

Why Do Patients Derogate Physicians Who Use a Computer-Based Diagnostic Support System?

Victoria A. Shaffer, PhD, C. Adam Probst, PhD, Edgar C. Merkle, PhD,
Hal R. Arkes, PhD, Mitchell A. Medow, MD, PhD

Objective. To better understand 1) why patients have a negative perception of the use of computerized clinical decision support systems (CDSSs) and 2) what contributes to the documented heterogeneity in the evaluations of physicians who use a CDSS. **Methods.** Three vignette-based studies examined whether negative perceptions stemmed directly from the use of a computerized decision aid or the need to seek external advice more broadly (experiment 1) and investigated the contributing role of 2 individual difference measures, attitudes toward statistics (ATS; experiment 2) and the Multidimensional Health Locus of Control Scale (MHLC; experiment 3), to these findings. **Results.** A physician described as making an unaided diagnosis was rated significantly more positively on a number of attributes than a physician using a computerized decision

aid but not a physician who sought the advice of an expert colleague (experiment 1). ATS were unrelated to perceptions of decision aid use (experiment 2); however, greater internal locus of control was associated with more positive feelings about unaided care and more negative feelings about care when a decision aid was used (experiment 3). **Conclusion.** Negative perceptions of computerized decision aid use may not be a product of the need to seek external advice more generally but may instead be specific to the use of a nonhuman tool and may be associated with individual differences in locus of control. Together, these 3 studies may be used to guide education efforts for patients. **Key words:** decision support techniques; computer assisted diagnosis; decision aids; patient satisfaction. (*Med Decis Making* 2013;33:108–118)

Computerized clinical decision support systems (CDSSs) are information systems, typically included in electronic medical records (EMRs), that provide clinicians with patient-specific assessments or recommendations to improve clinical decision making.^{1–3} CDSSs have been created for many different clinical issues, including chest pain, infertility,

and preventative medicine.¹ CDSSs provide several types of decision support, including reminder systems for vaccinations, screening tests and other types of preventative care, alerting systems for medication errors (e.g., drug interactions, overdoses, etc.), advice for medication prescription, critiques of existing orders, and recommendations for many health care issues. Although there have not been extensive studies of patient outcomes yet, many CDSSs have been shown to improve the performance of practitioners.¹ Yet despite the effectiveness and prevalence of CDSSs, they are among the least widely used features of EMRs.^{3,4}

There are several reasons that practitioners may be reluctant to embrace CDSSs.^{5–7} For example, physicians may be concerned about how patients will respond to the increasing role of technology in diagnostic medicine, or they may have concerns about how using diagnostic decision support will affect malpractice litigation. There has been some recent research designed to address these two issues. In 2007, Arkes and others⁸ compared judgments of

Received 12 July 2011 from Department of Health Sciences (VAS) and Department of Psychological Sciences, University of Missouri, Columbia, Missouri (VAS, ECM); Office of Patient Safety, Baylor Healthcare System, Dallas, Texas (CAP); Department of Psychology, Ohio State University, Columbus, Ohio (HRA); and College of Medicine, Boston University, Boston, Massachusetts (MAM). Data from this paper were presented at the annual meeting of the Society for Judgment and Decision Making; 16–19 November 2007; Long Beach, California. Revision accepted for publication 11 May 2012.

Address correspondence to Victoria A. Shaffer, PhD, Department of Health Sciences, School of Health Professions, University of Missouri, 504 Clark Hall, Columbia, MO 65221-4290, USA; e-mail: shafferv@health.missouri.edu.

DOI: 10.1177/0272989X12453501

physicians who used a decision aid with judgments of a physician who made an unaided diagnosis for fictitious patients presenting with a possible ankle fracture, a leg injury with risk for deep vein thrombosis, or a persistent cough. Across several studies, the authors demonstrated that a physician who uses a diagnostic aid for a routine medical event was perceived to be less capable than a physician using unaided judgment. When using the diagnostic aid, the physician was perceived to be less professional, to be less thorough, and to have less diagnostic ability. Similarly, Promberger and Baron⁹ reported that participants would be more likely to follow recommendations from physicians than from computers and reported greater trust in the recommendations from physicians than from computers. Furthermore, participants' preferences for their own decisions over the recommendation from a computer program were greater than preferences for their own decisions over the recommendation from a physician.

However, Pezzo and Pezzo¹⁰ painted a slightly more complex picture of patient responses. They reported that using a decision aid attenuated participants' evaluations of physicians. For a physician who was described as using a decision aid, a positive patient outcome was rated less positively and a negative patient outcome was rated less negatively when compared with a physician making an unaided decision. Furthermore, the authors reported that participants assigned less fault to physicians who followed the advice of an aid after a negative outcome. In contrast, greater fault was ascribed to physicians who defied the aid or made an unaided decision. In a related study, Arkes and others¹¹ reported that use of a simple decision aid by a physician did not affect judgments of malpractice in a mock jury trial, but aid use did affect mock jurors' ratings of punitiveness. The physician was deemed to be *least* deserving of punishment when he or she used the decision aid and heeded its advice but *most* deserving of punishment if he or she used the decision aid and defied its advice. Ratings of punitiveness when the physician used unaided judgment fell between these two conditions.

Thus, although diagnostic aid use does not affect judgments of malpractice, they appear to affect patients' perceptions of their physicians and evaluations of physicians' responsibility for both positive and negative patient outcomes. These findings could be quite significant; a patient's attitude toward his or her physician could affect larger health care issues such as compliance with preventative care and treatment recommendations. Therefore, the purpose of this article is to better understand why patients

have a negative perception of diagnostic aid use. To do so, we will present 3 vignette-based studies that are designed to explore this finding in more depth and examine the role of individual differences in patient response to practitioner decision aid use.

