

How Do Doctors in Different Countries Manage the Same Patient? Results of a Factorial Experiment

John McKinlay, Carol Link, Sara Arber, Lisa Marceau, Amy O'Donnell, and Ann Adams

Objective. To determine the relative contributions of: (1) patient attributes; (2) provider characteristics; and (3) health care systems to health care disparities in the management of coronary heart disease (CHD) and depression.

Data Sources/Study Setting. Primary experimental data were collected in 2001–2 from 256 randomly sampled primary care providers in the U.S. (Massachusetts) and the U.K. (Surrey, Southeast London, and the West Midlands).

Study Design. Two factorial experiments were conducted in which physicians were shown, in random order, two clinically authentic videotapes of “patients” presenting with symptoms strongly suggestive of CHD and depression. “Patient” characteristics (age, gender, race, and socioeconomic status [SES]) were systematically varied, permitting estimation of unconfounded main effects and the interaction of patient, provider, and system-level influences.

Data Collection/Data Extraction Methods. Analysis of variance was used to measure provider decision-making outcomes, including diagnosis, information seeking, test ordering, prescribing behavior, lifestyle recommendations, and referrals/follow-ups.

Principal Findings. There is a high level of consistency in decision making for CHD and depression between the U.S. and the U.K. Most physicians in both countries correctly identified conditions depicted in the vignettes, although U.S. doctors engage in more information seeking, are more likely to prescribe medications, and are more certain of their diagnoses than their U.K. counterparts. The absence of any national differences in test ordering is consistent for both of the medical conditions depicted. U.K. physicians, however, were more likely than U.S. physicians to make lifestyle recommendations for CHD and to refer those patients to other providers.

Conclusions. Substantively, these findings point to the importance of patient and provider characteristics in understanding between-country differences in clinical decision making. Methodologically, our use of a factorial experiment highlights the potential of these methods for health services research—especially the estimation of the influence of patient attributes, provider characteristics, and between-country differences in the quality of medical care.

Key Words. Clinical decision making, health disparities, clinical encounter

Disparities in the availability and quality of medical care *within* the United States have been extensively documented over the last several decades and are the subject of an Institute of Medicine report (2003). There is an interest in health care variations *between* different national systems, motivated in part by a desire to learn from the experience of others in order to inform U.S. health policy (Blendon et al. 2003, 2004; Schoen et al. 2004). Comparisons of the United States with other national health care systems, such as the United Kingdom or Canada, often lead to suggestions that too much is done in the United States (with its largely private insurance-based system) while too little is done elsewhere (in predominantly government-directed taxation based systems). Evidence-based medicine (EBM) has emerged as an international health care paradigm (Evidence-Based Medicine Working Group 1992) which promotes the use of tools, like clinical guidelines, to hopefully influence provider decision making, improve the quality of care and reduce both national and eventually international variations.

While there are doubtless geographic variations in health care depending on where the patient lives and the system in which care is received, a strict focus on system-level variation may miss important information about other sources of disparities. For example, much less attention has focused on the independent influence of patient attributes (e.g., gender, age, race/ethnicity and socioeconomic status) and provider characteristics (e.g., medical specialty, gender, age/clinical experience or type of employment), over and above geographic location. The influential Institute of Medicine Report (2003) identified “bias, stereotyping and clinical uncertainty on the part of health care providers” as contributing to disparities and calls for research on the prevalence and influence of these processes. The variable behavior of providers encountering different types of patients is increasingly viewed as an under-researched but important contributor to health care variations (Cooper, Hill, and Powe 2002; Paterson and Judge 2002; Van Ryn 2002; Van Ryn and Fu 2003). These bodies of research point to the question: Do health care disparities result primarily from geography (place or system), or from differences at

Addresses correspondence to John McKinlay, Ph.D., Senior Vice President and Chief Scientist, New England Research Institutes, 9 Galen Street, Watertown, MA 02472. Carol Link, Ph.D., Senior Statistician, Lisa Marceau, M.P.H., Director, Media and Communications, and Amy O'Donnell, M.P.H., Associate Research Scientist, are with the New England Research Institutes, Watertown, MA. Sara Arber, Ph.D., Professor, is with the University of Surrey, Guildford, U.K. and Ann Adams, Ph.D., Principal Research Fellow, is with the University of Warwick, Coventry, U.K.

the level of the doctor–patient encounter (i.e., patient attributes and provider characteristics)?

If exactly the same medical problem is managed differently when presented by different people in different geographic locations or in different systems of care, then health care variations are likely to eventually result. Therefore, the elimination of within- and even between-country health care variations should be sought as much through changes in provider behavior as through system-level changes in the organization and financing of health care. Profound implications could follow from this orientation to research. Rather than treating system-level variation as being in competition with variation from the doctor–patient encounter, these approaches may be viewed as complementing one another.

