NIH Public Access
Author Manuscript
Health Policy. Author manuscript; available in PMC 2010 March 1.

Published in final edited form as:
Health Policy. 2009 March ; 89(3): 279–287. doi:10.1016/j.healthpol.2008.06.007.

Is certainty more important than diagnosis for understanding race and gender disparities?: An experiment using coronary heart disease and depression case vignettes

Karen E. Lutfey, Ph.D.[2,1], **Carol L. Link, Ph.D.**[2], **Richard W. Grant, M.D., M.P.H.**[3], **Lisa D. Marceau, M.P.H.**[2], and **John B. McKinlay, Ph.D.**[2]

2 *New England Research Institutes, 9 Galen Street, Watertown, MA, 02472, USA*

3 *Division of General Medicine, Massachusetts General Hospital, and Harvard Medical School, Boston, MA*

Abstract

Objectives—To: (1) examine the influence of patient and provider attributes on physicians' diagnostic certainty and (2) assess the effect of diagnostic certainty on clinical therapeutic actions.

Methods—Factorial experiment of 128 generalist physicians using identical clinically authentic videotaped vignettes depicting patients with coronary heart disease (CHD) or depression.

Results—For CHD, physicians were least certain for Black patients (p = .003) and for younger female patients (p = .013). For depression, average certainty was higher than for the CHD presentation (74.0 vs. 57.9 on of scale of 0–100, p < .001) and there were no main effects of patient or provider characteristics. Increasing diagnostic certainty was a significant predictor of subsequent clinical actions, and these varied according to physician and patient characteristics across both conditions.

Conclusions—Physicians were least certain of their CHD diagnoses for Black patients and for younger women, but patient characteristics alone did not affect physician certainty of depression diagnoses. Physicians responded differentially to diagnostic certainty in terms of their clinical therapeutic actions such as test ordering and writing prescriptions. Physician responses to certainty may be as important as their responses to patient characteristics for understanding variation in clinical decision making.

Introduction

Health inequalities—disparities observed across gender, race, and socioeconomic boundaries—are an ongoing source of cost and worry in arenas of public health, healthcare delivery, and health policy. Differences in clinical decision making (CDM) have been observed for many conditions, ranging from coronary heart disease [1–5] to schizophrenia [6,7], and for several types of clinical decisions, such as making diagnoses, ordering tests [1,2,5], writing prescriptions [8,9], and assessing patient adherence [10,11]. Historically, explanations for the sources of these gradients have focused on individual patient factors such as health care access and utilization, health-related behaviors, and home and neighborhood environments. More

1Corresponding author. Tel: (617) 923-7747 ext. 313; email: E-mail: klutfey@neriscience.com.

There were no conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

recently, attention has expanded from patients to include healthcare providers and doctor-patient interaction as possible sources of inequality [12–17].

A sizeable literature has developed that focuses on physicians' social and psychological information processing during the patient-provider encounter as an important source of observed variation in CDM behavior. Previous work has evaluated how prejudice, stereotyping, and discrimination can affect providers' assessments of patients and thereby influence treatment decisions [14,15,18–23]. Even unprejudiced physicians may make biased clinical decisions, however, and the concept of uncertainty is used to describe another type of unintentional, implicit bias physicians may have in the process of cognitively organizing information. Balsa and McGuire (2003, 2005), for example, use the term "statistical discrimination" to describe bias resulting from prior probabilities (epidemiologic base rates) overwhelming presenting data (individual patients) when physicians are uncertain. Empirically, previous studies have largely focused on (1) which patient characteristics lead physicians to be most biased, with particular attention to racial differences [24], (2) which patient-physician pairs lead to the greatest uncertainty, especially whether racially-matched dyads have better outcomes [25], and (3) distinguishing between types of cognitive processes driving these differences, particularly evidence of prejudice versus uncertainty [11,24].

Implicit in this body of work is an assumption that bias in clinical decision making is primarily a function of patient characteristics, physician attributes, and how various combinations potentially influence physicians' cognitive processing. That is, if physicians treat patients unequally, it is due to a shortcoming in the collection and processing of patient information, not in the decisions they make once they have assessed patients. We extend this line of work by examining physicians' diagnostic certainty. We find that certainty is not only associated with the process of assessing and diagnosing patients, as has been shown in previous studies, but *also* exerts an independent influence on physicians' subsequent clinical actions. The implications of these results are critical, as they imply that the "fix" for variations in CDM may lie not only with reducing physicians' uncertainty with specific types of patients, but also in reducing variations in how they respond clinically to their own uncertainty.

Using data from a video vignette factorial experiment, we address the following research questions: (1) How certain are physicians of their diagnoses for conditions presented by two patients? (2) Which types of physicians have the highest diagnostic certainty? (3) Which types of patients elicit the highest certainty levels among physicians? (4) Are specific types of physicians more certain with specific types of patients? (5) How does certainty about diagnosis affect subsequent patient management, such as information seeking, test ordering, prescribing, lifestyle recommendations, and referrals/follow-up? (6) How do these processes vary according to the patients' presenting conditions (coronary heart disease [CHD] vs. depression)?

Materials and Methods

We used a factorial experiment to simultaneously measure the effects of: (a) patient attributes (age, gender, race and socioeconomic status) and (b) physician characteristics (gender and years of clinical experience) on physician diagnostic certainty and subsequent 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 (study details are described in McKinlay et al 2006). A full factorial of $2^4 = 16$ combinations of patient age (55 vs. 75), gender, race (White vs. Black) and socioeconomic status (SES) (lower vs. higher socioeconomic status, depicted by current or former employment as a janitor or 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). The vignette pairs were assigned such that each vignette depicted opposite social characteristics for each medical condition (for

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

example, if the depression vignette showed an older, black, low SES female, the CHD vignette that physician was shown would be a younger, white, high SES male). This random ordering was designed to avoid confounding by the order of presentation or by the social position of the "patient." IRB approval was obtained for the study, and signed informed consents were collected from each participating physician.

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 [26]; b) they represent examples of a well-defined organic medical condition and of a less-well-defined psychosocial phenomenon; and c) they admit a range of diagnostic, therapeutic and lifestyle actions. 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, for example, chest pain worsening with exertion, pain in the back between the shoulder blades, stress and elevated blood pressure). The depressed patient presented with six of the seven classic symptoms of depression (sleep disturbance, decreased interest, guilt, reduced energy, inability to concentrate, poor appetite and psychomotor retardation). Suicidal ideation was omitted as too indicative of the diagnosis [27]. Professional actors were selected for their comparability (in weight, attractiveness, etc.) and trained (under experienced physician supervision) to realistically and consistently portray a patient presenting with the signs/symptoms of disease to a primary care provider. Previous studies have used similar methods with success [8,28–32].

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 0–100. They were also asked how they would treat the patient in terms of asking for additional information, performing physical examinations, ordering tests, prescribing medications, giving lifestyle advice, and referring to other physicians.