EXPERIMENT 1

Although participants in past studies consistently reported that they perceived physicians using diagnostic aids to be less capable, it is unclear where their concern originates. That is, patients may be concerned that their physician needs to seek advice from an external source, regardless of whether that source is electronic or human. Or, the concern may stem from the incorporation of an electronic source into a traditionally human-centered practice, in which case, computer-based electronic aids may be seen as dehumanizing the patient-physician relationship. Thus, experiment 1 was designed to identify whether the source of the concern was simply the practitioner's need to rely on an external source or the introduction of computing into the doctor-patient interaction. If patients are concerned about their physician seeking additional advice during the diagnostic process, physicians using their unaided judgment should receive the highest evaluation. Moreover, evaluations of physicians seeking external advice, either from an expert or an aid, should not differ. In contrast, if patients are reacting to the introduction of a computer into the interactions with their physician, then a) physicians using diagnostic aids should be evaluated the *least* positively, and b) evaluations of physicians using unaided judgment should not differ from those of physicians who seek the advice of a human expert.

Method

Participants. Four hundred thirty-four students from 2 Midwestern universities participated in experiment 1 for extra credit or partial fulfillment of course credit. Participants were randomly assigned into 1 of 3 experimental conditions: control ($n = 145$), aid ($n = 148$), or expert ($n = 141$). Data were collected in the fall semester of 2006 and the spring semester of 2007.

Materials and procedure. Participants read 1 of 3 short vignettes that described an interaction between a physician and a patient who had a recent ankle injury; these materials were adapted from Arkes and others⁸ (see Appendix A). The vignettes differed in their description of the physician's

diagnostic process. The physician in the vignette made an unaided diagnosis (control condition), used a computer-based diagnostic aid (aid condition), or used the advice of an expert (expert condition). After reading the vignette, participants completed a brief questionnaire that asked them to evaluate the behavior of the physician described in the vignette on 5 dimensions: diagnostic ability, professionalism, overall satisfaction, thoroughness, and wait time at the physician's office. Participants rated each dimension on a 7-point Likert scale, where 1 was the most negative evaluation and 7 was the most positive evaluation.

Analyses. To test the 2 competing hypotheses, we compared ratings of the physician in the 3 conditions. Because the variances in these ratings differed significantly between conditions, we employed beta regression models (e.g., Smithson and Verkuilen¹²), which allowed us to separately model the mean and variance. These models employ the beta distribution to describe error, as opposed to a normal error distribution in traditional regression models. Along with allowing us to separately examine experimental effects on the mean and variance of our dependent measures, beta regression models explicitly account for the bounded nature of the Likert measures. This allows us to interpret effects on the variance as "polarization" effects.¹³ That is, larger variances on a Likert scale imply that individuals tend to be clustered at the extremes. In contrast, smaller variances imply that individuals tend to be clustered together at a single point on the scale.

For the beta regressions described below, we used the *betareg* package¹⁴ in R statistical software.¹⁵ We distinguish between regression weights for the mean and regression weights for the precision using different accented letters ($\hat{\alpha}$ for the mean, $\tilde{\alpha}$ for the precision). It is important to note that the $\hat{\alpha}$ s are estimated on the log-odds scale (similar to a logistic regression) and the $\tilde{\alpha}$ s are estimated on the log scale, so they are complicated to interpret at a glance. Thus, we supplement the regression weights with data and model predictions.

Results

We first studied the impact of the experimental conditions on mean physician ratings. We found that, compared with the aid group, the control group rated the physician more positively on overall patient satisfaction ($\hat{\alpha} = 0.31$, $z = 2.36$, $P < 0.05$) and professionalism ($\hat{\alpha} = 0.36$, $z = 2.70$, $P < 0.05$). On the 7-point

Likert scale, these results correspond to predicted increases of 0.46 and of 0.45 for the control group, respectively. There were no mean differences between the control group and expert group, with the corresponding z statistics ranging from -0.83 to 0.31 .

Next, we studied the impact of the experimental conditions on the variance in physician ratings. As mentioned previously, these results can inform us about the extent to which participants were polarized. We found that the aid group was more polarized than the control group on both overall patient satisfaction ($\tilde{\alpha} = 0.45$, $z = 3.23$, $P < 0.05$) and thoroughness ($\tilde{\alpha} = 0.48$, $z = 3.22$, $P < 0.05$). The effect on overall patient satisfaction was largely driven by the fact that, compared with the control condition, more than twice as many people in the aid condition rated the doctor a 1 or 2 (8% in the control group v. 18% in the aid group). Conversely, the effect on thoroughness was driven by polarization on both ends of the scale: 3% (aid) v. 0% (control) of participants rated the doctor a 1, and 16% (aid) v. 12% (control) rated the doctor a 7.

We also examined the impact of the expert condition on variance in ratings, finding that the expert group was also more polarized than the control group on thoroughness ($\tilde{\alpha} = 0.49$, $z = 3.15$, $P < 0.05$). This polarization is largely on the high end of the scale: 2% (expert) v. 0% (control) of participants rated the doctor a 1, and 23% (expert) v. 12% (control) rated the doctor a 7.

Discussion

Previous research demonstrated that physicians using diagnostic aids were perceived to be less capable than physicians making unaided diagnoses.⁸ Experiment 1 was designed to extend these findings by determining their source. Are physicians using diagnostic aids perceived to be less capable because they need to acquire additional information to make their diagnoses? Or does this finding simply reflect a bias against the introduction of technology into a traditionally human-oriented exchange?