In this paper, we simultaneously measure the effects of different health care systems, patient attributes, and physician characteristics on disparities in clinical decision making. Specifically, this paper examines the way in which primary care providers in two countries—the United States (with its largely private insurance-based health care system) and the United Kingdom (with its National Health Service [NHS] government-supported, taxation based system)—diagnose and manage two common medical problems (coronary heart disease and depression) when identically presented by “patients” of differing age, gender, race and socioeconomic status. Primary care providers (internists and family practitioners in the United States and general practitioners [GPs] in the United Kingdom) are viewed as “gatekeepers” to the rest of their health systems and more specialized levels of care. Thus, what occurs at the level of the medical encounter (the doctor–patient relationship) may contribute to observed health care variations both within and between countries.

RESEARCH METHODS

Experimental Study Design

The objective of this research is to estimate the unconfounded influence (either singly or in combination) of: (a) *patient attributes* (age, gender, race, and socioeconomic status); (b) *physician characteristics* (gender and years of clinical experience); and (c) *separate health care systems* (the United States or the United Kingdom) on medical decision making when providers are presented with identical signs and symptoms strongly suggestive of two common medical problems (coronary heart disease [CHD] and depression). Factorial experiments (which permit estimation of unconfounded main effects and interac-

tions of any two of the variables listed above) were conducted simultaneously in the United States (Massachusetts) and the United Kingdom (the West Midlands, SE London, and Surrey), focusing on a range of outcomes for each of the two medical problems (Cochran and Cox 1957; Fisher 1990). The rich potential of videotaped scenarios was demonstrated in a study showing that the race and sex of a patient independently influence how physicians manage chest pain (Schulman et al. 1999).

A full factorial of $2^4 = 16$ combinations of patient age (55 versus 75), gender, race (white versus black in the United States, or Afro Caribbean in the United Kingdom) and SES (lower versus higher social class—a cleaner/janitor versus a teacher) was used for the video scenarios. One of the 16 combinations was shown to each physician for each medical problem (2 videos per physician, in random order). The experiment was replicated twice. Eight strata of physician (gender, years of clinical experience [<12 or >22 years]) and country (United States/United Kingdom) characteristics were defined, to generate a total of $16 \times 2 \times 8 = 256$ physicians required to complete the design of both experiments.

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. The “patient” and “physicians” in the United States had American accents, while the very same “patients” and “physicians” in the United Kingdom had English accents. The believability of the accents was checked during field tests of the protocol, before beginning fieldwork. Care was taken to construct a culturally neutral set (U.S. physicians tend to have educational diplomas on the office wall while U.K. GPs have paintings or family photos and memorabilia). Immediately after viewing one selected video for each experiment (in random order), the experimental subjects (the sampled physicians) were asked a range of questions concerning their most likely diagnoses, certainty levels, test ordering, prescriptions, lifestyle recommendations they might make, and other information seeking they would engage in if they encountered the medical problem depicted on the video in their everyday clinical practice. Previous studies have used similar methods with success (McKinlay, Potter, and Feldman 1996; Feldman et al. 1997; McKinlay et al. 1997, 1998, 2002).

The medical conditions (CHD and depression) were selected because: (a) they are among the most common and costly problems presented by older patients to primary care providers (Cohen and Krauss 2003); (b) they represent examples of a well-defined organic medical condition and of a less-well-defined psychosocial phenomenon; (c) they admit a range of diagnostic,

therapeutic, and lifestyle actions; and (d) their reported prevalence differs between the United States and the United Kingdom. An advantage of videotapes (over written scenarios) is that potentially relevant nonverbal indicators (e.g., the “Levine fist” for CHD, or a dejected appearance for depression) can be embedded in the presentation. Scripts for the two medical problems were developed from several tape-recorded role-playing sessions with experienced clinical advisors. “Patients” in the CHD vignette presented with symptoms suggestive of CHD (including, e.g., heartburn, pain in the back between the shoulder blades, stress, and elevated blood pressure). The depressed “patient” presented with six of the seven SIGECAPS (sleep disturbance, decreased interest, guilt, reduced energy, inability to concentrate, poor appetite, and psychomotor retardation) and omitted suicidal ideation as too indicative (American Psychiatric Association 1994).

Physician Sample (The Experimental Subjects)

To be eligible for selection, physicians had to: (a) be internists or family practitioners (in the United States) or general practitioners (in the United Kingdom); (b) have ≤ 12 years clinical experience (graduated between 1989 and 1996) or ≥ 22 years experience (graduated between 1965 and 1979) in order to get clear separation by age; (c) be trained at an accredited medical school in either the United States or the United Kingdom (no foreign medical graduates were included); and (d) be currently working as doctors more than half-time. Screening telephone calls were conducted to identify eligible subjects and an appointment was scheduled for a 1-hour long in-person, one-on-one, structured interview. The required 256 interviews were conducted over a period of 9 months in 2001–2002 (128 throughout Massachusetts, 64 around Warwick and 64 throughout Surrey and SE London, U.K.). Each physician subject was provided a modest stipend to partially offset lost revenue and to acknowledge their participation. The response rates were 64.9 percent in the United States and 59.6 percent in the United Kingdom. Interviewers in each country were carefully trained and certified and frequent transatlantic telephone calls were conducted to ensure standardized interviewing and to minimize interviewer variability (Johannes, McKinlay, and Crawford 1997). Quality control interviews and site visits were conducted and selected tape-recorded interviews were reviewed by supervisors on a regular basis.