The original study included data from the United States and United Kingdom (as addressed in McKinlay et al 2006), but the present analyses focus only on the US data. To be eligible for selection, physicians had to: (a) be internists or family practitioners; (b) have ≤12 years clinical experience (graduated between 1989–96) or ≥22 years experience (graduated between 1965–79) in order to get clear separation by level of experience; (c) be trained at an accredited medical school in the US (no foreign medical graduates were included); and (d) be currently working in Massachusetts as physicians more than half-time. Screening telephone calls were conducted to identify eligible subjects and an appointment was scheduled for a one-hour long in-person, one-on-one, structured interview. The required 128 interviews (16 pairs of vignettes × 2 physician genders × 2 physician levels of experience × 2 replications) were conducted over a period of nine months in 2001–2. Each physician subject was provided a modest stipend ($100) to partially offset lost revenue and to acknowledge their participation. The response rate was 64.9% (completed appointments as a percentage of eligible physicians). Interviewers were carefully trained and certified, and quality control interviews were conducted and selected tape-recorded interviews were reviewed by supervisors on a regular basis. Physicians were blind to study hypotheses.

Every study represents a balance between internal and external validity. Because they are deliberately designed to control unwanted variability, experiments in particular are necessarily conducted under special circumstances or in an "unreal" setting. Determining whether findings from this "unreal" world of experiments are applicable to the "real world" of everyday social behavior is a perennial challenge to behavioral researchers. For the present study, external

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

validity concerns include not only whether physicians would diagnose vignette patients the same way they would actual patients, but also whether they would take the same clinical actions under experimental conditions as in their regular practice. In numerous peer-reviewed publications over a decade we [9,31,33] and other investigators [34–49] have presented results which show clinical vignettes work in clinical settings (i.e., they produce unconfounded estimates of the influence of systematically manipulated variables). In medical decision making, vignettes have been validated for predicting variation in the quality of preventive care [35] and in measuring the quality of physician practice [50]. In a direct comparison of vignettes, standardized patients, and chart abstraction, Peabody and colleagues [51] found that vignettes were a valid and comprehensive method for measuring quality of outpatient care, and that vignettes consistently produced results that were closer to standardized patients than chart abstraction results were. Additional studies comparing vignettes with standardized patients and other methods corroborate the result that vignettes are ecologically valid for studies of medical decision making [35,47,52,53].

For present purposes, a critical benefit of vignettes is that they 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” [53]. While standardized patients are considered especially useful for measuring real-time communication and physical examination skills [54], these topics are beyond the scope of our research questions. More importantly, such an approach would not permit exacting experimental control (absolutely identical presentations). We took four precautionary steps in an attempt to minimize possible 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—that is, to develop a script and videotaped vignette that was a clinically authentic presentation of the symptoms (e.g., CHD or depression) and the range of patients presenting them (e.g., men, women, black, white, older, younger). While there may be real-life differences in how various types of patients tend to present the same conditions, our vignette development process was designed to result in a presentation that would be plausible for the range of patient characteristics presented in the study. 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% 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.

Analysis of variance was used to test the main effects and two-way interactions of the design variables (patient gender, race, age, and SES, and physician gender and level of experience) on the diagnostic certainty (0–100, with 0 for complete uncertainty and 100 for complete certainty). If the physician did not make a CHD or depression diagnosis, his or her certainty for these diagnoses was set to 0. The balanced factorial design allows the unconfounded estimation of all main effects and two-way interactions using analysis of variance. Furthermore, the sample size of 128 allows us to detect a difference in certainty of 15 points with 80% power (that is, a true 15 point difference in certainty between two groups will be detected 80 percent of the time at $\alpha = 0.05$). Because the experiment was replicated, a pure error term with 128 degrees of freedom was used to test all effects using analysis of variance. To determine the effect of certainty on clinical decision making, logistic regression was used for dichotomous variables (e.g., whether or not an EKG was ordered) and analysis of covariance was used for

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

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 backwards elimination, non-significant effects (at the 0.05 level) other than certainty were removed from the model, leaving a parsimonious model. A paired t-test was used to test the difference between the certainty for the CHD vignette and the depression vignette. We acknowledge the limitations of multiple testing, but note that we observe consistency across the two conditions, and that the results we observe at the $p < .01$ level are unlikely to change. To facilitate interpretation, we present actual *p*-values, unadjusted for multiple testing, as this approach allows readers to choose their preferred level of significance.

Results

CHD vignette

For the CHD vignette, 95% of physicians made the correct diagnosis, and on a scale from 0–100, they had an average certainty of 57.9 for their CHD diagnoses (Figure 1). Because the vast majority of physicians correctly identified the condition, there were no significant main effects for identifying the correct diagnosis. While physician characteristics such as gender and level of experience did not, by themselves, predict variation in certainty, there were differences according to patient characteristics (Table 1). Physicians were less certain in making a CHD diagnosis for women than for men with otherwise identical presentations (53.5 vs. 62.3, $p = .048$). An interaction effect between age and gender resulted in physicians being least certain with younger women (47.4 for 55-year-old women vs. 67.3 for 55-year-old men, $p < .013$) (Figure 2a). The main effect for race was also significant, with physicians being less certain in making a diagnosis of CHD with Black than White patients (51.2 vs. 64.7, $p = .003$).

In turn, physicians' diagnostic certainty for CHD significantly influenced their subsequent diagnostic and therapeutic clinical actions (Tables 2a and 2b). For every 10-point increase in certainty, physicians were slightly less likely to seek additional information about patients' social circumstances (OR 0.86, $p < .035$). In terms of test ordering, physicians were more likely to order any CHD-appropriate test (OR 1.75, $p < .001$), a cardiac stress test (OR 1.39, $p < .001$), or an ECG/EKG (OR 1.75, $p < .001$) for each 10-point increase in certainty. Independent of certainty, male and female physicians ordered the same number of tests (3.1 vs. 3.3, $p = .664$). However, at 0 certainty, female physicians would order more tests (2.23 vs. 0.46, $p < .001$), and for every 10 point increase in certainty, the number of CHD tests ordered by male physicians increased more rapidly than for females (0.46 for males, $p < .001$ versus 0.18 for females, $p = .023$) (see Figure 2b). With increased diagnostic certainty, physicians were also more likely to write a CHD-appropriate prescription (OR 1.69, $p < .001$), and more likely to request a follow up visit sooner (.77 day sooner per 10 point increase in certainty, $p < .001$).

Depression vignette

Relative to CHD, a comparable number of physicians (93%) correctly diagnosed the condition depicted in the vignette, but they had higher average certainty for their depression diagnoses (74.0 vs. 57.9, $p < .0001$) (Figure 1). We found that certainty of depression diagnosis was not significantly influenced by either patient attributes (gender, age, race, and SES) or by physician characteristics (gender and level of experience) ($p > 0.05$ for all comparisons, Table 1). There was a significant interaction effect between patient SES and physician gender, with female physicians having greater certainty with high SES patients than with low SES (81.5 versus 68.3) while male physicians were less certain with high SES patients (67.9 versus 78.7) ($p = .010$) (Figure 2c).