In experiment 1, we replicated the findings of Arkes and colleagues.⁸ The physician who made an unaided diagnosis was perceived to be more professional, and participants in the control condition were more satisfied with the behavior of the physician. In addition, responses of the aid condition were more polarized than the control condition on overall satisfaction and professionalism. To address the source of this bias, we compared the ratings of

the control condition with those of the human expert condition. Ratings of the physician on the 5 dimensions did not differ between the control and expert conditions. However, responses of the expert group were more polarized than those of the control group on thoroughness of examination. Together, these results could indicate that physicians using a diagnostic aid are not perceived to be less capable simply because they need additional information to make a diagnosis. Instead, patients may simply object to the use of an impersonal technology. However, there are other alternative explanations for this finding. For example, patients might view physicians who need to consult a CDSS as “dumb” but consider a physician who consulted with a specialist as “smart.” Future research should focus on distinguishing between these alternative explanations.

It is also important to note that the results of experiment 1 are in contrast to physician behavior. A recent study by Medow and others¹⁶ demonstrated that physicians were more likely to change treatment course and to reduce confidence in their decision in response to hearing contradictory advice from a decision aid than from a colleague.

EXPERIMENT 2

In addition to understanding why physicians using diagnostic aids are deemed to be less capable, it is also important to determine for whom this perception exists. In the original experiments by Arkes and colleagues,⁸ the control and aid conditions differed in the amount of variability observed in physician ratings. There was little variance in the ratings across participants in the control group; however, there was significantly more variability in ratings among the aid group. Although the introduction of the computer-based diagnostic aid lowered ratings of diagnostic ability on average, there were also many participants who gave higher ratings when the physician used the aid. Thus, there appears to be a significant amount of individual variability in response to diagnostic aid use. With this in mind, the goal of experiment 2 and experiment 3 is to identify measures of individual difference that predict this variability. If such measures can be identified, it is possible to use this information to tailor how diagnostic aids are presented in practice.

Diagnostic aids are often based on statistical models that calculate the likelihood of a particular diagnosis for a given patient. Therefore, a large part of the resistance to these tools may stem from personal

beliefs about the role of statistics in individual patient care. Sieck and Arkes¹⁷ reported that decision aid use in a prediction task was related to an individual's attitude toward statistics. Participants in that study were asked to use 5 cues to predict whether a prospective juror favored physician-assisted suicide. They had the opportunity to evaluate the predictions of an equation that correctly classified 77% of the jurors, or they could attempt to classify the jurors without the use of the decision aid. The authors found that subjects who held a more favorable attitude toward statistics were more likely to use the decision aid provided.

Experiment 2 sought to extend this finding by investigating whether more favorable attitudes toward statistics were associated with more positive evaluations of physicians using diagnostic aids. Although the CDSS used in our scenario was not explicitly described as a “prediction rule,” as in Sieck and Arkes,¹⁷ we hypothesized that the use of a computerized aid to judgment would be looked upon more favorably by people with more positive attitudes toward statistics. To do so, we measured attitudes toward statistics of students in enrolled in psychology statistics courses. In addition, we asked the students to read a vignette describing an interaction between a doctor and a patient and evaluate the physician on several dimensions. The students viewed 1 of 2 versions of the vignette; the physician involved either used a computer-based diagnostic aid or made an unaided diagnosis.

We hypothesized that attitudes toward statistics would be positively related to physician ratings in the aid condition because we postulated that a greater appreciation for statistics would result in a greater appreciation for tools that use statistics, such as the prediction rule used by Sieck and Arkes¹⁷ and the CDSS used in this study. We had no specific hypotheses about how attitudes toward statistics would influence ratings of physicians in the control condition.

Method

Participants. One hundred nine students from a Midwestern university who were enrolled in a psychology statistics course participated in the experiment for extra credit. Individual classes were randomly assigned into 1 of 2 experimental groups, the control condition ($n = 60$) or the aid condition ($n = 49$). All members of a class were assigned to the same condition. Data were collected across 2 semesters in the fall of 2007 and the spring of 2008.

Materials. Two vignettes described an interaction between a physician and a patient who had a recent ankle injury. The physician in the vignette either made an unaided diagnosis (control condition) or used a computer-based diagnostic aid (aid condition). The vignettes were identical to the control and aid versions used in experiment 1. After reading the vignette, participants completed a brief questionnaire that asked them to evaluate the behavior of the physician described in the vignette on 5 dimensions: diagnostic ability, professionalism, overall satisfaction, thoroughness, and wait time at the physician's office. Participants rated each dimension on a 7-point Likert scale. Participants were also asked to complete the attitude toward statistics (ATS) measure,¹⁸ which includes 29 statements about statistics (e.g., "Statistics is an inseparable aspect of scientific research") to which participants must indicate the extent to which they agree, using 5-point Likert scales.

Procedure. Instructors of the statistics courses made in-class announcements about the study during the first 2 weeks of the semester. Interested students could access the study directly from the website for their statistics course. After accessing the study link, all participants were asked to read a brief description of a doctor-patient interaction. Half of the statistics courses read the control version, which described the physician as making an unaided diagnosis, whereas the other half of the courses read the aid version, in which the physician used a computer-based aid to make the diagnosis. After reading the vignette, participants rated the physician on the 5 dimensions and completed the ATS scale.¹⁸ Participants were asked to complete this entire procedure at 2 time points in the semester, during the first 2 weeks of the semester and the final week of the semester. Assignment to condition was the same for both data collection points.