As with all scientific experiments, we encountered the perennial trade-off between maintaining control of the experimental design and optimizing the generalizability of the results. Our study design required a total of 256

primary care physicians (128 from the United States and 128 from the United Kingdom). Such a modest number cannot reasonably be selected from both the United States and the United Kingdom and be expected to be representative of each country. Our sampling approach therefore represents a practical compromise. We include representation of rural/urban areas and health facilities of different types and sizes (including hospitals and community health centers) while retaining control and constraining project costs by limiting the geographical areas covered. An attempt to get nation-wide representation in each country with only 256 respondents would be prohibitively expensive.

Statistical Power and Analysis

The balanced factorial design allows the unconfounded estimation of all main effects and two-way interactions. The sample size of 128 in each experiment allows us to detect medium effect size differences of 0.5–0.7, with power exceeding 98 percent. Because the experiment was replicated, a pure error term with 128 degrees of freedom was used to test all effects. Analysis of variance was used to estimate all effects. In the absence of missing data, all effects are orthogonal. Logistic regression was not used for dichotomous variables since a complete model could not be specified without achieving complete separation of the data. Given the sample size, the assumptions of analysis of variance are met due to the central limit theorem (Miller 1986).

Validity of the Experimental Approach

Four precautionary steps were taken to protect against threats to external validity (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. This was achieved by basing the scripts on clinical experience, filming with experienced clinicians present, and by using professional actors/actresses. Second, the subjects (doctors) were specifically asked how typical the “patient” viewed on the videotape was compared with patients they encounter in everyday practice (92 percent considered them either very typical or reasonably typical). Third, the doctors viewed the tapes in the context of their practice day (not at a professional meeting, a course update, or in their home) so that it was likely they encountered real patients before and after they viewed the “patient” in the videotape. Fourth, 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.

RESULTS

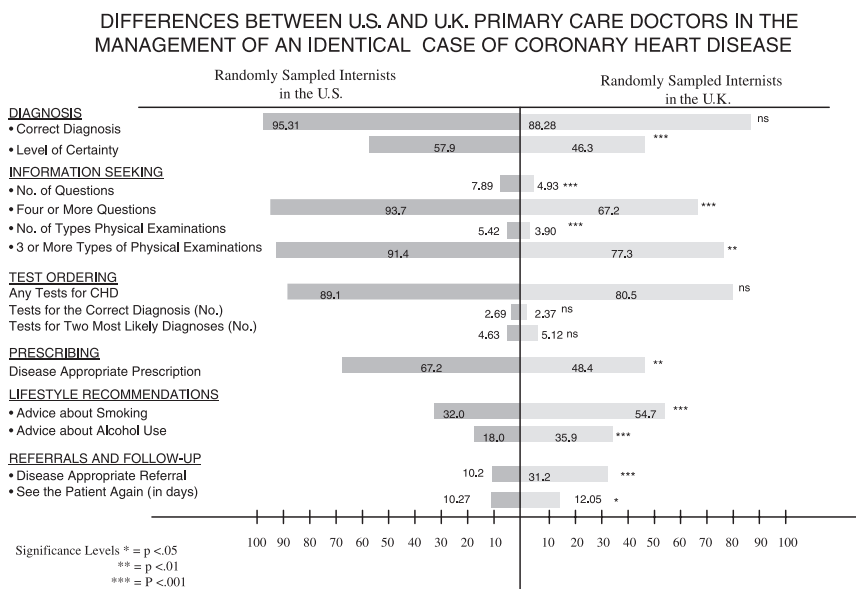
Main Effects

Major results are presented separately for each medical problem (CHD and depression): main effects are described first, followed by a discussion of higher order interactions and the consistency of findings. It should be emphasized that the physician subjects in each country (United States and United Kingdom) encountered (on videotape) exactly the same “patient” (with accents appropriately altered).

Coronary Heart Disease Experiment

Figure 1 summarizes major differences between randomly sampled internists in the United States and GPs in the United Kingdom in the management of an identical presentation of the signs and symptoms of CHD. While there was no

Figure 1: Differences between U.S. and U.K. Primary Care Doctors in the Management of an Identical Case of Coronary Heart Disease.



significant difference in the proportion of primary care doctors mentioning the *correct diagnosis* in each country (95 percent in the United States and 88 percent in the United Kingdom), there was a significant difference in the average *level of certainty* surrounding this diagnosis (58 percent in United States versus 46 percent for the United Kingdom). Between-country differences were also evident in physician *information seeking*, with U.S. internists asking significantly more questions (7.9 versus 4.9) and a greater proportion asking four or more questions of the presenting “patient” (94 versus 67 percent). U.S. physicians would also perform physical examinations on more parts of the body (5.4 versus 3.9 in the United Kingdom) and a higher proportion would perform three or more types of physical examinations (91 versus 77 percent). There were no significant differences in the *test ordering behavior* of the physicians in each country. In terms of *prescribing behaviors*, however, 67 percent of U.S. physicians would write a disease specific prescription, compared with only 48 percent of their GP counterparts in the United Kingdom.

