probability sample of physicians selected from within each of 4 strata within each country. To be eligible for selection, physicians had to 1) be internists or family practitioners in the United States and in Germany or general practitioners in the United Kingdom (to most accurately capture the types of nonspecialist physicians most likely to treat undiagnosed cases of CHD in each country), 2) be trained at an accredited medical school in the country in which they practiced (no international medical graduates were included), and 3) be currently in clinical practice more than half-time. Within each country, physicians were stratified into 4 equal cells by gender and level of experience, with "less" experience defined as those with ≤ 12 years since graduation from medical school in the United States or United Kingdom (≤ 7 years since licensure in Germany) and those with "more" experience having ≥ 22 years since graduation from medical school in the United States or United Kingdom (≥ 17 years since licensure in Germany). These cutoff dates were chosen to act as a proxy for clinical experience (which has been shown to affect clinical decision making, as discussed above), to standardize as much as possible the amounts of clinical experience across countries, and to create a clear separation between the strata. Twelve strata of physician characteristics (gender, years of clinical experience [< 12 or > 22 years], and health care system [United States/United Kingdom/Germany]) were defined, with 32 physicians included in each stratum from each country. This configuration generated a total of 384 physicians required to complete the design and the experiment (16 vignettes \times 12 physician strata \times 2 replications = 384; Table 1).

Screening telephone calls were conducted to identify eligible participants, and an hour-long

Table 1 Physician Enrollment in Study

	United States (<i>n</i> = 128)		United Kingdom (<i>n</i> = 64 from The Midlands; <i>n</i> = 64 from Surrey/Southeast London)		Germany (<i>n</i> = 128)		Total
	Level of Experience						
	Less	More	Less	More	Less	More	
Male	32	32	32	32	32	32	192
Female	32	32	32	32	32	32	192
Total	64	64	64	64	64	64	384
Response rate, %	64.9		59.6		65.0		

"Less" experience is defined as those with ≤ 12 years since graduation from medical school in the United States or United Kingdom (≤ 7 years since licensure in Germany). "More" experience is defined as having ≥ 22 years since graduation from medical school in the United States or United Kingdom (≥ 17 years since licensure in Germany).

appointment was scheduled for the interviewer to come to the physician's place of work to show the vignette and administer related questions. The interviews were conducted in 2001–2002 (128 throughout Massachusetts, 64 from The Midlands, and 64 throughout Surrey and Southeast London, England) and in 2005–2006 in Germany; every effort was made to select only one physician from each practice. Each physician participant was provided a modest stipend to partially offset lost revenue (\$100 [United States], £50 [United Kingdom], €100 [Germany]) and to acknowledge his or her participation. Quality control interviews and site visits were conducted, and selected tape-recorded interviews were reviewed by supervisors on a regular basis.

Analysis of variance was used to test the main effects and 2-way interactions of the design variables (patient gender, race, age, SES, and physician gender and level of experience) on the diagnostic certainty (0–100, with 0 for *not at all certain* and 100 for *completely certain*). If the physician did not make a CHD diagnosis, his or her certainty for these diagnoses was set to 0. To determine the effect of certainty on clinical decision making, we used logistic regression for dichotomous variables (e.g., whether or not an electrocardiogram [EKG] was ordered), and analysis of covariance was used for continuous variables (e.g., number of days to next appointment). Each model included as explanatory variables the design variables, certainty, and the interaction of the design variables and certainty. Using backward elimination, nonsignificant effects (at the 0.05 level) other than certainty were removed from the model, leaving a parsimonious model. Because of the challenges of multiple testing, we emphasize consistency across results and focus on identifying general patterns of physician certainty and treatment decisions.

Furthermore, the results we observe at the $P < 0.01$ level are unlikely to change. To facilitate interpretation, we present actual P values, unadjusted for multiple testing. To further facilitate interpretation of results, we indicate in Tables 2 to 4 the number of expected and observed significant results.

RESULTS

1. How Certain Were Physicians and How Did That Vary by Practice Setting (Country)?

Across all 3 countries, the vast majority of physicians correctly considered CHD (74.2% in Germany, 88.3% in the United Kingdom, 95.3% in the United States, and 85.9% overall, $P < 0.0001$), yet there were also significant differences in how many physicians failed to consider CHD in each health care system, with a 21.1% difference in rates between Germany and the United States.

Physicians' certainty levels for CHD diagnosis varied from 0 to 100, with an average of 52.1 (see Figure 1). Again, there was significant variation across countries, with the US physicians having the highest average certainty (57.9), followed by the United Kingdom (46.4) and Germany (38.9; $P < 0.0001$; Table 2). Using Tukey's method of multiple comparisons, we found that the level of certainty in the United States was significantly higher than that in either the United Kingdom or Germany, whereas the certainty levels were statistically comparable in the United Kingdom and Germany.