Analyses. As in experiment 1, variances of some of the dependent measures differed significantly between the 2 groups. To deal with the heterogeneous variances, we again employed beta regression models. To provide the cleanest test of the influence of ATS, we focused our analyses on time 1 measurements, which occurred during the first 2 weeks of the semester. In each regression, the following variables were used as predictors for the (logit-transformed) mean: condition, ATS score (standardized by the mean and standard deviation across conditions), and the 2-way interaction. ATS score was standardized to ease the interpretation of other model parameters.

Based on initial analyses, "condition" was used as the only predictor variable for the (log-transformed) precision/variance parameter.

Results

ATS scores in our participants ranged from 56 to 134 (mean [SD], 97.66 [16.45]). To test the hypothesis that attitudes toward statistics would be related to ratings of the physicians, we studied the impact of condition, ATS, and the ATS-by-condition interaction on the mean physician rating. For the thoroughness of examination measure, there was a negative effect of the aid condition ($\hat{a} = -0.46$, $z = 2.03$, $P < 0.05$) and a positive effect of ATS score ($\hat{a} = 0.21$, $z = 1.66$, $P < 0.05$) (see Figure 1). These effects translate to the 7-point Likert scale as follows. A participant in the aid condition who has a mean ATS score is predicted to lower his or her thoroughness rating by approximately half a point (0.54), as compared with an analogous participant in the control condition. Furthermore, for each standard deviation increase in a control group participant's ATS score, the participant's thoroughness rating is predicted to increase by about 0.2. (This increase varies because the prediction increases nonlinearly.)

For the diagnostic ability measure, there was a negative effect of the aid condition ($\hat{a} = -0.45$, $z = -2.27$, $P < 0.05$). For a participant with a mean ATS score, this effect represents a 0.64-point decrease on the 7-point scale. Finally, for the satisfaction measure, there was a significant, positive effect of ATS score ($\hat{a} = 0.27$, $z = 2.28$, $P < 0.05$) (see Figure 2). For a participant in the control group, each standard deviation increase in ATS score was predicted to result in a 0.4-point satisfaction increase. Effects (simple or interaction) of aid condition and ATS on other dependent measures were smaller, with z statistics ranging from -1.34 to 0.45 .

Finally, as noted above, we also studied the impact of condition on judgment polarization. We found significant effects of experimental condition on ratings of thoroughness ($\hat{a} = 0.63$, $z = 2.52$, $P < 0.05$), diagnostic ability ($\hat{a} = 1.09$, $z = 4.60$, $P < 0.05$), and overall satisfaction ($\hat{a} = 0.75$, $z = 3.20$, $P < 0.05$). In all 3 cases, individuals were more polarized (exhibited more variability) in the aid condition than in the control condition. For all 3 measures, the polarization largely involved the lower end of the scale: For thoroughness, 14% of aid group participants assigned a rating of 2 or less, compared with 3% of control group participants. The analogous proportions for the diagnostic ability and overall satisfaction measures are (14%

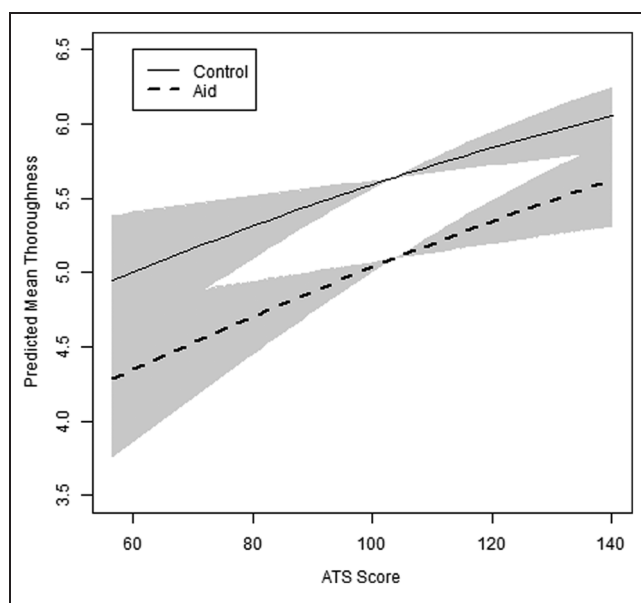


Figure 1 Model predictions of the relationship between condition and attitudes toward statistics (ATS) on ratings of physician thoroughness (experiment 2). The shaded areas represent variability in the slope associated with ATS, obtained by adding and subtracting 1 standard error from the point estimate.

aid; 5% control) and (29% aid; 10% control), respectively.

Discussion

In experiment 2, we replicated the findings of Arkes and colleagues.⁸ The physician who made an unaided diagnosis was perceived to be more thorough and to have greater diagnostic ability than the physician using the diagnostic aid. Moreover, judgments of thoroughness, diagnostic ability, and overall satisfaction were significantly more polarized in the aid condition than in the control condition; respondents in the aid condition were more likely to use the extreme lower end of the Likert rating scales. ATS score was also positively related to ratings of thoroughness and diagnostic ability, but contrary to our hypotheses, there were no significant condition-by-ATS interactions. Although Sieck and Arkes¹⁷ reported that attitudes toward statistics predicted decision aid use, we were unable to find evidence that ATS score was similarly related to participant ratings of physicians using decision aids. One potential limitation of this study is that by using psychology students, we may have restricted our range of ATS scores, thereby underestimating the influence