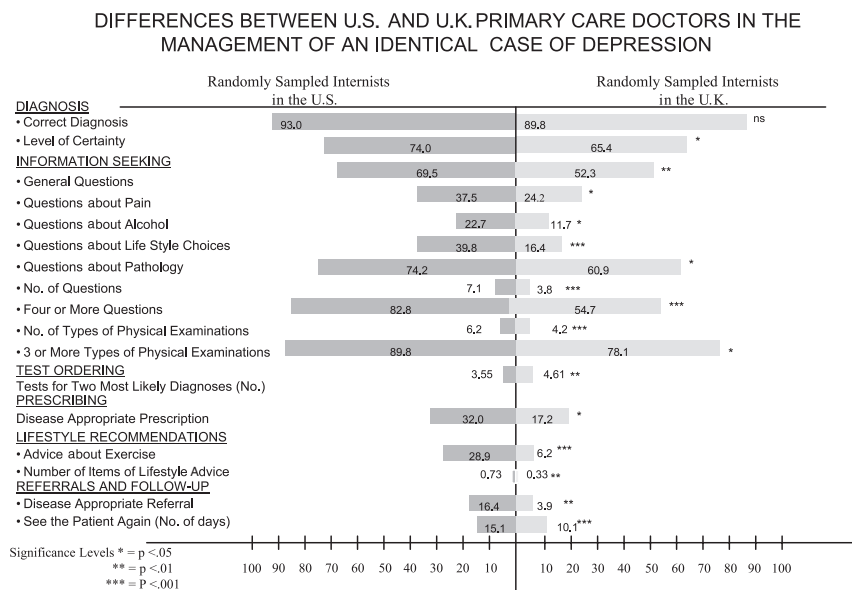
There were also significant differences between the two countries in physicians’ recommendations to patients concerning *lifestyles*, although in the direction of the U.K. physicians doing more than the U.S. physicians. GPs in the United Kingdom were much more likely to give advice about smoking (55 percent versus 32 percent), and twice as likely to offer advice regarding alcohol use (36 versus 18 percent). British GPs were three times more likely to *refer* the “patient” to an appropriate hospital specialist (31 versus 10 percent for U.S. internists) and would wish to see this patient again in significantly more time (12 days in the United Kingdom versus 10 days in the United States).

Overall, in the case of the “patient” with CHD, we find that U.S. physicians were significantly more likely than their U.K. counterparts to be more certain about the diagnosis, engage in more information seeking, and provide more prescriptions. U.K. physicians, on the other hand, were more likely than U.S. physicians to offer lifestyle recommendations, refer patients to other providers, and to wait longer before seeing the patient again.

Depression Experiment

Figure 2 presents the main effects as they pertain to the depression experiment. As with CHD, there was no significant difference in the high proportion of doctors in each country making the *correct diagnosis* (93 percent in the United States and 90 percent in the United Kingdom), but the U.S. physicians expressed greater *certainty* that it was correct (74 percent in the United States versus 65 percent in the United Kingdom). With respect to *information seeking*,

Figure 2: Differences between U.S. and U.K. Primary Care Doctors in the Management of an Identical Case of Depression.



U.S. physicians were again considerably more inquisitive than their GP counterparts in the United Kingdom: they asked significantly more general questions, more questions about specific topics (pain, alcohol, lifestyle choices, and pathology), more questions overall, and performed more types of physical examinations than their U.K. counterparts. The broader range of clinical actions mentioned here (compared with the case of CHD) probably reflects the more diffuse presentation of symptoms that often occurs with depression. Similar to the CHD experiment, there were few significant differences in the *test ordering* behavior of doctors in the two countries, although U.K. physicians were significantly more likely than U.S. physicians to test for the two most likely diagnoses. When encountering exactly the same “patient” presenting with the signs and symptoms of depression, GPs in the United Kingdom would be about half as likely to *write a disease specific prescription* compared with their primary care equivalents in the United States (17 versus 32 percent).

Unlike the case of the CHD findings, U.S. physicians were more likely than United Kingdom physicians to give exercise advice to depression “patients” as well as more items of lifestyle advice overall. U.S. and U.K. physicians also handled referrals and follow-up differently in the case of

depression compared with CHD. For depression, internists in the United States were four times more likely to refer the depressed “patient” to a mental health professional (16 versus 4 percent in the United Kingdom); U.S. physicians also suggested waiting a longer period of time before seeing the patient again (10 days in the United Kingdom versus 15 in the United States).

Overall, the depression experiment shows main effects that are largely similar to the findings from the CHD experiment. While both sets of physicians were very likely to have the correct diagnosis, U.S. physicians were significantly more certain of that diagnosis, would seek more types of information from the patient, and would be more likely to prescribe medication than their U.K. counterparts. While in the CHD experiment U.K. physicians were more likely to make lifestyle recommendations, refer patients to other providers, and to wait longer for follow-up, these patterns are the reverse for the depression experiment.