As with CHD, physicians' certainty levels influenced their subsequent clinical actions (Tables 2a and 2b). For each ten point increase in certainty, physicians were more likely to seek

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

additional information about patients' psychological states (OR 1.33, $p = .002$), and for the younger patients, ask more questions about social circumstances (OR 1.18, $p = .031$). For each 10-point increase in certainty, physicians were more likely to write a depression-specific prescription (OR 1.72, $p < .001$), more likely to refer the patient to a psychiatric professional (OR 2.08, $p = .007$) and less likely to refer the patient to another medical (non-psychiatric) professional (OR 0.83, $p = .045$).

DISCUSSION

We conducted a factorial experiment to derive unconfounded estimates of the simultaneous effects of patient and provider characteristics on the clinical decision-making process for two common diagnoses: CHD and depression. While the vast majority of physicians successfully diagnosed the conditions depicted in the vignettes, there was significant variation in their diagnostic certainty, which was the focus of our analyses. We found that diagnostic certainty was influenced by patient race and gender for CHD but not for depression, and that for both conditions the degree of diagnostic certainty directly influenced subsequent therapeutic actions.

The scope of this study required the exclusion of some topics that could provide a springboard for future research in this area. For various logistical reasons, the study does not include all the physician information that could potentially be of interest. Physician gender and level of experience were included as design factors because previous research shows that they are relevant to clinical decision making. While it would have been ideal to also vary physician race/ethnicity as a factor in the experimental design, the small numbers of such physicians available within the defined specialties, strata, and geographic area precluded this design. We therefore randomly sampled the specialties selected so that a representative sample of minority physicians would be included, thus mirroring the actual availability of racial and ethnic minority physicians available to the patient population. For similar reasons physician recruitment was random with regard to the type of healthcare setting in which a provider worked, and the interview did not collect information about the racial composition of physicians' typical patient caseloads or the diagnostic and treatment facilities available to patients.

Due to the experimental design of this study, its primary focus was on how physicians respond to the standardized stimulus of the patient in the vignette. For this reason, as discussed above, significant resources were invested in the development of a clinically authentic vignette. Furthermore, by randomly assigning physicians to a specific combination of patient characteristics in a given vignette, any physician differences based on the above factors that influenced selection into the study would also be randomly distributed in the results and therefore not systematically bias the results. Similarly, while we do not have information about differences between those who chose to participate in the study and those who declined, the random assignment of physicians to vignettes assures that selection biases do not affect the internal validity of the experiment. This missing information detracts from the generalizability of the results. Future studies, however, could complement the present study by examining the influenced of such influences on clinical decisions.

Based on these results, we identify three sets of policy and research implications. Our findings are consistent with earlier research showing that physicians have decreased certainty with some types of patients. However, we expand this earlier work by also showing that certainty also has an independent influence on subsequent clinical actions. The robustness of certainty in these models, particularly the differential tolerance of uncertainty exhibited by physicians, implies that the policy strategy of reducing gender- or race/ethnic differences in physicians' clinical assessments is not sufficient to reduce overall disparities. While two physicians may

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

be trained to have the same diagnostic certainty for a given type patient (say a younger female), our results show that their responses to that certainty are different, so that the same younger female patient may not receive treatment if she is seen by a physician whose threshold for clinical action is higher than his counterpart.

Therefore, a first policy implication is that, in order to achieve policy goals concerning the reduction of disparities, we need to understand and standardize variation in physicians' clinical responses to uncertainty (regardless of who the patient is). Such varied responses to uncertainty should be treated as a separate potential source of unequal treatment. Reaching this goal may entail deeper examination of sociological, cultural, economic, and organizational contexts that shape decision making. This type of approach is commensurate with Fennell's [17] call to look beyond the clinical encounter and consider the larger systems in which healthcare is structured. Intervention strategies may include establishing organizational motivators (such as insurance reimbursement) to encourage physicians to take clinical action even if their certainty is low, at least for potentially life-threatening conditions such as CHD. A second, clinical, policy implication is that if physicians are aware of this source of bias, they can make a conscious decision to be more liberal in their ordering of test and prescriptions that would identify patients who are at risk—rather than relying on high certainty levels to trigger clinical action, they could consider test ordering even when their certainty is lower. This strategy may help circumvent persistent difficulties in trying to teach physicians to eliminate their differential diagnostic uncertainty.

Physicians had a higher level of certainty for their diagnoses of depression relative to CHD. While it is difficult to compare certainty across different case vignettes, our results are consistent with the concept that physicians may be hesitant to commit to a single diagnosis (e.g. CHD) at the risk of missing other life-threatening conditions. In the case of the chest pain symptoms, the underlying condition is localized to one of several anatomical systems (cardiac, pulmonary, gastrointestinal, etc.) and therefore implicates several critical alternative diagnoses (e.g., pulmonary embolism, aortic dissection). In contrast, the relatively more benign course of depression, the less localized symptoms, and the decreased risk of incorrect diagnosis compared to the case of chest pain, may explain the greater certainty physicians were able to assign to their depression vignette diagnoses. This result implies that certainty varies differentially across conditions, and therefore physician uncertainty may be more of a problem for conditions with increased risk for catastrophic outcomes or with a greater number of alternative diagnoses. As a result, a third policy-related implication is that our CHD-related strategies should also be considered for other potentially catastrophic conditions; clinically and organizationally, physicians should be encouraged to take preventive action for these conditions even when their certainty is low, recognizing that low certainty may reflect biased responses to individual patients more than true physiological differences.

Finally, these results suggest possible limitations of clinical practice guidelines. Prior studies of CDM have demonstrated that when physicians are less certain, they are more likely to invoke stereotypical information—particularly knowledge associated with patients' social categories—to help fill the gaps in their knowledge with information that may be relevant to diagnostic and treatment decisions [19,24,55,56]. This is most likely to occur when the decision is complex or ambiguous, and the information is perceived as relevant to the clinical decision [55]. While physicians may draw from cultural stereotypes to supplement portions of missing information, "evidence-based" stereotypes (e.g., base rates) may be particularly likely to influence treatment decisions because they are seen as clinically relevant, legitimate sources of information, and providers are encouraged to use them under the rubric of Bayesian decision making [55]. Following from ethnographic research showing that evidence-based guidelines may actually morph through usage into collectively constructed "mindlines," [57] we are

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

concerned that guidelines may have the unintended potential to amplify and legitimate existing physician bias.

Acknowledgements

This study was supported by R01 AG16747 (PI: John B. McKinlay, Ph.D.)

Financial support for this study was provided entirely by a grant from National Institutes of Health, National Institute on Aging (Grant #AG16747). The funding agreement ensured our independence in designing the study, interpreting the data, writing, and publishing the report.