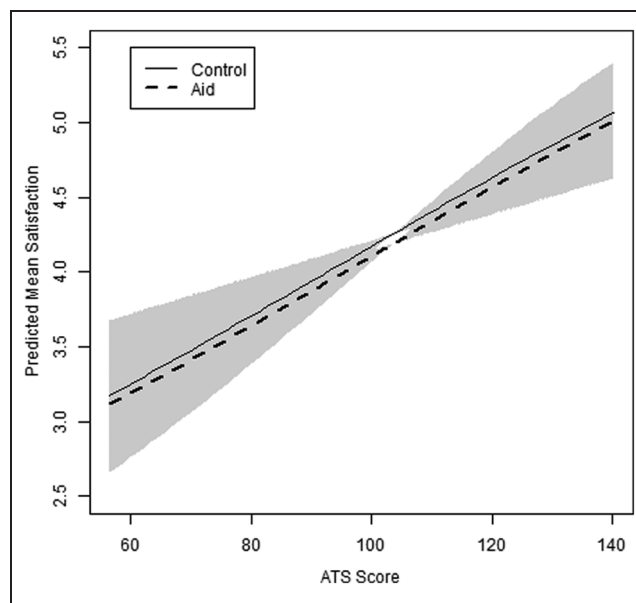


Figure 2 Model predictions of the relationship between condition and attitudes toward statistics (ATS) on ratings of satisfaction with the physician (experiment 2). The shaded areas represent variability in the slope associated with ATS, obtained by adding and subtracting 1 standard error from the point estimate.

of attitudes toward statistics on acceptance of CDSS use.

EXPERIMENT 3

In experiment 3, we examined the predictive power of another individual difference measure, locus of control. Locus of control is a construct that broadly measures the extent to which people believe they have control over events in their lives. People with an internal locus of control believe they have direct control over the events in their lives, whereas those with an external locus of control believe that external forces, such as chance or fate, control the events in their lives. We chose to focus on locus of control because it has been shown to be predictor of other health behaviors.^{19,20} In addition, we hypothesized that locus of control would be related to patients' satisfaction of care with physicians who use decision aids.

Two measures are used to assess locus of control in health care: Rotter's original Locus of Control (LOC) Scale²¹ and the Multidimensional Health Locus of Control (MHLC) Scale.²⁰ Rotter's LOC Scale is conceptualized as a single bipolar dimension where individuals are categorized along a continuum ranging

from an internal locus of control to an external locus of control. However, the MHLC Scale is based on 3 unipolar dimensions: Powerful Others (the belief that one's health outcomes are in the hands of powerful individuals such as physicians), Internality (the belief that one's health outcomes are within one's own control), and Chance (the belief that one's health outcomes are subject to random events). In this study, we chose to focus on the MHLC because of its specific relevance to the health context and the addition of the Powerful Others subscale.

We hypothesized that patients with a more internal orientation would feel less favorably toward clinicians who use computer-based decision aids. Thus, we predicted an interaction between the Internal dimension and condition on physician ratings, with a negative relationship between Internal locus of control and satisfaction with care only in the aid condition. We also hypothesized that the Powerful Others dimension would be positively related to ratings of satisfaction with care in both conditions; that is, participants with a strong belief in the role of powerful others would be more likely to report greater satisfaction with care regardless of practitioner aid use.

Method

Participants. One hundred eighty-nine students from a Midwestern university participated in experiment 3 for extra credit or partial fulfillment of course credit. Participants were randomly assigned to 1 of 2 conditions: control ($n = 101$) or aid ($n = 88$). Data were collected in the spring and fall semesters of 2009.

Materials and procedure. Participants read 1 of 2 short vignettes that described the experience of a patient who presented at the emergency room with possible appendicitis (see Appendix B). The 2 vignettes differed in their description of the physician's diagnostic process; the physician either used a diagnostic aid or made an unaided diagnosis (control condition). After reading the vignette, participants completed a brief questionnaire that asked them to evaluate the behavior of the physician described in the vignette on 8 dimensions: overall satisfaction, professionalism, quality of examination, knowledge of the physician, likelihood to use the physician themselves, willingness to visit the physician again, willingness to recommend the physician, and confidence in the physician. Participants rated each dimension on a 6-point Likert scale, with the exception of quality of examination, which was

measured on a 7-point Likert scale. Participants also completed the MHLC Scale.²⁰

Analyses. As in experiment 1 and 2, we employed beta regression models for data analysis. All 8 dependent measures (ratings of the physician) were strongly correlated; r s ranged from $+.62$ to $+.90$, all P s were <0.05 . In addition, the pattern of results was consistent across all 8 variables. Therefore, we created an aggregate physician rating by summing these 8 scores, with higher values indicating a more favorable opinion of the physician.

To test the hypothesis that locus of control would be related to ratings of the physicians, we estimated a beta regression model using the 3 locus of control dimensions (standardized), experimental condition, and 2-way interactions of the predictor variables on the mean parameter. Experimental condition was used as the sole predictor variable on the precision parameter, and the aggregate physician rating was the response variable. Boundaries were defined as the minimum possible score of 8 and the maximum possible score of 49.

Results

In our sample, scores on the Internal subscale ranged from 8 to 36 (mean [SD], 26.86 [4.14]). Scores on the Chance subscale ranged from 6 to 31 (18.86 [4.53]), and scores on the Powerful Others subscale ranged from 7 to 31 (17.43 [5.12]). Consistent with our hypotheses, aggregate physician ratings were negatively related to the Internal subscale in the aid condition ($\hat{a} = -0.66$, $z = -4.63$, $P < 0.05$). In contrast, aggregate physician ratings were positively related to the Internal subscale in the control condition ($\hat{a} = 0.44$, $z = 4.31$, $P < 0.05$) (see Figure 3). Translating these results to the aggregate physician ratings, a participant with a mean Internal subscale score is predicted to assign the physician a rating of 27.9 in the control condition and a rating of 28.1 in the aid condition. For a participant with an Internal subscale score that is 1 standard deviation above the mean, however, the predicted rating is 32.4 in the control condition and 25.9 in the aid condition. These predictions become more discrepant as the Internal subscale score increases past 1 standard deviation above the mean.