Interactions

Use of factorial experimentation yields not only the main effects as described, but also permits detection of higher-order interactions that may shape results. These higher-order effects are unconfounded—that is, the study design controls for the possible influence of all the other variables. Whereas the main effects address differences in provider behavior across countries, interaction effects allow us to consider the effects of patient attributes on physician decision making in both countries. Table 1 summarizes interaction effects concerning the influence of patient attributes on medical decision making for CHD and depression in the United States and the United Kingdom.

Coronary Heart Disease Experiment

Patient Age. While the main effects showed no significant differences in *test ordering* behavior across the two sets of physicians, interaction effects show that several types of test ordering varied depending on the age of the patient. While older patients in both the United States and United Kingdom have similar *likelihoods of providers ordering tests for CHD* (86 and 87 percent, respectively), that likelihood is higher for middle-aged patients in the United States compared with their U.K. counterparts (92 versus 73 percent). This differential effect of patient age on test ordering for CHD in the two countries is depicted in Figure 3a.

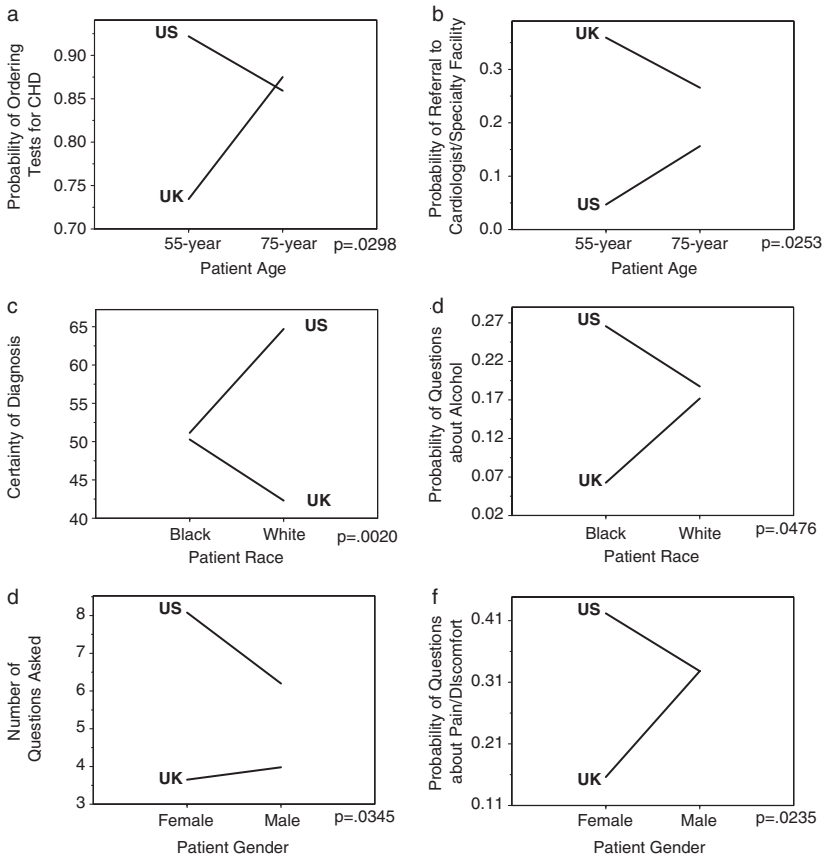
Older patients in the United Kingdom are also significantly more likely to have *more tests* for CHD ordered compared with their middle-aged

Table 1: The Influence of Patient Attributes (Age, Gender, Race) on Physician Decision Making for Coronary Heart Disease in the United States and United Kingdom

<i>Coronary Heart Disease (CHD)</i>					
	<i>Internists in the United States</i>		<i>GPs in the United Kingdom</i>		<i>p-Value</i>
	<i>Older</i>	<i>Middle Aged</i>	<i>Older</i>	<i>Middle Aged</i>	
<i>Patient's Age</i>					
Probability of ordering tests for CHD (%)	86	92	87	73	.0298
Tests ordered for CHD (average no.)	2.67	2.71	3.23	1.65	.0246
Tests ordered for two most likely diagnoses (average no.)	4.70	4.56	5.96	4.35	.0409
Probability of referral to cardiologist/specialty facility (%)	16	5	27	36	.0253
<i>Patient's Race</i>	<i>Black</i>	<i>White</i>	<i>Black</i>	<i>White</i>	<i>p-Value</i>
Certainty of diagnosis (%)	51	65	50	42	.0020
Probability of ordering tests for CHD (%)	83	95	86	75	.0124
Proportion asking questions concerning pain/discomfort (%)	52	73	67	47	.0009
Proportion asking questions about smoking (%)	67	47	47	50	.340
<i>Depression</i>					
	<i>Internists in the United States</i>		<i>GPs in the United Kingdom</i>		<i>p-Value</i>
	<i>Black</i>	<i>White</i>	<i>Black</i>	<i>White</i>	
<i>Patient's Race</i>					
Probability of asking about pain/discomfort (%)	47	28	22	27	.0452
Proportion asking about pathology (%)	80	69	50	72	.0032
Proportion asking about alcohol (%)	27	19	6	17	.0476
<i>Patient's Gender</i>	<i>Female</i>	<i>Male</i>	<i>Female</i>	<i>Male</i>	<i>p-Value</i>
Questions that would be asked (average no.)	8.08	6.20	3.65	3.98	.0345
Proportion asking questions about pain/discomfort (%)	42	33	16	33	.0235

Note: For purposes of clarity of presentation, numerous nonsignificant effects have been omitted in order to focus on the significant influences of interest.