2. Which Physicians Were the Most Certain?

We next examined whether physician characteristics (gender and years of experience) were associated

Table 2 Analysis of Variance Results for Certainty Regarding Coronary Heart Disease: Main Effects and Significant ($P < 0.05$) 2-Way Interactions

Experimental Factor			Mean Certainty	P Value
Patient gender				0.0058
	Male		51.8	
	Female		43.7	
Patient age, y				0.0003
	55		42.4	
	75		53.1	
Patient race (United Kingdom/United States only)				0.4024
	Black		50.7	
	White		53.6	
Patient socioeconomic status				0.1998
	Lower		45.9	
	Upper		49.6	
Physician gender				0.7221
	Male		48.3	
	Female		47.2	
Physician experience				0.1575
	Less		45.7	
	More		49.8	
Practice setting (country)				< 0.0001
	Germany		38.9	
	United Kingdom		46.4	
	United States		57.9	
Patient Gender × Patient Age				0.0086
	Male	55	50.3	
		75	53.3	
	Female	55	34.5	
		75	52.9	
Patient Race × Practice Setting (Country) (United Kingdom/United States only for P value)				0.0021
	Germany	White	38.9	
	United Kingdom	Black	50.3	
	United Kingdom	White	42.5	
	United States	Black	51.2	
	United States	White	64.7	
Patient Race × Level of Experience (United Kingdom/United States only)				0.0500
	Black	Less	44.5	
	Black	More	56.9	
	White	Less	54.1	
	White	More	53.0	
Patient Age × Practice Setting (Country)				0.0100
	55	Germany	27.6	
	55	United Kingdom	42.2	
	55	United States	57.3	
	75	Germany	50.3	
	75	United Kingdom	50.5	
	75	United States	58.5	
Expected significant				1.4
Observed significant				7

Table 3 Effect of Certainty Regarding Coronary Heart Disease (CHD) on Clinical Decision Making: Continuous Variables (Change per 10-Point Increase in Certainty)

Variable	Change per 10-Point Increase in Certainty	95% Confidence Interval	P
Number of questions	-0.04	-0.15, 0.07	0.4178
Number of examinations	0.06	-0.02, 0.14	0.1137
Number of tests for CHD			
Germany	0.02	-0.06, 0.10	0.5815
United Kingdom	0.56	0.48, 0.64	< 0.0001
United States	0.39	0.31, 0.47	< 0.0001
Time to next appointment (days)	-0.44	-0.68, -0.20	0.0003
Number of pieces of lifestyle advice	-0.02	-0.07, 0.03	0.5360
Expected significant			0.25
Observed significant			3

with certainty, independent of the health care system in which providers practiced (Table 2). Although there were no significant main effects for physician characteristics, we did observe an interaction between patient race and physician level of experience (Figure 2a). With white patients, physicians had comparable certainty levels (54.1 v. 53.0, less and more experienced, respectively). With black patients, however, more experienced physicians had increased certainty (56.9), whereas those with less experience were less certain (44.5).

3. With Which Patients Were Physicians Most Certain?

Independent of health care system differences, physician certainty varied significantly according to the gender of the patient, with physicians reporting higher average certainty levels with male v. female patients (51.8 v. 43.7, $P=0.0058$) and by age, with higher certainty for older patients (53.1 v. 42.4, $P=0.0003$). We also observed an interaction between patient age and gender, such that physicians were much less certain in making a CHD diagnosis for younger women (34.5) with otherwise identical symptom presentation ($P=0.0086$; Figure 2b). There were no main effects of patient race or SES on certainty (Table 2).

4. How Did Patient and Provider Effects Vary Across Countries?

In addition to the main effect differences between countries, we also observed variation between countries according to characteristics of the presenting patient (Table 2). First, we observed an interaction

between patient age and health care system, with US physicians having comparable diagnostic certainty for younger (55-year-old) and older (77-year-old) patients (57.3 v. 58.5), German physicians having the greatest difference in certainty for the 2 types of patients (27.6 v. 50.3), and UK physicians falling in between (42.2 v. 50.5; $P=0.0100$; Figure 2c). Second, there was an interaction between health care system and patient race. Although physicians in the United States and United Kingdom had similar certainty levels for black patients (51.2 v. 50.3), their certainty levels diverged when the patient was white, with US physicians having increased certainty and UK physicians being less certain (42.5 v. 64.7, $P=0.0015$; Figure 2d).

5. What Was the Effect of Certainty on Clinical Decision Making?

In turn, physicians' diagnostic certainty for CHD significantly influenced their subsequent diagnostic and therapeutic clinical actions (Tables 3 and 4). Logistic regression results showed that for each 10-point increase in certainty for a CHD diagnosis, physicians were less likely to ask questions about the patient's psychological state (odds ratio [OR] = 0.91, $P=0.0163$) or social environment (OR = 0.88, $P=0.0062$), but they were more likely to ask about other general information (OR = 1.10, $P=0.0232$; Table 4). We also observed an interaction of patient SES and certainty on the odds of asking questions about smoking. As the certainty of the CHD diagnosis increased, physicians were more likely to ask lower SES patients about smoking (OR = 1.14, $P=0.0065$), but certainty had no effect on the likelihood of physicians asking upper SES patients about smoking.