Neither the Chance nor Powerful Others subscale was significantly related to the aggregate physician ratings in either condition. We also examined the impact of condition on rating polarization. For the aggregate ratings reported in this experiment,

individuals were more polarized in the control condition than in the aid condition ($\tilde{a} = 0.89$, $z = 4.81$, $P < 0.05$). Ratings in the control condition tended to be closer to the scale boundaries than did ratings in the aid condition.

Discussion

In experiment 3, we examined the relationship between locus of control and attitudes toward decision aid use in emergency medicine. In the control condition, higher scores on the Internal subscale were positively related to ratings of satisfaction with care. In contrast, higher scores on the Internal subscale were negatively related to ratings of satisfaction with care in the aid condition. This suggests that compared with less internally oriented people, those with a stronger internal locus of control may feel more positively about the care they receive from clinicians who do not use diagnostic aids but more negatively about care when decision aids are used. However, contrary to predictions, the Powerful Others dimension was not related to ratings of satisfaction with care in either condition.

GENERAL DISCUSSION

Previous research by Arkes and colleagues⁸ demonstrated that physicians using computer-based diagnostic aids in primary care were perceived to be less professional, to be less thorough in examinations, and to have less diagnostic ability. This article was designed to extend these findings in 2 ways. First, we hoped to identify whether this derogation of the physician was simply due to the need to seek advice from an outside source (colleague or computer) or if the distrust was specific to the use of a computerized decision aid (experiment 1). Second, we sought to identify important individual differences that might help to predict patients who would object to the use of a computer-based diagnostic aid, examining the role of attitudes toward statistics (experiment 2) and health locus of control (experiment 3).

In experiment 1, the physician who was described as using a computer-based diagnostic aid in a brief vignette was given significantly lower ratings of diagnostic ability, professionalism, and overall satisfaction than the physician who made an unaided diagnosis. In contrast, ratings of the physician who sought advice from a colleague were not significantly different from ratings of the physician who made the unaided diagnosis. Although people may not expect their physicians to be

knowledgeable about all topics, they may still have concerns about the use of computer-based aids. One possible explanation is that patients may highly value the human component of the patient-clinician interaction and may have concerns about how CDSSs will affect this relationship. Alternatively, patients may simply view physicians who use CDSSs as less intelligent, whereas they may consider consulting a specialist as a wise choice. Or this finding may simply reflect a general distrust of computing systems to replace clinical judgment in important diagnostic tasks. The latter interpretation is consistent with previous research by Promberger and Baron.⁹ In any case, this finding is significant as it may help direct attempts to educate patients about the benefits of computerized diagnostic tools in health care.

The second aim of this research was to identify measures of individual difference that could help to predict patients who would respond more negatively to the use of computer-based decision aids. Experiment 2 examined the role of attitudes toward statistics. Because diagnostic aids are often based on statistical models that calculate the likelihood of a particular diagnosis for a given patient, a large part of the resistance to these tools may stem from personal beliefs about the role of statistics in individual patient care. We replicated the original findings of Arkes and colleagues⁸ and found that several of the dependent measures were significantly more polarized in the aid condition than in the control condition. ATS score was also positively related to ratings of thoroughness and diagnostic ability, but contrary to our hypotheses, there were no significant condition-by-ATS interactions. Therefore, even though attitudes toward statistics have been shown to be an important predictor of decision aid use in another domain,¹⁷ it did not help explain attitudes toward physicians using decision aids in this study.

Experiment 3 examined the role of locus of control using the MHLC Scale, which includes the Internal, Chance, and Powerful Others subscales. Consistent with our hypotheses, people with higher scores on the Internal subscale gave more positive evaluations of the physician and the care received when the physician made an unaided diagnosis but gave lower evaluations to physicians who used a computerized diagnostic aid. This study is significant because it helps to explain the heterogeneity in reactions to decision aid use. This work could be used to identify patients who may be more receptive to decision aid use in health care (those with a more external orientation) and those who may be less receptive to the use of diagnostic aids (those with a more internal orientation).

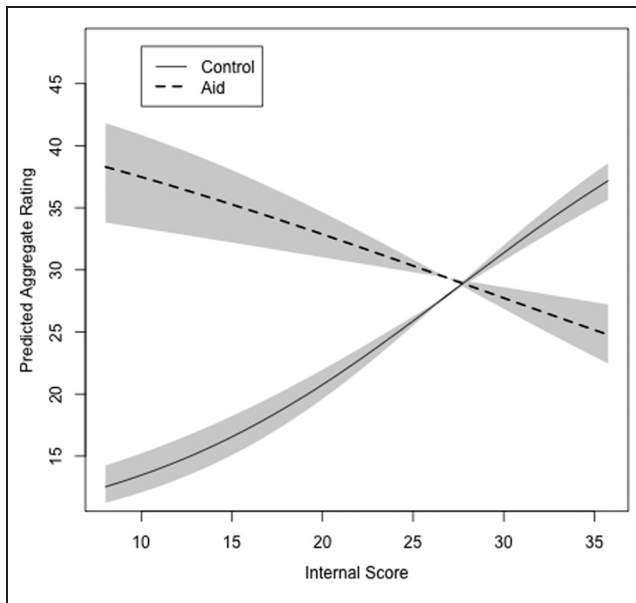


Figure 3 Model predictions of the relationship between condition and internal health locus of control on aggregate physician rating (experiment 3). The shaded areas represent variability in the slopes associated with Internal scores, obtained by adding and subtracting 1 standard error from the point estimates.