Figure 3: Graphical Representation of Interaction Effects: (a) Coronary heart disease (CHD) experiment: the interaction effect of patient age and country setting on test ordering. (b) CHD experiment: the interaction effect of patient age and country setting on referral behavior. (c) CHD experiment: the interaction effect of patient race and country setting on physician certainty. (d) Depression experiment: the interaction effect of patient gender and country setting on physician questioning. (e) Depression experiment: the interaction effect of patient gender and country setting on physician questioning. (f) Depression experiment: the interaction effect of patient gender and country setting on questions asking about pain/discomfort.



counterparts (3.23 and 1.65, respectively), although the average number of tests that a U.S. physician would order for CHD is about the same for middle-aged and older patients. When considering the *two most likely diagnostic possibilities* mentioned, U.K. physicians would again order significantly more tests for their older patients than would U.S. providers (6.0 in the United Kingdom versus 4.7 in the United States), while middle-aged patients would have about the same number of tests ordered in both countries. *Referral behavior* also varies between countries according to patient age. Main effects show that, irrespective of their age, patients presenting with CHD in the United Kingdom are more likely to be referred to a cardiologist or specialty facility. Between each country, however, a patient's age has a different effect: while middle-aged patients are more likely to be referred in the United Kingdom than in the United States, these likelihoods converge among older patients (see Figure 3b). Overall, the effect of patient age on provider behavior is one wherein U.K. physicians are more likely than U.S. physicians to order tests for elderly relative to middle-aged patients, and are also less likely to refer those patients out relative to U.S. physicians. Patient age did not co-vary with diagnosis behavior, information seeking, prescribing, or lifestyle recommendations.

Patient Race. Interaction effects show that providers' diagnostic certainty, information seeking, and test ordering also vary between countries depending on the race of the patient. While doctors in the United States and the United Kingdom show similar *levels of certainty* with black patients, U.S. physicians are significantly more certain with their white patients, while U.K. physicians are less certain with whites (see Figure 3[c]). Consistent results are also evident with respect to the *probability of test ordering*: while black patients have about the same probability in the United States and the United Kingdom, white patients in the United States are significantly more likely to have tests ordered for CHD than their counterparts in the United Kingdom. *Information seeking behavior* also varies by race, with white patients in the United States being more likely to be asked questions concerning pain/discomfort relative to blacks, while the reverse holds true for the United Kingdom. Finally, while physician questioning about smoking varied little according to patient race in the United Kingdom, black patients in the United States are significantly more likely than white to be questioned about smoking. The overall pattern in these interaction effects is one where white patients in the United States experience greater physician certainty about

their diagnosis, have more tests ordered on their behalf, and receive more questions about their discomfort; in the United Kingdom, these patterns are reversed. There were no significant interactions between country setting and either age or socioeconomic status in the management of CHD.

Depression Experiment

The significant second-order interactions between a patient's race or gender and their country setting (the United States or the United Kingdom) in the management of depression are also summarized in Table 1.

Patient Race. With respect to the effect of a patient's race, we observe generally consistent patterns in *information seeking* behavior. The same proportions of white depressed patients in both countries would be asked questions about pain/discomfort, pathology and alcohol consumption, while black depressed patients in the United States would be significantly more likely to be asked such questions. Figure 3d summarizes this interaction with respect to questions concerning alcohol.

Patient Gender. A depressed patient's gender also affects the *information seeking* behavior of physicians across countries. Physicians in the United States would ask significantly more questions overall and female patients would be asked more questions than their male counterparts. Patients in the United Kingdom would be asked fewer questions overall, with no apparent differences by gender. These between and within-country differences in number of questions asked are depicted in Figure 3(e). Questions asking about pain/discomfort do not vary by country if the patient is male, but female patients in the United Kingdom are much less likely to be questioned about pain/discomfort—a depressed female patient in the United States is nearly three times more likely to be questioned on this subject (see Figure 3[f]).