References

1. Ayanian JZ, Epstein AM. Differences in the use of procedures between women and men hospitalized for coronary heart disease. N Engl J Med 1991 Jul 25;325(4):221–5. [PubMed: 2057022]
2. Healy B. The Yentl syndrome. N Engl J Med 1991 Jul 25;325(4):274–6. [PubMed: 2057027]
3. Kannel, WB.; Abbott, RD. Incidence and prognosis of myocardial infarction in women: the Framingham Study. In: Eaker, ED.; Packard, B.; Wenger, NK.; Clarkson, TB.; Tyroler, HA., editors. Coronary Heart Disease in Women Proceedings of an NIH Workshop. New York: Haymarket-Doyma; 1987. p. 208-14.
4. Schwartz LM, Fisher ES, Tosteson NA, Woloshin S, Chang CH, Virnig BA, et al. Treatment and health outcomes of women and men in a cohort with coronary artery disease. Arch Intern Med 1997 Jul 28;157(14):1545–51. [PubMed: 9236556]
5. 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 Jun 13;297(22):2489–95. [PubMed: 17565083]
6. Nasrallah HA, Meyer JM, Goff DC, McEvoy JP, Davis SM, Stroup TS, et al. Low rates of treatment for hypertension, dyslipidemia and diabetes in schizophrenia: data from the CATIE schizophrenia trial sample at baseline. Schizophr Res 2006 Sep;86(1–3):15–22. [PubMed: 16884895]
7. Kelly DL, Dixon LB, Kreyenbuhl JA, Medoff D, Lehman AF, Love RC, et al. Clozapine utilization and outcomes by race in a public mental health system: 1994–2000. J Clin Psychiatry 2006 Sep;67(9): 1404–11. [PubMed: 17017827]
8. McKinlay J, Link C, Arber S, Marceau L, O'Donnell A, Adams A, et al. How do Doctors in Different Countries Manage the Same Patient? Results of a Factorial Experiment. Health Services Research 2006;41(6):2182–200. [PubMed: 17116115]Erratum in 41(6):303
9. 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. Journal of General Internal Medicine 2007;22(3):289–96. [PubMed: 17356957]
10. Lutfey K, Freese J. Toward Some Fundamentals of Fundamental Causality: Socioeconomic Status and Health in the Routine Clinic Visit for Diabetes. American Journal of Sociology 2005;110(5): 1326–72.
11. Lutfey KE, Ketcham JD. Patient and provider assessments of adherence and the sources of disparities: evidence from diabetes care. Health Serv Res 2005 Dec;40(6 Pt 1):1803–17. [PubMed: 16336549]
12. Aberegg SK, Terry PB. Medical decision-making and healthcare disparities: The physician's role. J Lab Clin Med 2004 Jul;144(1):11–7. [PubMed: 15252402]
13. 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 Summer; 14(3):360–71. [PubMed: 15328937]
14. van Ryn M. Research on the provider contribution to race/ethnicity disparities in medical care. Med Care 2002 Jan;40(1 Suppl):I140–51. [PubMed: 11789627]
15. van Ryn M, Fu SS. Paved with good intentions: do public health and human service providers contribute to racial/ethnic disparities in health? Am J Public Health 2003 Feb;93(2):248–55. [PubMed: 12554578]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

16. LaVeist TA, Arthur M, Morgan A, Plantholt S, Rubinstein M. Explaining racial differences in receipt of coronary angiography: the role of physician referral and physician specialty. Med Care Res Rev 2003 Dec;60(4):453–67. [PubMed: 14677220]discussion 96–508

17. Fennell ML. Racial disparities in care: looking beyond the clinical encounter. Health Serv Res 2005 Dec;40(6 Pt 1):1713–21. [PubMed: 16336544]

18. Miller LG, Liu H, Hays RD, Golin CE, Beck CK, Asch SM, et al. How well do clinicians estimate patients' adherence to combination antiretroviral therapy? J Gen Intern Med 2002 Jan;17(1):1–11. [PubMed: 11903770]

19. Balsa AI, McGuire TG. Statistical discrimination in health care. J Health Econ 2001 Nov;20(6):881–907. [PubMed: 11758051]

20. Balsa AI, McGuire TG. Prejudice, clinical uncertainty and stereotyping as sources of health disparities. J Health Econ 2003 Jan;22(1):89–116. [PubMed: 12564719]

21. Institute of Medicine. Unequal Treatment: Confronting Racial and Ethnic Disparities in Healthcare. Washington, D.C: The National Academies Press; 2003.

22. Snowden LR. Bias in mental health assessment and intervention: theory and evidence. Am J Public Health 2003 Feb;93(2):239–43. [PubMed: 12554576]

23. van Ryn M, Burke J. The effect of patient race and socio-economic status on physicians' perceptions of patients. Soc Sci Med 2000 Mar;50(6):813–28. [PubMed: 10695979]

24. Balsa AI, McGuire TG, Meredith LS. Testing for statistical discrimination in health care. Health Serv Res 2005 Feb;40(1):227–52. [PubMed: 15663711]

25. Cooper LA, Roter DL, Johnson RL, Ford DE, Steinwachs DM, Powe NR. Patient-centered communication, ratings of care, and concordance of patient and physician race. Ann Intern Med 2003 Dec 2;139(11):907–15. [PubMed: 14644893]

26. Cohen JW, Krauss NA. Spending and Service Use Among People with the Fifteen Most Costly Medical Conditions. Health Affairs 2003;22(2):129–38. [PubMed: 12674416]

27. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Vol. 4th Revised. Washington, DC: American Psychiatric Association; 2000.

28. Feldman HA, McKinlay JB, Potter DA, Freund KM, Burns RB, Moskowitz MA, et al. Nonmedical influences on medical decision making: an experimental technique using videotapes, factorial design, and survey sampling. Health Serv Res 1997 Aug;32(3):343–66. [PubMed: 9240285]

29. McKinlay JB, Burns RB, Durante R, Feldman HA, Freund KM, Harrow BS, et al. Patient, physician and presentational influences on clinical decision making for breast cancer: results from a factorial experiment. J Eval Clin Pract 1997 Feb;3(1):23–57. [PubMed: 9238607]

30. McKinlay JB, Burns RB, Feldman HA, Freund KM, Irish JT, Kasten LE, et al. Physician variability and uncertainty in the management of breast cancer. Results from a factorial experiment. Med Care 1998 Mar;36(3):385–96. [PubMed: 9520962]

31. McKinlay JB, Lin T, Freund K, Moskowitz M. The unexpected influence of physician attributes on clinical decisions: results of an experiment. J Health Soc Behav 2002 Mar;43(1):92–106. [PubMed: 11949199]

32. McKinlay JB, Potter DA, Feldman HA. Non-medical influences on medical decision-making. Soc Sci Med 1996 Mar;42(5):769–76. [PubMed: 8685745]

33. McKinlay J, Link C, Arber S, Marceau L, O'Donnell A, Adams A, et al. How do Doctors in Different Countries Manage the Same Patient? Results of a Factorial Experiment Health Services Research 2006;41(6):2182–200.Erratum in 41(6):303