There are limitations to these studies. First, these studies involve hypothetical scenarios. Reactions to actual events may produce different or stronger reactions, which may depend on the relationship to the patient (e.g., significant other, family member, acquaintance). Second, our participants were all undergraduate students. As such, they are unlikely to have extensive interactions with physicians or much experience making serious medical decisions. In addition, they are more likely to have experience with technology than older adults, which could potentially attenuate negative attitudes toward decision aid use. Third, experiment 2 used only students from a statistics course. This may have restricted the range of ATS scores, thus obscuring any possible relation between this factor and the dependent variables. Sieck and Arkes,¹⁷ who did find a relation between ATS scores and decision aid use, sampled a more general group of participants for their study.

Future research designed to further understand patients' reluctance to accept the use of decision support should seek to establish the boundary conditions of this finding. For example, researchers could examine how variations in the clinical context or properties of the CDSS (e.g., accuracy, transparency, complexity) affect attitudes toward aid use. Future work should also

include samples with greater variability in age and medical history. Given the effectiveness of CDSSs, it is important to improve patient attitudes toward their use in the hopes that the decision support features of EMRs would be used more frequently. To achieve this goal, clinicians and researchers could partner to develop and test patient education interventions designed to alleviate patients' negative attitudes toward CDSS use.

The 2 broad aims of this article provide a better understanding about what motivates the negative reaction to physicians who use diagnostic aids. Specifically, this effect does not appear to be driven by the general need to seek advice, and this effect is moderated by individual differences (e.g., health locus of control). With the recent US initiatives to promote EMRs, there will likely be a steadily increasing number of computerized clinical decision support systems available to practitioners. Furthermore, CDSS use is an important part of meeting the meaningful use criteria that enable US physician practices and hospitals to become eligible for federal reimbursement when investing in computerized systems. Thus, whether patients are comfortable or not with the use of these computer-based decision aids, they will be increasingly used in health care, which could increase patient dissatisfaction. These 3 studies may help to guide education efforts for patients not ready to make this transition. In addition, clinicians may use this information to tailor how information is presented to individual patients.

APPENDIX A

On Saturday afternoon, during an informal game of softball at the local park, you hurt your left ankle. You jumped up to catch a line drive, and when you landed, your ankle turned in. You fell to the ground and were unable to get up or walk because of the pain. Your teammates, one of whom is a physical therapist, helped you to the side and got some ice for your ankle from the concession stand. Your friend gave you a ride home and helped you to your couch. She recommended that you see your physician tomorrow. Until then, you kept your ankle elevated, used ice, and took some ibuprofen for the pain.

Early the next morning, at 7:30 AM, you called your physician's office. The recorded message said they do not take calls for appointments for another hour and a half. You called back promptly 90 minutes later and got an appointment later that afternoon. You managed to go to work, using an ornamental walking stick you brought back from a trip.

In the afternoon, you left work early to get to your appointment. After a 30-minute wait, a nurse takes you to an examination room. The nurse asks you what the problem is, and you respond that you injured your left ankle yesterday playing softball. The nurse takes your temperature (98.7 degrees), measures your blood pressure (122/78), takes your pulse (78 beats per minute), measures your respiratory rate (16 breaths per minute), and asks you to step on the scale. The nurse writes this information in the chart and then leaves. You want to ask her if you should get an X-ray to help move things along, but she left before you had a chance.

About 15 minutes later, the doctor comes into the room and asks you a number of questions:

1. When did this happen?
(Last evening.)
2. What were you doing?
(Playing softball.)
3. Describe the accident:
(When I landed, my ankle turned in. I could not walk on it. My friends helped me to the side and got some ice for it. The ankle swelled up and there seems to be a bruise over the outer part of it.)
4. Can you walk on it now?
(No, I cannot put my full weight on it. I must use this walking stick.)
5. Have you ever injured your ankle before?
(Yes, but only minor twists.)

In addition, the doctor asks you some more questions:

- Do you have any drug allergies? (*No.*)
Do you have any other major health problems? (*No.*)

The doctor uses her stethoscope to listen to your lungs as you breathe deeply and to your heart while you lie back quietly. She then examines your ankles: Your left ankle is puffy. There is a visible black and blue mark on the outside, over the bone. When she pushes over the bone on the outside of the ankle, it is tender. She moves it side to side and back and forth. It is beginning to throb. She asks you to walk on it without the walking stick. You are unable to. The doctor asks you to sit down again.

Control Condition

Your doctor explains to you that she is concerned that you might have fractured your ankle during the injury. So she orders an X-ray of your ankle.

Aid Condition

The doctor then indicates she is going to use a decision aid [computer program] for ankle injuries to

decide whether to order an X-ray of your ankle. When she returns, she says that, according to this decision aid [computer program], you should have an X-ray of your ankle to see if it is fractured. So she orders an X-ray of your ankle.

Expert Condition

The doctor then indicates she is going to speak with a specialist in ankle injuries to decide whether to order an X-ray of your ankle. When she returns, she says that, according to the specialist, you should have an X-ray of your ankle to see if it is fractured. So she orders an X-ray of your ankle.

APPENDIX B

You are out on a Friday night with your friends. Your group begins the evening by going to a fast-food restaurant for dinner. After finishing dinner, your friends comment that there is a new movie they have wanted to go see and you agree to go along.

When buying your ticket for the movie, you suddenly experience abdominal pain in the lower right portion of your abdomen. The pain feels as though you have indigestion or are in need of passing gas. You attribute the pain to the dinner you ate and continue on into the theater to find your seat. You observe the pain gets considerably worse when you move or cough, and you begin to feel nauseous. When you feel your forehead, you notice you are running a slight fever.