Table 4 Effect of Certainty Regarding CHD on Clinical Decision Making: Dichotomous Variables (Odds Ratio per 10-Point Increase in Certainty)

Variable	Odds Ratio per 10-Point Increase in Certainty	95% Confidence Interval	P
Information seeking			
Four or more questions	1.00	0.92, 1.08	0.9332
Questions about			
Pathology	0.96	0.89, 1.03	0.2704
Medical history	1.06	0.98, 1.14	0.1675
Pain	1.03	0.96, 1.11	0.4437
Smoking			
Lower SES patient	1.14	1.04, 1.25	0.0065
Upper SES patient	1.00	0.91, 1.09	0.9417
Alcohol	0.96	0.87, 1.05	0.3215
Psychological state	0.91	0.84, 0.98	0.0163
Social questions	0.88	0.81, 0.96	0.0062
General questions	1.10	1.01, 1.19	0.0232
Physical examination			
Complete physical	0.99	0.90, 1.09	0.8828
Test ordering			
Order tests for CHD			
Germany	1.06	0.97, 1.17	0.1714
United Kingdom	1.86	1.53, 2.24	< 0.0001
United States	1.91	1.56, 2.33	< 0.0001
Stress test	1.34	1.22, 1.47	< 0.0001
EKG			
Germany	1.46	1.28, 1.66	< 0.0001
United Kingdom	1.31	1.18, 1.45	< 0.0001
United States	1.95	1.61, 2.36	< 0.0001
Prescription writing			
CHD appropriate prescription	1.53	1.38, 1.70	< 0.0001
Referrals			
Referral to cardiologist	1.23	1.12, 1.36	< 0.0001
Referral to other medical professional	0.93	0.85, 1.00	0.0641
Advice giving			
Advice about:			
Diet	0.97	0.91, 1.04	0.4291
Smoking	1.00	0.92, 1.09	0.9896
Alcohol	0.94	0.86, 1.02	0.1446
Relaxation	0.91	0.81, 1.03	0.1434
Exercise	1.07	0.94, 1.22	0.3009
Weight	0.99	0.85, 1.17	0.9424
Expected significant			1.1
Observed significant			12

CHD, coronary heart disease; SES, socioeconomic status; EKG, electrocardiogram.

Certainty was also significant for test ordering behavior. For example, as certainty of a CHD diagnosis increased, physicians from the United States and United Kingdom were significantly more likely to order at least one diagnostic test for CHD ($P < 0.0001$), whereas this relationship did not hold for German physicians (Table 4). Furthermore, as

certainty increased, UK and US physicians ordered greater numbers of CHD-related tests ($P < 0.0001$), whereas certainty did not significantly affect the number of CHD-related tests that the German physicians ordered (Table 3). However, increased certainty was associated with increased likelihood of physicians from all 3 countries ordering stress tests

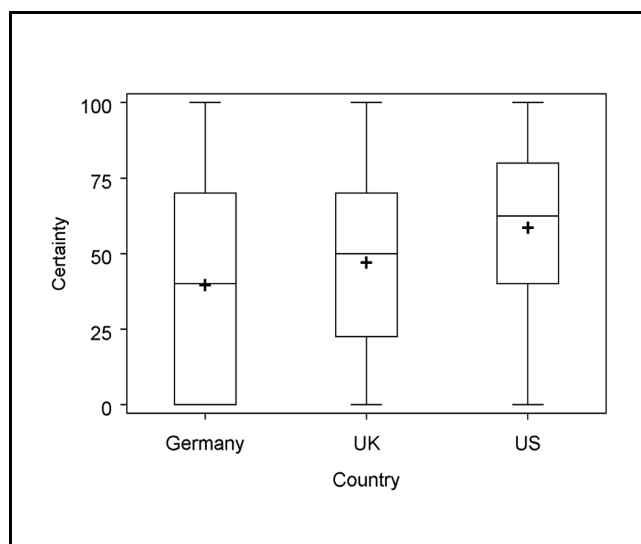


Figure 1 Physicians' certainty of coronary heart disease (CHD) diagnoses in 3 countries ($P < .0001$). Certainty ranges from 0 (complete uncertainty) to 100 (complete certainty). The box and whiskers plot can be interpreted as follows: the whiskers extend to the data minimum and maximum, the bottom of the box is at the 25th percentile, the top of the box is at the 75th percentile, the middle line is the median, and the + is at the mean.

and EKG tests ($P < 0.0001$), and this effect was the strongest in the United States (OR=1.95, $P < 0.0001$; Table 4).

Increased certainty was associated with higher odds of writing a CHD-appropriate prescription (beta-blockers, calcium channel blockers, aspirin, and short-acting nitrates; OR=1.53, $P < 0.0001$) in all 3 countries. Higher certainty of a CHD diagnosis also increased the odds of a physician referring the patient to a cardiologist (OR=1.23, $P < 0.0001$). Finally, increased certainty was significant for predicting how soon a physician would request to see the patient again ($P=0.0003$), with a repeat visit requested 0.44 days sooner per 10-point increase in certainty.