34. Kales HC, Neighbors HW, Valenstein M, Blow FC, McCarthy JF, Ignacio RV, et al. Effect of race and sex on primary care physicians' diagnosis and treatment of late-life depression. J Am Geriatr Soc 2005 May;53(5):777–84. [PubMed: 15877552]

35. 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 Nov;15(11):782–8. [PubMed: 11119170]

36. Aberegg SK, Haponik EF, Terry PB. Omission bias and decision making in pulmonary and critical care medicine. Chest 2005 Sep;128(3):1497–505. [PubMed: 16162749]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

37. Barnhart JM, Wassertheil-Smoller S. The effect of race/ethnicity, sex, and social circumstances on coronary revascularization preferences: a vignette comparison. Cardiol Rev 2006 Sep–Oct;14(5): 215–22. [PubMed: 16924161]
38. Currin L, Schmidt U, Waller G. Variables that influence diagnosis and treatment of the eating disorders within primary care settings: a vignette study. Int J Eat Disord 2007 Apr;40(3):257–62. [PubMed: 17167756]
39. Epstein SA, Gonzales JJ, Weinfurt K, Boekeloo B, Yuan N, Chase G. Are psychiatrists' characteristics related to how they care for depression in the medically ill? Results from a national case-vignette survey. Psychosomatics 2001 Nov–Dec;42(6):482–9. [PubMed: 11815683]
40. Kales HC, DiNardo AR, Blow FC, McCarthy JF, Ignacio RV, Riba MB. International medical graduates and the diagnosis and treatment of late-life depression. Acad Med 2006 Feb;81(2):171–5. [PubMed: 16436580]
41. Kales HC, Neighbors HW, Blow FC, Taylor KK, Gillon L, Welsh DE, et al. Race, gender, and psychiatrists' diagnosis and treatment of major depression among elderly patients. Psychiatr Serv 2005 Jun;56(6):721–8. [PubMed: 15939950]
42. Landon BE, Reschovsky J, Reed M, Blumenthal D. Personal, organizational, and market level influences on physicians' practice patterns: results of a national survey of primary care physicians. Med Care 2001 Aug;39(8):889–905. [PubMed: 11468507]
43. Li L, Wu Z, Zhao Y, Lin C, Detels R, Wu S. Using case vignettes to measure HIV-related stigma among health professionals in China. Int J Epidemiol 2007 Feb;36(1):178–84. [PubMed: 17175545]
44. Mitchell TB, Dyer KR, Peay ER. Patient and physician characteristics in relation to clinical decision making in methadone maintenance treatment. Subst Use Misuse 2006;41(3):393–404. [PubMed: 16467013]
45. Morita T, Akechi T, Sugawara Y, Chihara S, Uchitomi Y. Practices and attitudes of Japanese oncologists and palliative care physicians concerning terminal sedation: a nationwide survey. J Clin Oncol 2002 Feb 1;20(3):758–64. [PubMed: 11821458]
46. Peabody JW, Liu A. A cross-national comparison of the quality of clinical care using vignettes. Health Policy Plan 2007 Sep;22(5):294–302. [PubMed: 17660225]
47. Robra BP, Kania H, Kuss O, Schonfisch K, Swart E. [Determinants of hospital admission--investigation by case vignettes]. Gesundheitswesen 2006 Jan;68(1):32–40. [PubMed: 16463243]
48. Sirovich BE, Gottlieb DJ, Welch HG, Fisher ES. Variation in the tendency of primary care physicians to intervene. Arch Intern Med 2005 Oct 24;165(19):2252–6. [PubMed: 16246991]
49. Sox CM, Koepsell TD, Doctor JN, Christakis DA. Pediatricians' clinical decision making: results of 2 randomized controlled trials of test performance characteristics. Arch Pediatr Adolesc Med 2006 May;160(5):487–92. [PubMed: 16651490]
50. Peabody JW, Luck J, Glassman P, Jain S, Hansen J, Spell M, et al. Measuring the quality of physician practice by using clinical vignettes: a prospective validation study. Ann Intern Med 2004 Nov 16;141 (10):771–80. [PubMed: 15545677]
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 Apr 5;283(13):1715–22. [PubMed: 10755498]
52. Braspenning J, Sergeant J. General practitioners' decision making for mental health problems: outcomes and ecological validity. J Clin Epidemiol 1994 Dec;47(12):1365–72. [PubMed: 7730845]
53. Veloski J, Tai S, Evans AS, Nash DB. Clinical vignette-based surveys: a tool for assessing physician practice variation. Am J Med Qual 2005 May–Jun;20(3):151–7. [PubMed: 15951521]
54. Cochran, WG.; Cox, GM. Experimental Designs. New York: John Wiley and Sons; 1957.
55. Burgess DJ, van Ryn M, Crowley-Matoka M, Malat J. Understanding the provider contribution to race/ethnicity disparities in pain treatment: insights from dual process models of stereotyping. Pain Med 2006 Mar–Apr;7(2):119–34. [PubMed: 16634725]
56. Balsa AI, Seiler N, McGuire TG, Bloche MG. Clinical uncertainty and healthcare disparities. Am J Law Med 2003;29(2–3):203–19. [PubMed: 12961805]
57. Gabbay J, le May A. Evidence based guidelines or collectively constructed "mindlines?" Ethnographic study of knowledge management in primary care. Bmj 2004 Oct 30;329(7473):1013. [PubMed: 15514347]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