After the movie is over, you tell your friends you are going to turn in early and explain that you are having abdominal pain and are feeling nauseous. One of your friends tells you that her brother had similar symptoms and had to see a doctor. You agree to allow your friends to take you by the emergency room to make sure there is nothing seriously wrong.

You enter the emergency room and inform the nurse at the desk that you are having abdominal pain and nausea. She asks you to fill out some standard paperwork and to take a seat to wait for your name to be called. After about 20 minutes, you are called into an examination room. After another 15 minutes, the emergency room physician, Dr. Crawford, enters and asks you to explain to him what symptoms you are experiencing.

Dr. Crawford performs a physical examination on you and orders several lab tests. He takes your temperature and tells you that you have a slight fever of 101.5 F, your pulse rate is 106, and your blood pressure is 132/76. He then asks you to lie down and

begins to feel your abdomen. You inform Dr. Crawford that you feel slightly better when you lie down. He orders a blood test to be drawn and a urine test to be performed as well as an X-ray of your abdomen. He explains he will return when the results come back and leaves the room to order the tests.

When he returns 45 minutes later, he informs you that your white blood count is slightly elevated to a count of 10,500 and that your urine test appears normal. Your X-ray is clear. He then performs a CT scan of your abdomen, which is inconclusive.

The physician then explains he initially thought you had appendicitis but that you are not exhibiting all the symptoms that usually accompany appendicitis. For example, your neutrophil count is not more than 75% and your pain did not begin in the center of your abdomen and radiate to the right. In your case, the pain began, and lingered, in the lower right quadrant. Dr. Crawford also mentions that he observed no "rebound pain" when examining you (pain that occurs when an abdomen has pressure applied to it and then released).

Control Condition

Dr. Crawford explains that he was initially unsure whether to arrange for a surgeon to perform an appendectomy because he was not certain concerning the diagnosis of appendicitis. However, he subsequently decides to call for a surgeon, the appendectomy is quickly scheduled, and the surgeon begins the operation within the hour.

Aid Condition

Dr. Crawford explains he is going to type your symptoms into a computer-based diagnostic aid. He explains that this aid will calculate your "Alvarado" score and will then recommend a course of treatment. He explains that the Alvarado score is a clinical coding system used to calculate appendicitis. He takes a few moments to enter in your score and then reads the output the device gives him. He tells you your score is 7, which is "probable appendicitis." Dr. Crawford informs you that he thought that there was some chance that you did not have appendicitis, but he has decided to follow the recommendation of the Alvarado score. Therefore, he calls for a surgeon, the appendectomy is quickly scheduled, and the surgeon begins the operation within the hour.

REFERENCES

- Garg AX, Adhikari NKJ, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. *JAMA*. 2005;293(10):1223–38.
- Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*. 2005;330(7494):765–73.
- Simon JS, Rundall TG, Shortell SM. Adoption of order entry with decision support for chronic care by physician organizations. *J Am Med Inform Assoc*. 2007;14:432–9.
- Kaplan B. Evaluating informatics applications: clinical decision support systems literature review. *Int J Med Inform*. 2001;64:15–37.
- Keeffe B, Subramanian U, Tierney WM, et al. Provider response to computer-based care suggestions for chronic heart failure. *Med Care*. 2005;43(5):461–5.
- Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? *JAMA*. 1999;282(15):1458–65.
- Meehl PE. Causes and effects of my disturbing little book. *J Pers Assess*. 1986;50(3):370–5.
- Arkes HR, Shaffer VA, Medow MA. Patients derogate physicians who use a computer-assisted diagnostic aid. *Med Decis Making*. 2007;27:189–202.
- Promberger M, Baron J. Do patients trust computers? *J Behav Decis Making*. 2006;19:455–68.
- Pezzo MV, Pezzo SP. Physician evaluation after medical errors: does having a computer decision aid help or hurt in hindsight? *Med Decis Making*. 2006;26:48–56.
- Arkes HR, Shaffer VA, Medow MA. The influence of a physician's use of a diagnostic decision aid on the malpractice verdicts of mock jurors. *Med Decis Making*. 2008;28:201–8.
- Smithson M, Verkuilen J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychol Methods*. 2006;11:54–71.
- Smithson M, Merkle EC, Verkuilen J. Beta regression finite mixture models of polarization and priming. *J Educ Behav Stat*. 2011;36:804–31.
- Cribari-Neto F, Zeileis A. Beta regression in R. *J Stat Software*. 2010;34(2):1–24.
- R: A language and environment for statistical computing. Team RDC, editor. Vienna, Austria: R Foundation for Statistical Computing; 2001.
- Medow MA, Arkes HR, Shaffer VA. Are residents' decisions influenced more by a decision aid or a specialist's opinion? A randomized controlled trial. *J Gen Intern Med*. 2010;25(4):316–20.
- Sieck WR, Arkes HR. The recalcitrance of overconfidence and its contribution to decision aid neglect. *J Behav Decis Making*. 2005;18:29–53.
- Wise SL. The development and validation of a scale measuring attitudes toward statistics. *Educ Psychol Meas*. 1985;45:401–5.
- Wallston KA. The validity of the multidimensional health locus of control scales. *J Health Psychol*. 2005;10(5):623–31.
- Wallston KA, Wallston BS, DeVellis R. Development of the multidimensional health locus of control (MHLC) scales. *Health Educ Monographs*. 1978;6(2):160–70.
- Rotter JB. Social learning and clinical psychology. Englewood Cliffs, NJ: Prentice Hall; 1954.