DISCUSSION

We observed significant differences between health care systems, with US physicians having the highest levels of certainty. As expected, physicians were least certain of their CHD diagnoses with younger female patients.^{12,44} In addition, there was racial variation depending on the physician's level of experience and variation by health care system

depending on patient age and race. Increased certainty was associated with differences in information seeking as well as increased test ordering, prescriptions, referrals to cardiologists, and shorter time to follow-up.

Previous studies of medical practice variation have largely focused on either system-level social and economic patterns in CDM or on physicians' individual-level cognitive processing as mechanisms that may generate such variation. The factorial design of our experiment allowed for unconfounded estimates of the simultaneous effects of patient characteristics, provider attributes, and health care systems (as represented by country) on physicians' certainty of CHD diagnoses and their subsequent clinical actions. With our analytic approach, we were able to capture physicians at all points on the certainty spectrum rather than excluding those who did not list CHD among their differential diagnosis selections. Although those who did not consider CHD among their differential diagnosis selections did not provide an explicit certainty value, we know their certainty concerning the presence of CHD was low because they were allowed to list a full set of diagnoses (we recognize that physicians may have high certainty that CHD is absent from the vignette, but this is beyond the scope of the present analysis). Therefore, if they were treated as missing, they would not be missing at random. By increasing the variability, our estimates are rendered more conservative because of a decrease in power.

The sample size of 384 allows us to detect a difference in certainty of 12 points with 97% power, assuming a standard deviation of 30 (which, assuming a unimodal beta distribution of 0–1, is the upper bound of a standard deviation). That is, a true 12-point difference in certainty between 2 groups will be detected 97% of the time at $\alpha = 0.05$. Because the experiment was replicated, a pure error term with 192 degrees of freedom was used to test all effects using analysis of variance, because of the omission of the race factor for the German experiment (explained above), only the US and UK data are considered for the effect of patient race on certainty. For all other analyses, all data were considered.

The main effects related to patient characteristics (decreased certainty with female and younger patients) partially corroborate Bayesian perspectives and studies of uncertainty and statistical discrimination,^{31,45} suggesting that when physicians are uncertain, they are likely to make diagnostic decisions that are consistent with existing epidemiologic base rates. In this case, prior assumptions