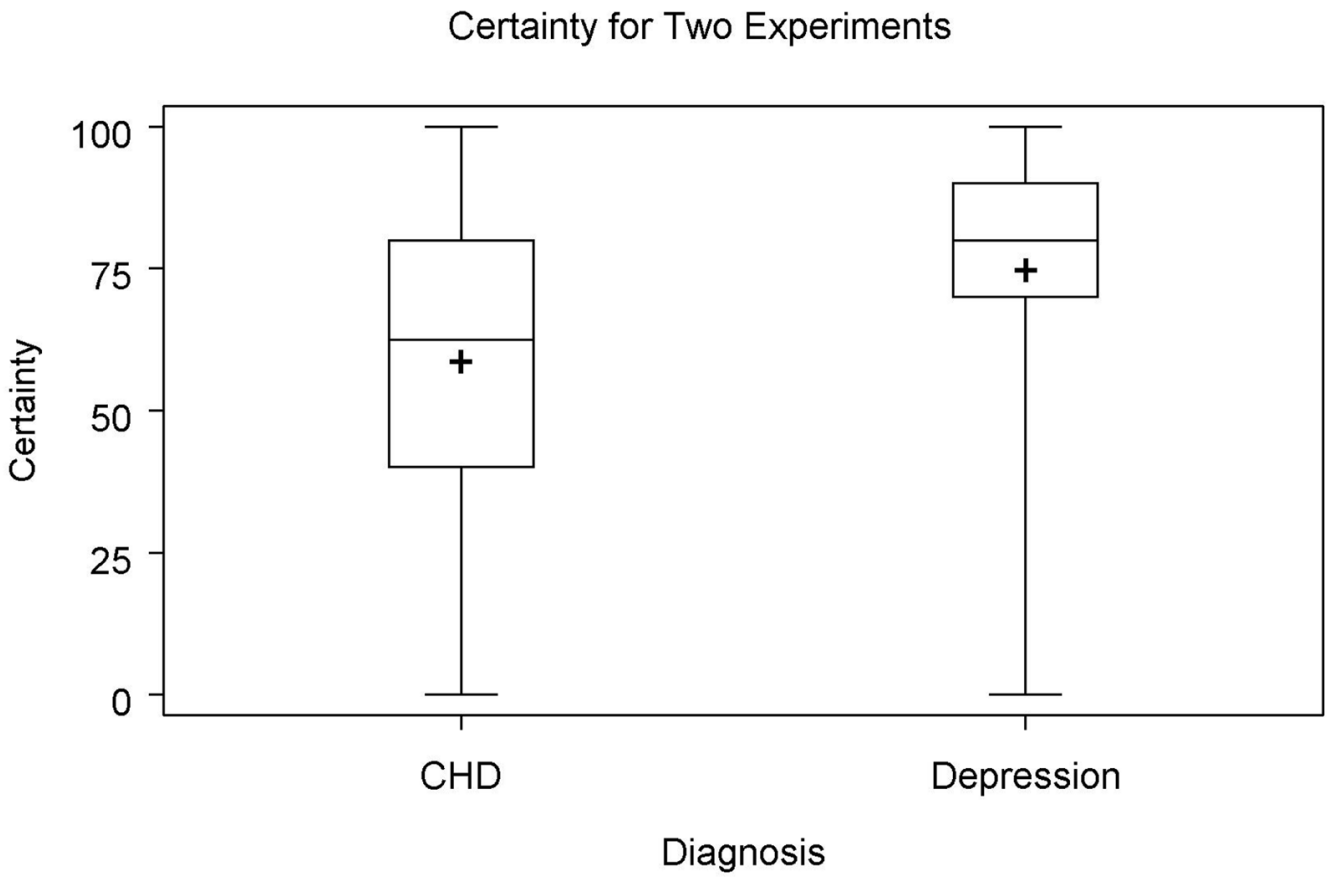
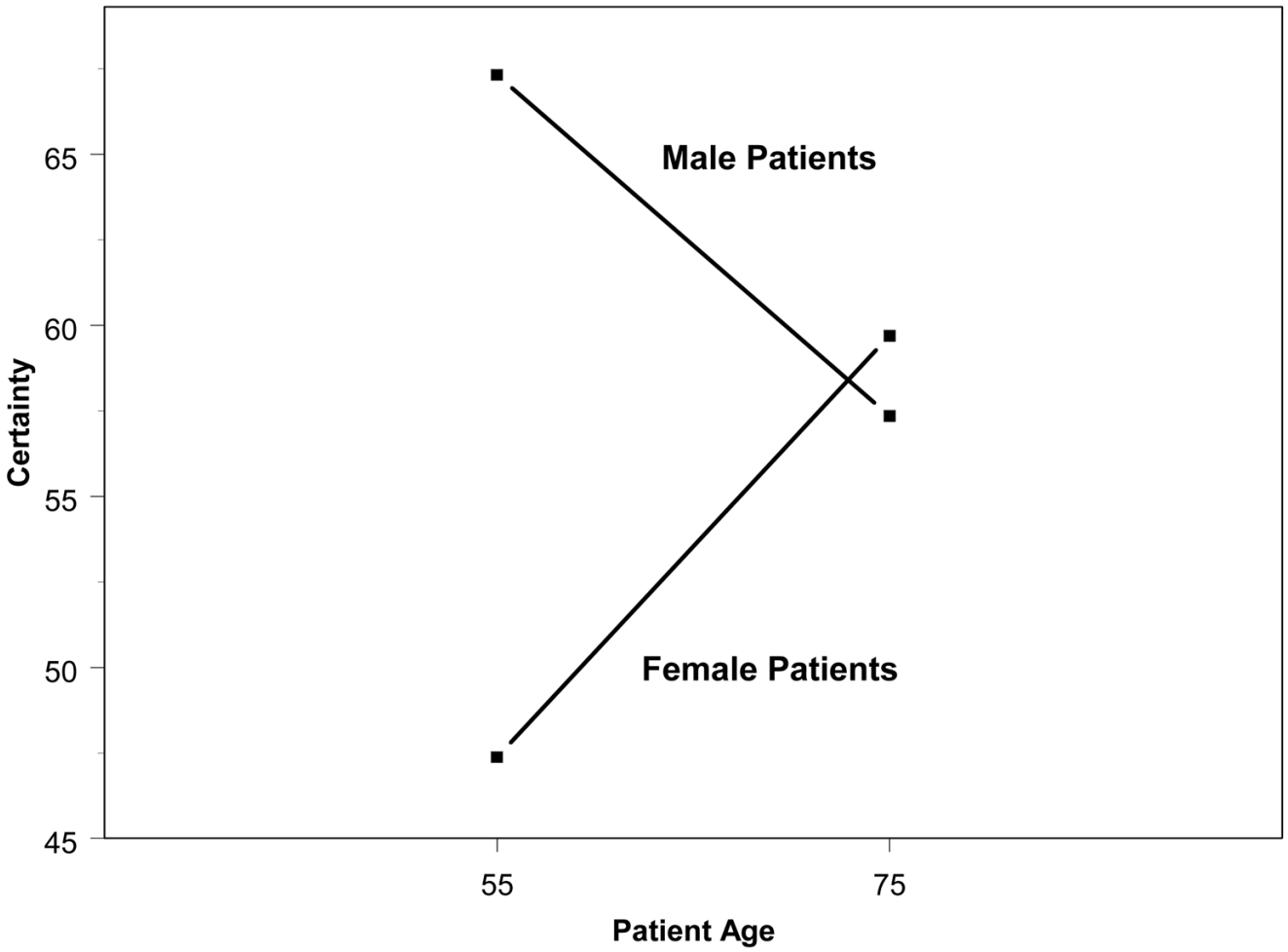


Figure 1. Physician certainty of CHD and Depression diagnoses ($p < .0001$)
Certainty ranges from 0 (complete uncertainty) to 100 (complete certainty). The box and whiskers plot can be interpreted as follows: the whiskers extend to the data minimum and maximum, the bottom of the box is at the 25th percentile,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