SUMMARY AND IMPLICATIONS

Overall, there is a high level of consistency in decision making for the two different medical problems (CHD and depression) and between the two countries studied (see Table 2). For both CHD and depression a very high proportion (around 90 percent) of doctors in each country selected the correct diagnosis as

Table 2: Consistency of Principal Results across the Two Experiments (CHD and Depression) and between the Two Countries (United States and United Kingdom)

	Experiment No. 1 (CHD)		Experiment No. 2 (Depression)		Consistency across the Two Experiments
	United States	United Kingdom	United States	United Kingdom	
Diagnosis					
Correct diagnosis (%)	>90	>85	>90	>85	✓
Level of certainty	Higher	Lower	Higher	Lower	✓
Information seeking					
Number of questions	More	Fewer	More	Fewer	✓
Four or more questions	Higher	Lower	Higher	Lower	✓
Number of things examined	More	Fewer	More	Fewer	✓
Three or more things examined	Higher	Lower	Higher	Lower	✓
Test ordering					
Any tests for the problem	No difference		No difference		✓
Tests for the correct diagnosis	No difference		No difference		✓
Tests for two most likely diagnoses	No difference		Lower	Higher	X
Prescribing					
Disease specific prescription	Higher	Lower	Higher	Lower	✓
Lifestyle recommendations	Not comparable		Across experiments		
Advice about smoking	—		—		
Advice about alcohol	—		—		
Advice about exercise	—		—		
Referrals and follow-up					
Disease appropriate referral	Lower	Higher	Higher	Lower	X
See the patient again	No	Difference	Later	Sooner	X

✓, consistent results between the experiments; X, inconsistent results between the experiments; —, no comparable results between the experiments.

one of the possible diagnoses. Doctors in the United Kingdom, however appear to be less certain of their diagnoses—a difference evident for both medical conditions. U.S. internists also appear to be more inquisitive than their GP counterparts in the United Kingdom. Regardless of the illness condition they ask more questions of the “patient” and would examine more things. The absence of any national differences in test ordering is consistent for both of the medical conditions depicted. U.S. internists are significantly more likely to prescribe medications for both of the illness presentations. The tendency for U.K. doctors to make significantly more lifestyle recommendations for CHD was not pursued for depression because contributing risk behaviors are less well understood.

The tendency for U.K. doctors to wish to see the “patient” in a shorter period of time is not consistent for both illness conditions. Interestingly, while U.S. internists are four times more likely to refer the depressed “patient” for specialist care (mainly to a psychiatrist or psychologist), they are three times *less* likely to refer the case of CHD to a cardiologist or for specialist care. This apparent inconsistency may be explained by national differences in health care financing and physician competition. Internists in the United States may view a case of depression as burdensome and costly in time and other resources (several months of repeat visits could be required). Referral of these cases may reflect economic expediency. In the highly competitive U.S. health care system (especially with physician oversupply in the Northeast) failure to refer the “patient” with CHD to specialists may reflect a fear of losing the case, thereby costing business for a physician’s employing organization.

Important implications follow from both the methodology and the results presented in this paper. Substantively, these results show that patient and provider characteristics are critical for explaining variations in health care and clinical decision making, even in a cross-cultural context. The use of a factorial experiment permits simultaneous examination of different types of influence—patient, provider, and health care system. Consistent with earlier research (Arber et al. 2004, 2006; Adams et al. 2006), we find that selected patient attributes (e.g., age, race, gender but not SES) and provider characteristics (e.g., physician gender) influence decision making for the conditions we studied in both countries. Over and above these influences, however (and controlling for them) we detect significant differences between the two national health care systems. How a “patient” with either CHD or depression is managed depends not only on who they are (patient attributes) and who they encounter (doctor’s gender and years of experience), but even more on the health care system in which the interaction occurs.

Methodologically, our focus on patient variables, provider characteristics, and health system influences within a single study represents a somewhat new direction in health services research on clinical decision making. When attempts are made to examine the contribution of these different influences on CDM, they typically employ multivariate analyses of large observational datasets (e.g., Medicare Outcomes data). Unfortunately, such analyses are unable to produce unconfounded estimates of the relative contribution of patient influences. This is only possible through the type of factorial experimentation illustrated in this paper. To the extent that health services research continues to be interested in estimating patient, provider, and organizational influences on clinical decision making, it is important to extend beyond a focus on a particular type or level of influence (especially patient attributes) at the expense of understanding other potentially important influences (provider characteristics and organizational/system features).

ACKNOWLEDGMENTS

The authors are grateful to Timothy Guiney, M.D., Alan Goroll, M.D., Theodore Stern, M.D., John Stoeckle, M.D. (Massachusetts General Hospital, Harvard Medical School, Boston), and David Armstrong, M.D., and Mark Ashworth, M.D. (United Medical School of Guys and St. Thomas', London) and Diane Ackerley, M.D. (Guildford and Waverley Primary Care Trust). Ann Adam's post is funded by a Department of Health NCCRC Primary Career Scientist Award.

Grant Support: This project is supported by Grant No. AG 16747 from the National Institute on Aging, NIH.

Disclosures: All authors attest that they have no financial interest conflicting with complete and accurate reporting of the study findings.

Disclaimers: None

REFERENCES

- Adams, A., C. D. Buckingham, S. Arber, J. B. McKinlay, L. D. Marceau, and C. Link. 2006. "The Influence of Patient Age on Clinical Decision-Making about Coronary Heart Disease in the US and the UK." *Ageing and Society* 26(2): 303–21.
- American Psychiatric Association. 1994. *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edition. Washington, DC: American Psychiatric Association.
- Arber, S., J. B. McKinlay, A. Adams, L. D. Marceau, C. Link, and A. O'Donnell. 2004. "Influence of Patient Characteristics on Doctors' Questioning and Lifestyle Advice for Coronary Heart Disease: A US/UK Video Experiment." *British Journal of General Practice* 54 (506): 673–8.