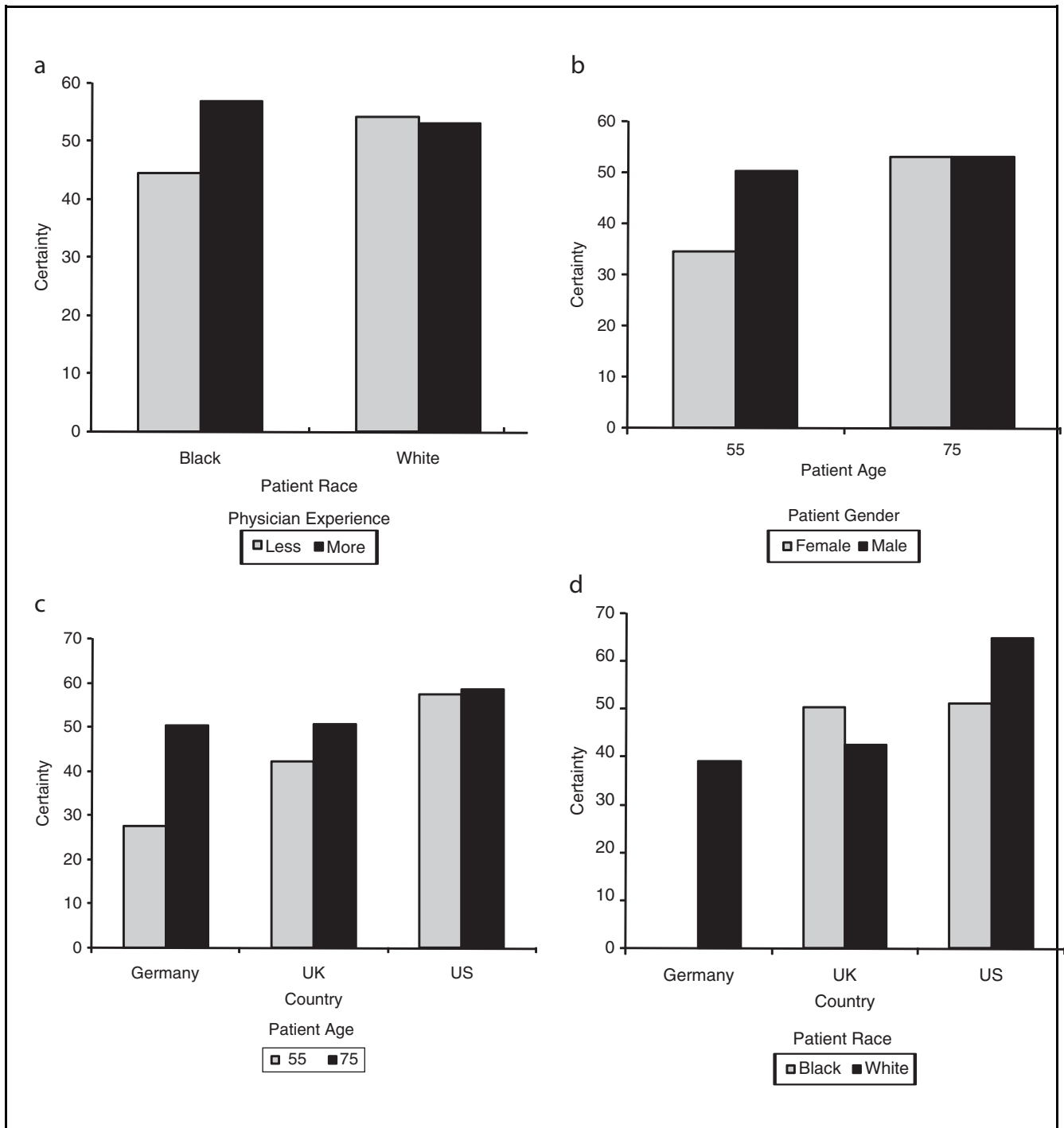


Figure 2 (a) Two-way interaction effect of patient race and physician level of experience on physicians' diagnostic certainty for coronary heart disease (CHD), 0 to 100 (United States and United Kingdom only; $P = 0.0500$). (b) Two-way interaction effect of patient age and patient gender on physicians' diagnostic certainty for CHD, 0 to 100 ($P = 0.0086$). (c) Two-way interaction effect of patient age and country on physicians' diagnostic certainty for CHD, 0 to 100 ($P = 0.0100$). (d) Two-way interaction effect of patient race and country on physicians' diagnostic certainty for CHD, 0 to 100 ($P = 0.0021$).

overwhelm the presenting patient-specific data, thereby contributing to the reification of some types of existing health statistics. Most important, these findings extend previous work by showing that certainty—not simply identifying a diagnosis but having diagnostic certainty about that condition—has an independent effect on clinical actions.⁴⁶ Therefore, these results suggest that having CHD on the differential diagnosis list is necessary but not sufficient for physicians to take appropriate therapeutic actions; this result is consistent with the notion that physicians need to pass certainty thresholds to order more tests or treat a patient.²⁸ By extension, improving disparities in CHD outcomes is not just a matter of physicians learning to consider CHD more appropriately in specific populations (e.g., women) but also to be able to do so with sufficient certainty to trigger appropriate clinical actions to improve morbidity and mortality outcomes.

However, these perspectives do not fully explain our results, and the persistence of some between-health care system differences implies that features of the broader sociological, cultural, and organizational environments are also relevant to decision making. Beyond diagnostic certainty, observed cross-national variation in CHD diagnosis and treatment may also be a function of differences in a series of influences that are beyond the scope of the present analysis or our study more generally. These include both patients' and physicians' cultural expectations for medical practice and treatment as well as biological variations in the prevalence of CHD and related conditions at a population level, such that physicians in different health care systems may be differentially equipped to identify CHD with certainty. For example, previous research has shown cross-cultural differences in the relationship between symptoms and underlying conditions.^{47,48} Other possible explanations include funding mechanisms, expectations for physicians to achieve diagnostic certainty in a brief period of time, modes of practice, and access to resources across the 3 countries. For example, economic reimbursement policies may translate to more pressure on US physicians to achieve a firm diagnosis and management plan during the initial patient consultation so they can be paid, whereas in the United Kingdom and Germany, physicians may tend to make these decisions over a series of consultations close together in time. Similarly, ready availability of technological equipment in Massachusetts in conjunction with a fear of lawsuits for missed diagnoses may lead to increased testing in the United States relative to

physicians in the United Kingdom or Germany, whereas increased rates of referral among UK and German physicians²⁵ may explain lower rates of testing and prescription treatment relative to the United States. Increased regulation in the United Kingdom with pay-for-performance may also contribute to the UK physicians' higher certainty relative to their German counterparts, where professional pressures and regulations are not yet as explicit as in the United States.⁴⁹ Most generally, physician learning is known to be related to the local context of practice, such that physicians may either self-select into environments with practice styles similar to their own or adapt to the local culture—either type of pattern could exist within the local contexts that were selected from each country for this study.