NIH-PA Author Manuscript

NIH-PA Author Manuscript

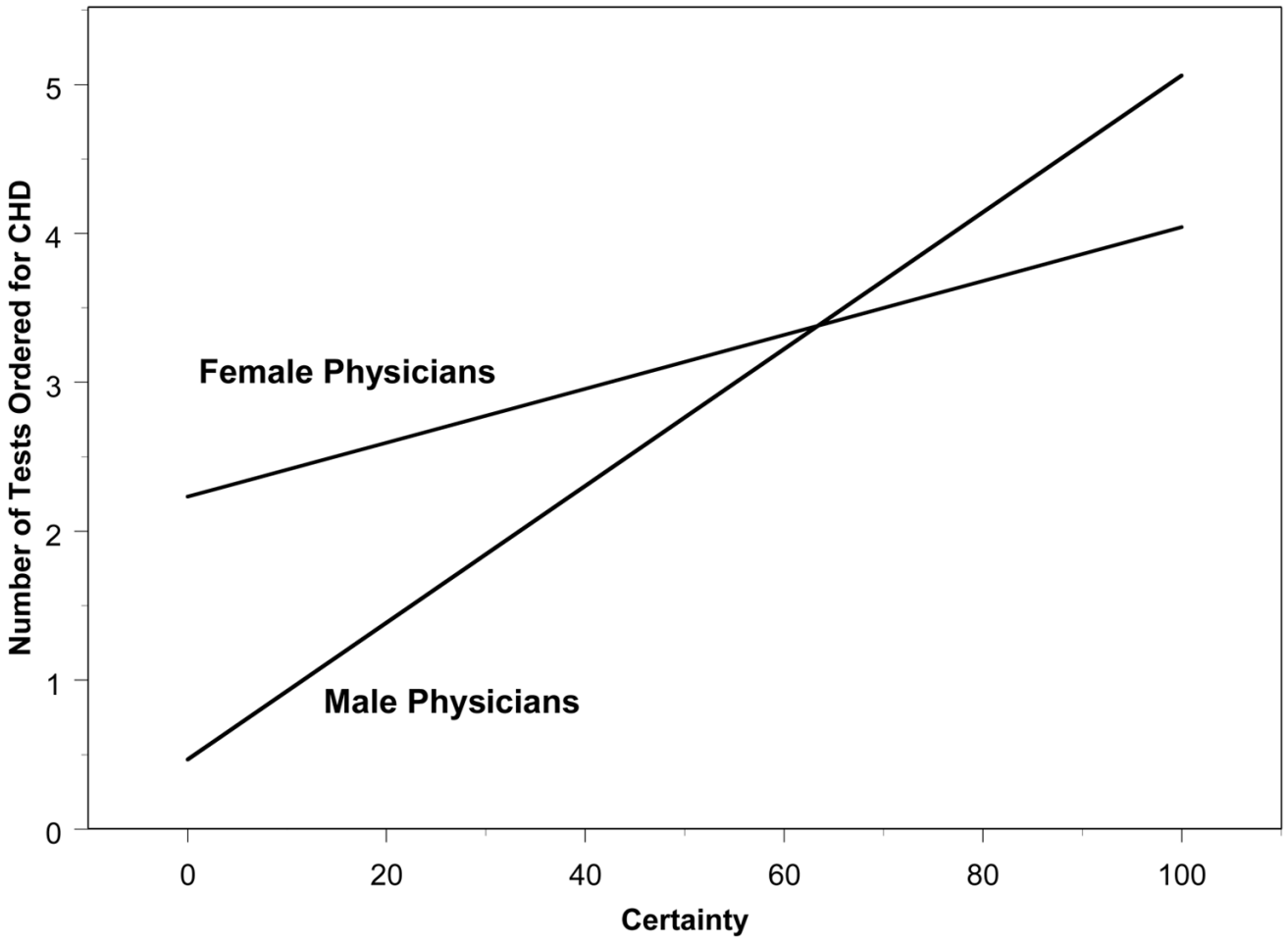
NIH-PA Author Manuscript



NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

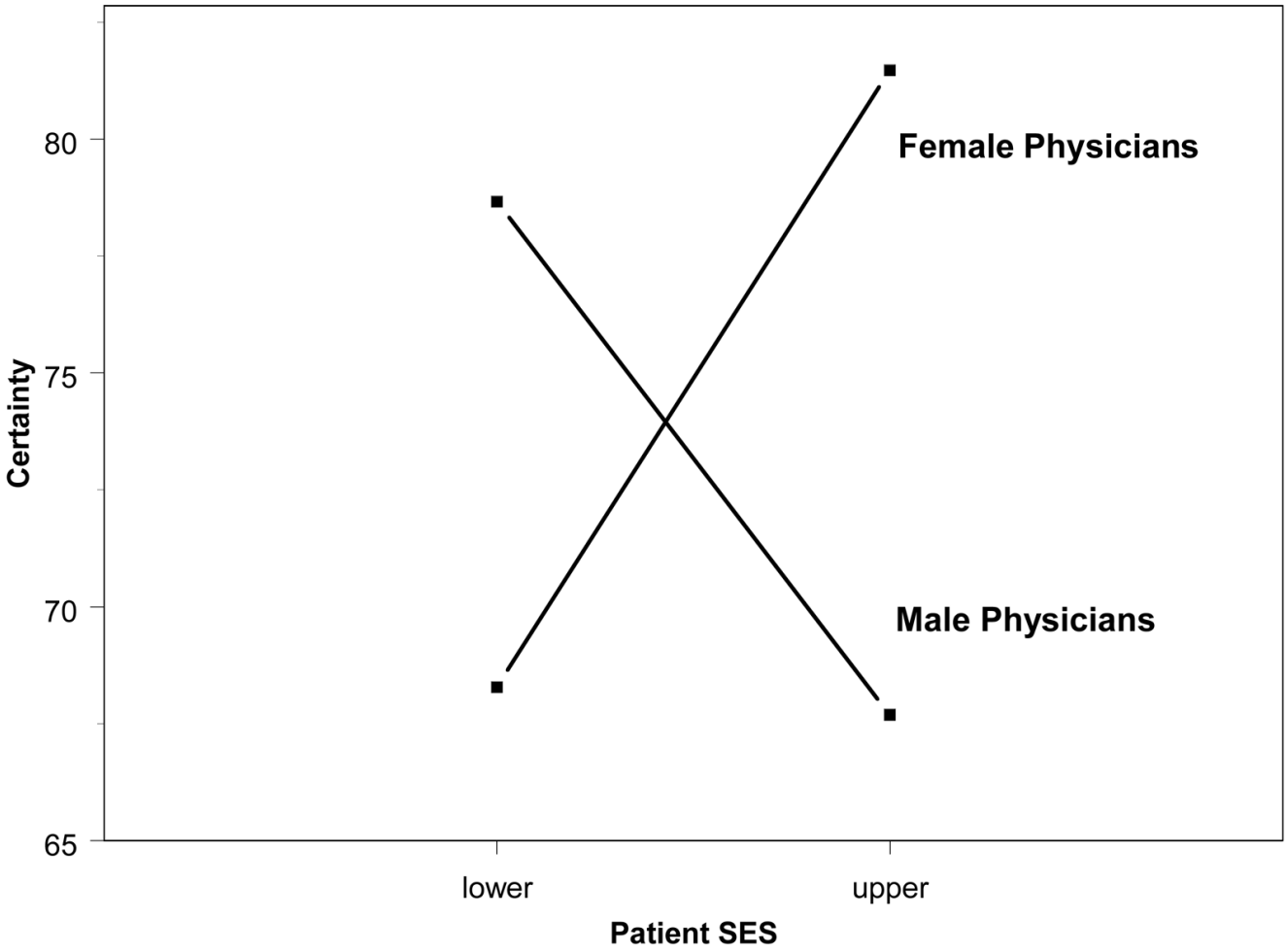


Figure 2.