- . 2006. "Patient Characteristics and Inequalities in Doctors Diagnostic and Management Strategies Relating to CHD: A Video-Simulation Experiment." *Social Science and Medicine* 62(1): 103–15.
- Blendon, R. J., C. Schoen, C. DesRoches, R. Osborn, and K. Zapert. 2003. "Common Concerns amid Diverse Systems: Health Care Experiences in Five Countries." *Health Affairs* 22 (3): 106–21.
- Blendon, R. J., C. Schoen, C. M. DesRoches, R. Osborn, K. Zapert, and E. Raleigh. 2004. "Confronting Competing Demands to Improve Quality: A Five-Country Hospital Survey." *Health Affairs* 23 (3): 119–35.
- Cochran, W. G., and C. M. Cox. 1957. *Experimental Designs*. New York: John Wiley & Sons Inc.
- Cohen, J. W., and N. A. Krauss. 2003. "Spending and Service Use among People with the Fifteen Most Costly Medical Conditions." *Health Affairs* 22 (2): 129–38.
- Cooper, L. A., M. N. Hill, and N. R. Powe. 2002. "Designing and Evaluating Interventions to Eliminate Racial and Ethnic Disparities in Health Care." *Journal of General Internal Medicine* 17 (6): 477–86.
- Evidence-Based Medicine Working Group. 1992. "Evidence-Based Medicine. A New Approach to Teaching the Practice of Medicine." *Journal of the American Medical Association* 268 (17): 2420–5.
- Feldman, H. A., J. B. McKinlay, D. A. Potter, K. M. Freund, R. B. Burns, and M. A. Moskowitz. 1997. "Non-Medical Influences on Medical Decision Making: An Experimental Technique Using Videotapes, Factorial Design, and Survey Sampling." *Health Services Research* 32 (3): 343–65.
- Fisher, R. A. 1990. *Statistical Methods, Experimental Design and Scientific Inference*. New York: Oxford University Press.
- Institute of Medicine. 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: The National Academies Press.
- Johannes, C. B., J. McKinlay, and S. Crawford. 1997. "Interviewer Effects in a Cohort Study." *American Journal of Epidemiology* 146: 429–38.
- McKinlay, J. B., R. B. Burns, R. Durante, H. A. Feldman, K. M. Freund, B. S. Harrow, J. T. Irish, and L. E. Kasten. 1997. "Patient, Physician and Presentational Influences on Clinical Decision Making for Breast Cancer: Results from a Factorial Experiment." *Journal of Evaluation in Clinical Practice* 3 (1): 23–57.
- McKinlay, J. B., R. Burns, H. A. Feldman, K. M. Freund, J. T. Irish, L. E. Kasten, M. A. Moskowitz, D. A. Potter, and K. Woodman. 1998. "Physician Variably and Uncertainty in the Management of Breast Cancer." *Medical Care* 36 (3): 385–96.
- McKinlay, J. B., L. Ling, K. Freund, and M. Moskowitz. 2002. "The Unexpected Influence of Physician Attributes on Clinical Decisions: Results of an Experiment." *Journal of Health and Social Behavior* 43: 92–106.
- McKinlay, J. B., D. Potter, and H. Feldman. 1996. "Non-Medical Influences on Medical Decision-Making." *Social Science and Medicine* 42: 769–76.
- Miller, R. J. 1986. *Beyond ANOVA, Basics of Applied Statistics*. New York: John Wiley & Sons.

- Paterson, I., and K. Judge. 2002. "Equality of Access to Health Care." In *Reducing Inequalities in Health. A European Perspective*, edited by J. Mackenbach and M. Bakker, pp. 169–87. London: Routledge.
- Schoen, C., R. Osborn, P. T. Huynh, M. Doty, K. Davis, K. Zapert, and J. Peugh. 2004. "Primary Care and Health System Performance: Adults' Experiences in Five Countries." *Health Affairs*: W4: 487–503.
- Schulman, K. A., J. A. Berlin, W. Harless, J. F. Kerner, S. Sistrunk, B. J. Gersh, R. Dube, C. K. Taleghani, J. E. Burke, S. Williams, J. M. Eisenberg, and J. J. Escarce. 1999. "The Effect of Race and Sex on Physicians' Recommendations for Cardiac Catheterization." *New England Journal of Medicine* 340 (8): 618–26.
- Van Ryn, M. 2002. "Research on the Provider Contribution to Race/Ethnicity Disparities in Medical Care." *Medical Care* 40 (1): 140–51.
- Van Ryn, M., and S. S. Fu. 2003. "Paved with Good Intentions: Do Public Health and Human Service Providers Contribute to Racial/Ethnic Disparities in Health?" *American Journal of Public Health* 93 (2): 248–55.