Every study represents a balance between internal and external validity. We recognize that our vignette-based approach has some limitations compared to studies of behavior in natural interaction—the most obvious being that physicians do not directly interact with the patient in the vignette. For present purposes, vignettes offer several key advantages over alternative methods: they 1) allow for the manipulation of several variables at once and the measurement of unconfounded effects, thereby “isolating physicians' decision making from other factors in the environment”⁵⁰; 2) provide a standardization of case mix; 3) allow vignette-based studies to collect a large amount of information simultaneously from a large number of subjects; 4) make efficient use of time; and 5) are cost-effective (e.g., standardized patients would have been prohibitively expensive in this context). In a direct comparison of vignettes, standardized patients, and chart abstraction, Peabody and colleagues⁵¹ validated the use of vignettes for studying quality of outpatient care, and studies comparing vignettes with standardized patients and other methods corroborate the result that vignettes are ecologically valid for studies of medical decision making.^{50,52,53}

We took 4 precautionary steps in an attempt to minimize possible threats to external validity and compensate for the artificial aspects of the experimental situation (i.e., that physicians may behave differently with a videotaped patient under experimental conditions compared with real patients in an everyday clinical setting). First, considerable effort was devoted to ensuring the clinical authenticity of the videotaped presentation. Expert clinical consultants were actively involved in all stages of the process, from early stages of role-playing and script

development to final stages of film shooting, where they oversaw vignette filming to determine face and content validity. Second, the doctors viewed the tapes in the context of their practice day (often during their lunch periods) to maximize the likelihood that they encountered real patients before and after they viewed the patient in the videotape and also so they were in a physical setting they associated with decision making. Third, the doctors were specifically instructed at the outset to view the patient as one of their own patients and to respond as they would typically respond in their own practice. In the United States, 90.6% of the doctors considered the vignettes to be very or reasonably typical, with 91.4% in the United Kingdom and 81.3% in Germany.

Our study has limitations that underscore the need for additional research. First, questions remain about how physicians cognitively process these cues from patients. Previous research in cognitive psychology has suggested that physicians often rely on pattern recognition as well as more analytic types of processing, such as Bayesian decision making, when evaluating patient cues.^{54,55} However, these questions are beyond the scope of the present study, and these data do not allow us to specify the exact cognitive and psychological processes physicians use when interpreting information from the vignettes. A similar study of US physicians (with the same CHD vignette) primed physicians to determine whether the underdiagnosis of CHD in some patient populations was due to physicians not considering that diagnosis or considering it and then eliminating it from their differential diagnosis.⁵⁶ Also beyond the scope of this study, but potentially related, is the question of differences in clinical decision-making practices between family practitioners and internists, despite both groups being likely to treat the type of patient depicted in the vignette.

Although the cross-national component of our study identified persistent differences between countries, there remains limited generalizability from each group of physicians to the entire population of physicians in their respective countries. In terms of statistical and clinical significance, our results are relatively modest and therefore limited in their ability to explain the wide range of cross-national variations that have been identified in existing literature. However, the possible explanations outlined above build on our current results and are promising avenues for future inquiry.

In summary, our findings underscore the role of uncertainty during the clinical decision-making

process in contributing to or amplifying CHD-related disparities. To the extent that inequalities are generated from within health care systems, researchers and policy makers should continue to develop interventions targeted at the level of the patient-physician encounter while also considering which broader, system-level factors influence individual physician behaviors. Because diagnostic certainty is so important for understanding subsequent clinical actions, our results also highlight the need for interventions to not only increase diagnostic accuracy but also increase certainty to lead to optimal therapeutic actions.

APPENDIX

Illustrative Excerpt from Vignette Script for CHD

Doctor: So, what brings you here today?

Patient: Actually, I've been having a fair amount of indigestion.

Doctor: Indigestion?

Patient: Yes.

Doctor: Can you describe it?

Patient: Yes, I just sort of get this feeling right here (*rubs chest/stomach area*) . . . it's not really a pain, actually. It's just really uncomfortable. It's usually after a big meal, if I've eaten too much, or if the food's particularly spicy, maybe. I sometimes feel kind of sick, too.

Doctor: Spicy foods, mostly, or does this happen after other foods too?

Patient: Actually, it's probably more common after a big meal, and especially if I have to rush to get somewhere afterwards. I get sort of queasy. It's really uncomfortable.

Doctor: How long has this been going on?

Patient: About 3 months.

Doctor: Would you say this is something new, or has this kind of thing happened before?

Patient: No, this is new. Now, sometimes I do get heartburn, but that usually goes away with antacids.

Doctor: How is this similar or different from the heartburn?

Patient: Well, the heartburn's right here too (*massaging the same spot*), but it burns. This other stuff is just, um, just a bad feeling. It's hard to describe.

Doctor: Tell me about the last time this happened. Where were you?

Patient: (*pauses, thinking*) The last time? We had just gone out for a meal and were walking to the car, quickly I remember, it was chilly. And I started to get the feeling.

Doctor: What did you do?

Patient: I stopped. And tried to get a deep breath in.

Doctor: Did that seem to help?

Patient: Yes, a bit. My [wife/husband] came over and asked if I was OK. I said, sure, I'm fine and we got into the car. After a few minutes of driving, the feeling began to go away.