Figure 2a: Two-way interaction effect of patient gender and patient age on physicians' certainty (0–100) of a CHD diagnosis (p = .013).

Figure 2b: Two-way interaction effect of physician gender and diagnostic certainty on number of CHD tests ordered (p = .0125).

Figure 2c: Two-way interaction effect of patient SES and physician gender on physicians' certainty (0–100) of a depression diagnosis (p = .0103).

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

Analysis of variance results for physicians' certainty of CHD and depression diagnoses (0–100), including main effects and significant ($p < 0.05$) two way interactions.

	CHD		Depression	
	Mean Certainty	**p**	**Mean Certainty**	**p**
Main Effects				
Patient Gender		**.048**		.347
Male	62.3		76.2	
Female	53.5		71.9	
Patient Age		.789		.340
55	57.3		76.2	
75	58.5		71.8	
Patient Race		**.003**		.584
Black	51.2		72.8	
White	64.7		75.3	
Patient SES		.272		.809
Lower	55.5		73.5	
Upper	60.3		74.6	
Physician Gender		.991		.711
Male	57.9		73.2	
Female	58.0		74.9	
Physician Experience		.179		.997
Less	55.0		74.0	
More	60.9		74.0	
Two-Way Interactions				
Patient Gender × Patient Age		**.013**		

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

		CHD		Depression	
		Mean Certainty	**p**	**Mean Certainty**	**p**
Male	55	67.3			
	75	57.3			
Female	55	47.4			
	75	59.7			
Patient SES × Physician Gender					**.010**
Lower	Male			78.7	
	Female			68.3	
Upper	Male			67.9	
	Female			81.5	

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 2

Effect of physicians' certainty of CHD and depression diagnoses (0–100) on subsequent clinical actions (CDM).

2A. Effect of diagnostic certainty on clinical actions (CDM) measured as continuous variables (change per 10 point increase in certainty).

	CHD			Depression		
Variable	**Change per 10 point increase in certainty**	**95% confidence interval**	**p**	**Change per 10 point increase in certainty**	**95% confidence interval**	**p**
Information seeking						
Number of questions	0.00	−0.28, 0.28	.998	0.11	−0.06, 0.28	.535
Physical Examination						
Number of examinations	0.02	−0.15, 0.19	.828	0.12	−0.10, 0.34	.290
Test ordering						
Number of tests for CHD/ Depression	-	-	-	0.08	0.00, 0.16	**.046**
Male physician	0.46	0.31, 0.61	**<.001**	-	-	-
Female physician	0.18	0.03, 0.33	**.023**	-	-	-
Follow-up						
Time to next appt. (days)	−0.77	−1.19, −0.35	**.004**	0.29	−0.33, 0.91	.360
Advice giving						
Number of pieces of lifestyle advice	−0.06	−0.17, 0.06	.274	0.09	0.02, 0.11	**.010**

2B. Effect of diagnostic certainty on clinical actions (CDM) measured as dichotomous variables (Odds Ratio [OR] per 10 point increase in certainty).

	CHD			Depression		
Variable	**OR per10 point increase in certainty**	**95% confidence interval**	**p**	**OR per10 point increase in certainty**	**95% confidence interval**	**p**
Information seeking						
4 or more questions	0.78	0.56, 1.09	.149	0.97	0.81, 1.15	.701
Questions about:						
pathology	0.92	0.80, 1.07	.283	0.90	0.76, 1.07	.240

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

2B. Effect of diagnostic certainty on clinical actions (CDM) measured as dichotomous variables (Odds Ratio [OR] per 10 point increase in certainty).

	CHD			Depression		
Variable	**OR per10 point increase in certainty**	**95% confidence interval**	**p**	**OR per10 point increase in certainty**	**95% confidence interval**	**p**
medical history	1.10	0.92, 1.31	.317	0.93	0.80, 1.08	.370
pain	0.97	0.85, 1.11	.675	0.91	0.80, 1.04	.188
smoking	1.06	0.91, 1.22	.456	0.89	0.66, 1.20	.445
alcohol	0.96	0.84, 1.10	.569	1.04	0.88, 1.22	.675
psychological state	0.91	0.79, 1.04	.146	1.33	1.12, 1.59	**.002**
social questions	0.86	0.74, 0.99	**.035**	-	-	-
55 year old patient	-	-	-	1.18	1.02, 1.38	**.031**
75 year old patient	-	-	-	0.97	0.83, 1.13	.676
general questions	1.05	0.89, 1.24	.553	0.87	0.73, 1.03	.105
Physical Examination						
Complete physical	-	-	-	1.00	0.87, 1.15	.961
less experience physicians	0.83	0.65, 1.07	.155	-	-	-
more experienced physicians	1.10	0.93, 1.29	.274	-	-	-
Test ordering						
Order tests for correct diagnosis	1.75	1.35, 2.27	**<.001**	1.24	1.03, 1.39	**.021**
Stress test	1.39	1.18, 1.63	**<.001**	-	-	-
ECG/EKG	1.75	1.36, 2.26	**<.001**	-	-	-
Prescriptions						
Disease appropriate prescription	1.69	1.39, 2.05	**<.001**	1.72	1.24, 2.38	**.001**
Referrals						
Referral to cardiologist/ psychiatric professional	1.20	0.94, 1.53	.142	2.08	1.23, 3.51	**.006**
Referral to other medical professional	0.92	0.71, 1.20	.548	0.83	0.69, 1.00	**.045**

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

2B. Effect of diagnostic certainty on clinical actions (CDM) measured as dichotomous variables (Odds Ratio [OR] per 10 point increase in certainty).

	CHD			Depression		
Variable	**OR per10 point increase in certainty**	**95% confidence interval**	**p**	**OR per10 point increase in certainty**	**95% confidence interval**	**p**
Advice giving						
Diet	0.98	0.85, 1.12	.717	1.22	0.91, 1.62	.183
Smoking	-	-	-	0.96	0.77, 1.21	.747
Male patients	1.05	0.91, 1.21	.490	-	-	-
Female patients	0.89	0.75, 1.06	.185	-	-	-
Alcohol	0.88	0.74, 1.03	.116	0.96	0.78, 1.19	.732
Relaxation	0.82	0.62, 1.09	.170	1.30	0.82, 2.08	.265
Exercise	0.90	0.70, 1.15	.402	1.16	0.96, 1.41	.115
Weight	0.97	0.64, 1.45	.868	1.87	0.24, 14.74	.548