REFERENCES

- Institute of Medicine. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Healthcare*. Washington, DC: National Academies Press; 2003.
- Ayanian JZ, Epstein AM. Differences in the use of procedures between women and men hospitalized for coronary heart disease. *N Engl J Med*. 1991;325:221–5.
- Schwartz LM, Fisher ES, Tosteson NA, et al. Treatment and health outcomes of women and men in a cohort with coronary artery disease. *Arch Intern Med*. 1997;157:1545–51.
- Popescu I, Vaughan-Sarrazin MS, Rosenthal GE. Differences in mortality and use of revascularization in black and white patients with acute MI admitted to hospitals with and without revascularization services. *JAMA*. 2007;297:2489–95.
- Holmes JS, Arispe IE, Moy E. Heart disease and prevention: race and age differences in heart disease prevention, treatment, and mortality. *Med Care*. 2005;43(suppl):I33–41.
- James TL, Feldman J, Mehta SD. Physician variability in history taking when evaluating patients presenting with chest pain in the emergency department. *Acad Emerg Med*. 2006;13:147–52.
- Martin R, Gordon EE, Lounsbury P. Gender disparities in the attribution of cardiac-related symptoms: contribution of common sense models of illness. *Health Psychol*. 1998;17:346–57.
- McKinlay JB, Potter DA, Feldman HA. Non-medical influences on medical decision-making. *Soc Sci Med*. 1996;42:769–76.
- Armstrong DL, Strogatz D, Wang R. United States coronary mortality trends and community services associated with occupational structure, among blacks and whites, 1984–1998. *Soc Sci Med*. 2004;58:2349–61.
- Fincher C, Williams JE, MacLean V, Allison JJ, Kiefe CI, Canto J. Racial disparities in coronary heart disease: a sociological view of the medical literature on physician bias. *Ethn Dis*. 2004;14:360–71.
- McKinlay JB, Link CL, Freund KM, Marceau LD, O'Donnell AB, Lutfey KE. Sources of variation in physician adherence with clinical guidelines: results from a factorial experiment. *J Gen Intern Med*. 2007;22:289–96.
- Arber S, McKinlay JB, Adams A, Marceau LD, Link CL, O'Donnell AB. Patient characteristics and inequalities in doctors' diagnostic and management strategies relating to CHD: a video-simulation experiment. *Soc Sci Med*. 2006;62:103–15.
- Bird CE, Fremont AM, Bierman AS, et al. Does quality of care for cardiovascular disease and diabetes differ by gender for enrollees in managed care plans? *Womens Health Issues*. 2007;17:131–8.
- Fremont AM, Correa-de-Araujo R, Hayes SN. Gender disparities in managed care: it's time for action. *Womens Health Issues*. 2007;17:116–9.
- Wexler DJ, Grant RW, Meigs JB, Nathan DM, Cagliero E. Sex disparities in treatment of cardiac risk factors in patients with type 2 diabetes. *Diabetes Care*. 2005;28:514–20.
- Britt H, Bhasale A, Miles DA, Meza A, Sayer GP, Angelis M. The sex of the general practitioner: a comparison of characteristics, patients, and medical conditions managed. *Med Care*. 1996;34:403–15.
- Collins E, Katona C, Orrell M. Management of depression in the elderly by general practitioners: II. Attitudes to ageing and factors affecting practice. *Fam Pract*. 1995;12:12–7.
- Armstrong D, Fry J, Armstrong P. Doctors' perceptions of pressure from patients for referral. *BMJ*. 1991;302:1186–8.
- Curoe A, Kralewski J, Kaissi A. Assessing the cultures of medical group practices. *J Am Board Fam Pract*. 2003;16:394–8.
- Brook RH, Park RE, Chassin MR, Koseoff J, Keesey J, Solomon D. Do Patient, Physician, and Hospital Characteristics Affect Appropriateness and Outcome of Selected Procedures? Santa Monica, CA: RAND; 1991.
- Goodman DC, Grumbach K. Does having more physicians lead to better health system performance? *JAMA*. 2008;299:335–7.
- Wennberg JE, Fisher ES, Skinner JS. Geography and the debate over Medicare reform. *Health Aff (Millwood)*. 2002;Suppl Web Exclusives:W96–114.
- Weisz D, Gusmano MK, Rodwin VG. Gender and the treatment of heart disease in older persons in the United States, France, and England: a comparative, population-based view of a clinical phenomenon. *Gen Med*. 2004;1:29–40.
- Pilote L, Saynina O, Lavoie F, McClellan M. Cardiac procedure use and outcomes in elderly patients with acute myocardial infarction in the United States and Quebec, Canada, 1988 to 1994. *Med Care*. 2003;41:813–22.
- McKinlay J, Link C, Arber S, et al. How do doctors in different countries manage the same patient? Results of a factorial experiment. *Health Serv Res*. 2006;41:2182–200. [Erratum in 41(6):2303.]
- Gurjeva OS, Bukhman G, Murphy S, Cannon CP. Treatment and outcomes of eastern Europeans with coronary syndromes in OPUS-TIMI 16. *Int J Cardiol*. 2005;100:101–7.
- Kramer JM, Newby LK, Chang WC, et al. International variation in the use of evidence-based medicines for acute coronary syndromes. *Eur Heart J*. 2003;24:2133–41.
- Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med*. 1980;302:1109–17.
- Pauker SG, Kassirer JP. Decision analysis. *N Engl J Med*. 1987;316:250–8.
- van Ryn M. Research on the provider contribution to race/ethnicity disparities in medical care. *Med Care*. 2002;40(suppl):I140–51.
- Balsa AI, McGuire TG, Meredith LS. Testing for statistical discrimination in health care. *Health Serv Res*. 2005;40:227–52.
- Lutfey KE, Ketcham JD. Patient and provider assessments of adherence and the sources of disparities: evidence from diabetes care. *Health Serv Res*. 2005;40(pt 1):1803–17.
- Villaire M, Mayer G. Chronic illness management and health literacy: an overview. *J Med Pract Manage*. 2007;23:177–81.
- Safeer RS, Keenan J. Health literacy: the gap between physicians and patients. *Am Fam Physician*. 2005;72:463–8.
- Bittner V. Angina pectoris: reversal of the gender gap. *Circulation*. 2008;117:1505–7.

36. Hemingway H, Langenberg C, Damant J, Frost C, Pyorala K, Barrett-Connor E. Prevalence of angina in women versus men: a systematic review and meta-analysis of international variations across 31 countries. *Circulation*. 2008;117:1526–36.
37. Shaw LJ, Lewis JF, Hlatky MA, et al. Women's ischemic syndrome evaluation: current status and future research directions: report of the National Heart, Lung and Blood Institute workshop: October 2-4, 2002: Section 5: gender-related risk factors for ischemic heart disease. *Circulation*. 2004;109:e56–8.
38. McKinlay JB. Some contributions from the social system to gender inequalities in heart disease. *J Health Soc Behav*. 1996;37:1–26.
39. Marrugat J, Gil M, Sala J. Sex differences in survival rates after acute myocardial infarction. *J Cardiovasc Risk*. 1999;6:89–97.
40. Light DW. Universal health care: lessons from the British experience. *Am J Public Health*. 2003;93:25–30.
41. Altenstetter C. Insights from health care in Germany. *Am J Public Health*. 2003;93:38–44.
42. Fisher RA. *Statistical Methods, Experimental Design and Scientific Inference*. New York: Oxford University Press; 1990.
43. Cohen JW, Krauss NA. Spending and service use among people with the fifteen most costly medical conditions. *Health Affairs*. 2003;22:129–38.
44. Arber S, McKinlay JB, Adams A, Marceau LD, Link CL, O'Donnell AB. Influence of patient characteristics on doctors' questioning and lifestyle advice for coronary heart disease: a UK/US video experiment. *Br J Gen Pract*. 2004;54:673–8.
45. Balsa AI, McGuire TG. Statistical discrimination in health care. *J Health Econ*. 2001;20:881–907.
46. Lutfey KE, Link CL, Grant RW, Marceau LD, McKinlay JB. Is certainty more important than diagnosis for understanding race and gender disparities? An experiment using coronary heart disease and depression case vignettes. *Health Policy*. In press.
47. Staniland JR, Clamp SE, de Dombal FT, et al. Presentation and diagnosis of patients with acute abdominal pain: comparisons between Leeds, U.K. and Akershus county, Norway. *Ann Chir Gynaecol*. 1980;69:245–50.
48. Wigton RS. Social judgement theory and medical judgement. *Thinking Reasoning*. 1996;2:175–90.
49. Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. Quality of primary care in England with the introduction of pay for performance. *N Engl J Med*. 2007;357:181–90.
50. Veloski J, Tai S, Evans AS, Nash DB. Clinical vignette-based surveys: a tool for assessing physician practice variation. *Am J Med Qual*. 2005;20:151–7.
51. Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA*. 2000;283:1715–22.
52. Dresselhaus TR, Peabody JW, Lee M, Wang MM, Luck J. Measuring compliance with preventive care guidelines: standardized patients, clinical vignettes, and the medical record. *J Gen Intern Med*. 2000;15:782–8.
53. Robra BP, Kania H, Kuss O, Schonfisch K, Swart E. [Determinants of hospital admission: investigation by case vignettes]. *Gesundheitswesen*. 2006;68:32–40.
54. Brooks LR, LeBlanc VR, Norman GR. On the difficulty of noticing obvious features in patient appearance. *Psychol Sci*. 2000;11:112–7.
55. Eva KW, Brooks LR. The underweighting of implicitly generated diagnoses. *Acad Med*. 2000;75(suppl):81–3.
56. Lutfey KE, Eva KW, Gerstenberger E, Link CL, McKinlay JB. The cognitive basis of diagnostic and treatment disparities in coronary heart disease: results of a factorial experiment. Unpublished manuscript, 2008.